Genetic And Functional Approaches To Understanding Autoimmune And Inflammatory Pathologies

Abbas Raza
University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

Part of the Genetics and Genomics Commons, Immunology and Infectious Disease Commons, and the Pathology Commons

Recommended Citation
Raza, Abbas, "Genetic And Functional Approaches To Understanding Autoimmune And Inflammatory Pathologies" (2020). Graduate College Dissertations and Theses. 1175.
https://scholarworks.uvm.edu/graddis/1175

This Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
GENETIC AND FUNCTIONAL APPROACHES TO UNDERSTANDING AUTOIMMUNE AND INFLAMMATORY PATHOLOGIES

A Dissertation Presented

by

Abbas Raza

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy Specializing in Cellular, Molecular, and Biomedical Sciences

January, 2020

Defense Date: August 30, 2019
Dissertation Examination Committee:

Cory Teuscher, Ph.D., Advisor
Jonathan Boyson, Ph.D., Chairperson
Matthew Poynter, Ph.D.
Ralph Budd, M.D.
Dawei Li, Ph.D.
Dimitry Krementsov, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

Our understanding of genetic predisposition to inflammatory and autoimmune diseases has been enhanced by large scale quantitative trait loci (QTL) linkage mapping and genome-wide association studies (GWAS). However, the resolution and interpretation of QTL linkage mapping or GWAS findings are limited. In this work, we complement genetic predictions for several human diseases including multiple sclerosis (MS) and systemic capillary leakage syndrome (SCLS) with genetic and functional data in model organisms to associate genes with phenotypes and diseases.

Focusing on MS, an autoimmune inflammatory disease of the central nervous system (CNS), we experimentally tested the effect of three of the GWAS candidate genes (SLAMF1, SLAMF2 and SLAMF7) in the experimental autoimmune encephalomyelitis (EAE) mouse model and found a male-specific locus distal to these loci regulating CNS autoimmune disease. Functional data in mouse suggests this male-specific locus modulates the frequency of immune cells including CD11b+, TCRαβ+CD4+Foxp3+, and TCRαβ+CD8+IL-17+ cells during EAE disease. Orchiectomy experiments demonstrate that this male specific phenotype is dependent on testis but not on testosterone (T) or 5α-dihydrotestosterone (DHT). Using a bioinformatic approach, we identified SLAMF8 and SLAMF9 along with other differentially expressed genes in linkage with MS-GWAS predictions, as potential positional candidates regulating CNS autoimmune disease.

Using SCLS, an extremely rare disorder of unknown etiology characterized by recurrent episodes of vascular leakage, we identified and modeled this disease in an inbred mouse strain, SJL, using susceptibility to histamine-and infection-triggered vascular leak as the major phenotypic readout. This trait “Histamine hypersensitivity” (Histh) was mapped to a region on Chr-6. Remarkably, Histh is syntenic to the genomic locus (3p25.3) most strongly associated with SCLS in humans. Subsequent studies found that the Histh locus is not unique to SJL but additional mouse strains also exhibit Histh phenotype. Considering GWAS studies in SCLS are limited by sample size, we utilized interval-specific SNP-based association testing to predict Histh candidates. Furthermore, to dissect the complexity of Histh QTL, we developed network-based functional prediction methods to rank genes in this locus by predicting functional association with multiple Histh-related processes. The top-ranked genes include Cxcl12, Ret, Cacna1c, and Ctnm3, all of which have strong functional associations and are proximal to SNPs segregating with Histh.

Lastly, we utilized the power of integrating genetic and functional approaches to understand susceptibility to Bordetella pertussis and pertussis toxin (PTX) induced histamine sensitization (Bphs), a sub-phenotype with an established role in autoimmunity. Congenic mapping in mice had earlier linked Bphs to histamine H1 receptor gene (Hrh1/H1R) and demonstrated that H1R differs at three amino acid residues in Bphs-susceptible and -resistant mice. Our subsequent studies identified eight inbred mouse strains that were susceptible to Bphs despite carrying a resistant H1R allele. Genetic analyses mapped the locus complementing Bphs to mouse Chr-6, in linkage disequilibrium with Hrh1; we have designated this Bphs-enhancer (Bphse). Similar to the approaches used for Histh, we utilized interval-specific SNP based association testing and network-based functional enrichment to predict nine candidate loci for Bphse including Atp2b2, Atg7, Pparg, Syn2, Ift122, Raf1, Mkcn2, Timp4 and Gt(Rosa)26Sor. Overall, these studies demonstrate the power of integrating genetic and functional methods in humans and animal models to predict highly plausible loci underlying QTL/GWAS data.
CITATIONS

Material from this dissertation has been accepted for publication to *Comm Biol* on September 10, 2019 in the following form:


AND

Material from this dissertation has been submitted for publication to *Genes, Genomes and Genetics* on July 29, 2019 in the following form:

DEDICATION

I dedicate my work to my late father, Shakil Raza Ghazali, and my mother, Shabana Shakil, who sacrificed their goals, dreams and aspirations so I could be successful.
ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my advisor Dr. Cory Teuscher for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Mathew Poynter, Dr. Jonathan Boyson, Dr. Ralph Budd, Dr. Dimitry Krementsov and Dr. Dawei Li for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank my ex-fellow lab mates Dr. Devin Champagne, Dr. Victoria DeVault, Dr. Phyu Thwe and Dr. Mike Secinaro for the stimulating discussions, and for all the fun we have had in the last five years. Also, I would like to thank my lab mom, Ms. Cheryl Collins, for keeping me sane and motivated in research with her wonderful baking and lots of care. My sincere thanks to Dr. Roxana del Rio for guiding me throughout flow cytometry and helping me whenever I needed her support.

Last but not the least, I would like to thank my family: my late father, my mother and my brothers and sister for supporting me spiritually throughout writing this thesis and my life in general.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CITATIONS</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER 1: COMPREHENSIVE LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Introduction to Multiple Sclerosis</td>
<td>1</td>
</tr>
<tr>
<td>Clinical course</td>
<td>2</td>
</tr>
<tr>
<td>Epidemiology</td>
<td>4</td>
</tr>
<tr>
<td>Therapies available</td>
<td>5</td>
</tr>
<tr>
<td>Pathogenesis</td>
<td>7</td>
</tr>
<tr>
<td>Etiology</td>
<td>15</td>
</tr>
<tr>
<td>1.2. Genome wide association studies (GWAS) in MS</td>
<td>20</td>
</tr>
<tr>
<td>1.3. Experimental autoimmune encephalomyelitis (EAE)</td>
<td>27</td>
</tr>
<tr>
<td>Active EAE</td>
<td>27</td>
</tr>
<tr>
<td>Passive EAE</td>
<td>28</td>
</tr>
<tr>
<td>Immunology of EAE</td>
<td>29</td>
</tr>
<tr>
<td>Genetics and sex differences in EAE</td>
<td>36</td>
</tr>
<tr>
<td>1.4. The signaling lymphocytic activation molecule (SLAM) family of receptors</td>
<td>39</td>
</tr>
<tr>
<td>1.5. SLAM family receptor 1, 2 and 7 in MS-GWAS studies</td>
<td>43</td>
</tr>
</tbody>
</table>
CHAPTER 2: SEX-SPECIFIC REGULATION OF CNS AUTOIMMUNITY BY SIGNALING LYMPHOCYTIC ACTIVATION MOLECULE (SLAM) LOCUS

2.1. Abstract........................................................................................................................................ 77
2.2. Introduction.................................................................................................................................. 79
2.3. Results.......................................................................................................................................... 82
  2.3.1. EAE susceptibility in male mice is controlled by a locus in linkage disequilibrium with Slamf1, Slamf2, and Slamf7 ............................................................. 82
  2.3.2. Immune profiling of c1 and B6 mice ....................................................................................... 83
  2.3.3. Increased IL-10 by male c1 T cells co-segregates with resistance to EAE ...... 84
  2.3.4. Gene-by-testis specific interactions modify EAE disease in c1 male mice..... 84
  2.3.5. Lack of evidence for Slamf1, 2 and 7 as the true GWAS candidates regulating CNS autoimmune disease ...................................................................................................... 86
  2.3.6. Polymorphic candidate genes within the refined locus whose expression is testis-dependent are novel candidates regulating CNS autoimmunity, .......... 86
2.4. Discussion..................................................................................................................................... 88
2.5. Material and Methods .................................................................................................................. 93
  2.5.1. Animals ................................................................................................................................... 93
  2.5.2. DNA extraction and Genotyping ............................................................................................ 93
  2.5.3. Induction and evaluation of actively induced EAE ............................................................... 94
  2.5.4. Orchiectomy and hormone administration ............................................................................. 95
  2.5.5. Cytokine and proliferation assays ............................................................................................. 96
  2.5.6. CNS-infiltrating mononuclear cell isolation .............................................................................. 96
  2.5.7. Antibodies and flow cytometric analysis .................................................................................. 97
  2.5.8. In Silico analysis and Data Mining ........................................................................................... 97
2.5.9. Statistics .................................................................................................................. 99
2.6. Acknowledgements ..................................................................................................... 100
2.7. References .................................................................................................................. 101
2.8. Figures ......................................................................................................................... 107
2.9. Tables .......................................................................................................................... 116
2.10. Supplementary Data .................................................................................................. 118

CHAPTER 3: CONCLUSION AND FUTURE DIRECTIONS. ............................................ 121

COMPREHENSIVE BIBLIOGRAPHY .............................................................................. 134

APPENDIX A: A NATURAL MOUSE MODEL REVEALS GENETIC DETERMINANTS OF SYSTEMIC CAPILLARY LEAK SYNDROME (CLARKSON DISEASE) ............................................................................................................. 172

A.1. Abstract ....................................................................................................................... 173
A.2. Introduction .................................................................................................................. 174
A.3. Results ........................................................................................................................ 177
  A.3.1. Dermal vasculature of SCLS patients exhibit hyper-responsive to leak provocateurs ................................................................................................................................. 177
  A.3.2. SJL mice exhibit traits that phenocopy human SCLS ........................................... 177
  A.3.3. SJL mice exhibit age- and inflammation-dependent hypersensitivity to systemic administration of histamine ........................................................................................................... 179
  A.3.4. Histh, the locus controlling susceptibility to vascular hypersensitivity to histamine, exhibits maximal linkage to mouse chromosome 6 ................................................. 180
  A.3.5. Congenic mapping of Histh .................................................................................... 181
  A.3.6. Dermal vasculature of mice harboring an Histh susceptibility allele is hyperresponsive to histamine ................................................................................................................ 181
  A.3.7. Susceptibility to histamine-induced systemic vascular leak is genetically controlled by Histh ............................................................................................................................ 183
A.3.8. Acute viral infectious trigger exacerbates genetically-controlled vascular hyperpermeability. ................................................................. 184

A.3.9. Synteny of Histh locus and SCLS GWAS candidates. ....................... 184

A.4. Discussion .......................................................................................... 186

A.5. Material and Methods ........................................................................ 191
A.5.1. Patients and skin testing ................................................................. 191
A.5.2. Animals .......................................................................................... 191
A.5.3. Histh Phenotyping. ........................................................................ 191
A.5.4. DNA extraction and genotyping. ...................................................... 192
A.5.5. Data resources in the Mouse Phenome Database ............................ 192
A.5.6. Miles assay ..................................................................................... 193
A.5.7. Influenza virus infection. ................................................................. 193
A.5.8. Systemic vascular leak analysis. ....................................................... 194
A.5.9. Linkage analysis and generation of Histh congenic mice ................. 194
A.5.10. Synteny mapping between Histh locus and SCLS GWAS candidates. .... 194

A.6. Acknowledgements ............................................................................ 196
A.7. References .......................................................................................... 197
A.8. Figures ................................................................................................ 201
A.9 Tables .................................................................................................. 208

APPENDIX B: NETWORK-BASED FUNCTIONAL PREDICTION AUGMENTS GENETIC ASSOCIATION TO PREDICT CANDIDATE GENES FOR HISTAMINE HYPERSENSITIVITY IN MICE ........................................... 211

B.1. Abstract ............................................................................................. 212
B.2. Introduction ....................................................................................... 213
B.3. Results ........................................................................................................................................... 216

B.3.1. Generation of Interval Specific Recombinant Congenic Lines (ISRCL) across the Histh locus ......................................................................................................................... 216

B.3.2. Inbred strain survey of Histh ....................................................................................................... 216

B.3.3. Targeted genetic association analysis for Histh .......................................................................... 217

B.3.4. Network-based prediction of Histh-associated genes .................................................................. 217

B.3.5. Integration of functional enrichment with genetic association ................................................. 220

B.4. Discussion ...................................................................................................................................... 222

B.5. Material and Methods ..................................................................................................................... 227

B.5.1. Animals ....................................................................................................................................... 227

B.5.2. Histh Phenotyping ....................................................................................................................... 227

B.5.3. DNA extraction and genotyping .................................................................................................. 227

B.5.4. Generation of Histh congenic lines and GigaMUGA .................................................................. 229

B.5.5. Targeted genetic association testing ............................................................................................ 229

B.5.6. Trait-related gene sets ................................................................................................................ 233

B.5.7. FNTM network ........................................................................................................................... 234

B.5.8. Clustering gene sets .................................................................................................................... 234

B.5.9. Machine learning ...................................................................................................................... 235

B.5.10. Positional Candidate Scoring .................................................................................................... 236

B.5.11. Combined Gene Score ............................................................................................................ 237

B.6. Acknowledgements ......................................................................................................................... 238

B.7. References ...................................................................................................................................... 239

B.8. Figures ............................................................................................................................................ 243
APPENDIX C: GENETIC ANALYSIS OF BPHSE: A NOVEL GENE COMPLEMENTING RESISTANCE TO BORDETELLA PERTUSSIS-INDUCED HISTAMINE SENSITIVITY ........................................................................................................ 252

C.1. Abstract .......................................................................................................................... 254

C.2. Introduction .................................................................................................................. 255

C.3. Results .......................................................................................................................... 259

C.3.1. H1R is highly conserved in mice. .............................................................................. 259

C.3.2. Identification of co-adaptation complementing BphsR. ......................................... 260

C.3.3. A functional linkage disequilibrium domain on Chr 6 regulates susceptibility to Bphs. .............................................................................................................. 261

C.3.4. Identification of candidate genes for Bphse. .......................................................... 262

C.4. Discussion .................................................................................................................... 266

C.5. Materials and Methods ............................................................................................... 272

C.5.1. Animals. ..................................................................................................................... 272

C.5.2. Bphs Phenotyping. .................................................................................................. 272

C.5.3. DNA sequencing of third intracellular loop of Hrh1. ............................................ 273

C.5.4. DNA isolation and Genotyping................................................................................ 273

C.5.5. H1RKO mice genotyping. ....................................................................................... 275

C.5.6. Linkage analysis and generation of Bphs congenic. ................................................. 275

C.5.7. Targeted genetic association testing. ..................................................................... 275

C.5.8. Trait-related gene sets. ........................................................................................... 280

C.5.9. Functional enrichment and ranking of Bphs associated genes. .............................. 280
C.5.10. Combined gene score................................................................. 281
C.6. Acknowledgements ........................................................................... 282
C.7. References ......................................................................................... 283
C.8. Figures ............................................................................................... 289
C.9. Tables ................................................................................................ 292
C.10. Supplementary data ......................................................................... 297
LIST OF FIGURES

CHAPTER 1: COMPREHENSIVE LITERATURE REVIEW
Figure 1.1. MS disease course. ................................................................. 3
Figure 1.2. Likelihood of identifying genetic variants by risk allele frequency and
strength of genetic effect............................................................................ 22

CHAPTER 2: SEX-SPECIFIC REGULATION OF CNS AUTOIMMUNITY BY
SIGNALLING LYMPHOCYTIC ACTIVATION MOLECULE (SLAM) LOCUS
Figure 2.1. B6 mice with overlapping intervals encompassing the 129-derived Slam
locus used in this study. ........................................................................... 107
Figure 2.2. Male c1 mice exhibit increased resistance to EAE. ....................... 108
Figure 2.3. Ex-vivo production of IL-10 by MOG35-55 specific T cells.............. 110
Figure 2.4. EAE severity was enhanced in orchietomized (ORX) c1 mice compared
to sham mice while neither Testosterone (T) nor 5α-dihydrotestosterone (DHT)
replacement ameliorate disease susceptibility. ........................................ 111
Figure 2.5. Generation and EAE testing of interval specific recombinant congenic
(ISRC) lines across c1 interval confirms that EAE susceptibility in male mice is
controlled by a locus in linkage disequilibrium to Slamf1, Slamf2 and Slamf7........ 114
Figure 2.6. Polymorphic candidate genes within the c1 locus whose expression is
modulated by orchiecetomy................................................................. 115

CHAPTER 3: CONCLUSIONS AND FUTURE DIRECTIONS
Figure 3.1. A comparison of GWAS and functional follow-up studies.............. 123

APPENDIX A: A NATURAL MOUSE MODEL REVEALS GENETIC
DETERMINANTS OF SYSTEMIC CAPILLARY LEAK SYNDROME
(CLARKSON DISEASE)
Figure A.1. SCLS patients are hyper-responsive to histamine and morphine......... 201
Figure A.2. SJL mice phenocopy SCLS ..................................................... 202
Figure A.3. Dermal vasculature of mice containing the identified Histh allele is
hyper-responsive to histamine. .................................................................. 203
Figure A.4. Systemic administration of histamine induces vascular leak in SJL and
B10.S-HisthSJL but not B10.S mice............................................................... 204
Figure A.5. Influenza-associated vascular leak in SJL mice as a means to model
SCLS........................................................................................................... 205
Figure A.6. Shared genetic and phenotypic alignment between Histh and SCLS........ 207

APPENDIX B: NETWORK-BASED FUNCTIONAL PREDICTION
AUGMENTS GENETIC ASSOCIATION TO PREDICT CANDIDATE GENES
FOR HISTAMINE HYPERSENSITIVITY IN MICE.
Figure B.1. Interval specific recombinant congenic (ISRC) mapping places Histh
candidates in four genetic loci. ............................................................... 243
Figure B.2. Targeted genetic association analysis for Histh................................ 244
Figure B.3. Workflow Overview. ................................................................. 245
Figure B.4. Two axes of gene scoring. ......................................................... 246
Figure B.5. Final gene scores................................................................. 247
Figure B.6. Network-based machine learning for functionally annotating genes. .......... 249

APPENDIX C: GENETIC ANALYSIS OF BPHSE: A NOVEL GENE COMPLEMENTING RESISTANCE TO BORDETELLA PERTUSSIS-INDUCED HISTAMINE SENSITIVITY.

Figure C.1. Congenic mapping of Bphse confirms linkage to Bphs/Hrh1. ............... 289
Figure C.2. Integration of genetic and functional mapping approaches to predict candidates for Bphse. ........................................................................... 290
Supplementary Figure C.1. Mouse family tree showing the phylogenetic relationships among 102 inbred and wild-derived inbred strains. .............................................. 298
LIST OF TABLES

CHAPTER 1: COMPREHENSIVE LITERATURE REVIEW
Table 1.1. Disease modifying therapies approved for use in MS.............................. 6
Table 1.2. Published GWAS studies in MS................................................................. 24
Table 1.3. EAE loci in mouse models.......................................................................... 37
Table 1.4. The SLAM family and signaling adaptors................................................. 40

CHAPTER 2: SEX-SPECIFIC REGULATION OF CNS AUTOIMMUNITY BY
SIGNALLING LYMPHOCYTIC ACTIVATION MOLECULE (SLAM) LOCUS
Table 2.1. SLAM locus is highly conserved between human and mouse with natural
 genetic variation segregating between B6 and 129 mice................................. 116
Table 2.2. Immune profile of c1 and B6 mice............................................................... 117
Supplementary Table 2.1. Clinical disease metrics of c1, c2 and B6 mice.............. 118
Supplementary Table 2.2. Primers used in genotyping B6.129 congenic lines........ 119

APPENDIX B: NETWORK-BASED FUNCTIONAL PREDICTION
AUGMENTS GENETIC ASSOCIATION TO PREDICT CANDIDATE GENES
FOR HISTAMINE HYPERSENSITIVITY IN MICE.
Table B.1. A survey of Histh phenotypes across 23 inbred mouse strains............ 250

APPENDIX C: GENETIC ANALYSIS OF BPHSE: A NOVEL GENE
COMPLEMENTING RESISTANCE TO BORDETELLA PERTUSSIS-
INDUCED HISTAMINE SENSITIVITY.
Table C.1. Distribution of H1R₅ and H1R₆ alleles in inbred laboratory and wild-
derived mouse strains................................................................. 292
Table C.2. Bphs susceptibility of mice with the H1R₆ allele.................................. 293
Table C.3. Bphs susceptibility in (H1R₅ × H1RKO) F₁ hybrids................................. 294
Table C.4. Linkage of chromosome 6 marker loci to Bphse.................................. 295
Table C.5. Predicted candidates for Bphse and their annotation in Bphs functional
 networks................................................................................................. 296
Supplementary Table C.1. List of 50 inbred mouse strains used in genetic association
 testing........................................................................................................ 297
CHAPTER 1: COMPREHENSIVE LITERATURE REVIEW

1.1. Introduction to Multiple Sclerosis

Multiple sclerosis (MS) is the most common progressive neurologic disease of young adults worldwide (Feigin, 2017). Earliest records suggestive of MS date back to 1300s in Schiedam, Holland about a woman named Lidwina the Virgin (Orrell, 2005). In year 1395 at age 16 years she developed a strange illness that exhibited as blindness in one eye, weakness and pain that subsequently led her to fall while skating on a frozen canal. She died in 1433. Moreover, records from the personal diary of Sir Augustus d'Esté (1794-1848), grandson of King George III of England, reveals almost certainly that he suffered from MS. His symptoms started at age 28 with a transient visual loss. Later as the disease progressed, he experienced double vision (diplopia), weakness and stiffness of the legs (spastic paraparesis), lack of control over defecation (anal incontinence) and erectile dysfunction (impotence). He spent the rest of his life wheelchair bound until his death in 1848 (Landtblom et al., 2010b).

The first illustration of pathological findings in MS is attributed to Robert Carswell, a Scottish professor of pathology, in his work Illustrations of the Elementary Forms of Disease published in 1837 in which he drew the macroscopical appearances of the plaques seen in multiple sclerosis (Murray, 2009). Jean Martin Charcot, a French neurologist at the Hôpital de Salpêtrière in Paris, is attributed to having coined the term sclérose en plaques disseminées in 1868, which later became known as ‘disseminated sclerosis’ and in the 1960s as ‘multiple sclerosis’. He described through a series of seminal
lectures the pathology: demyelination (“nervous tubes...have lost their myelin cylinder”) and axonal loss (“in the plaque center, nervous tubes are reduced and thinner”), diagnostic criteria: nystagmus (involuntary eye movement), intention tremor (rhythmic shaking) and scanning speech; and disease onset (young adults with predominance in women) of MS (Charcot, 1868; Lubetzki, 2018). The diagnostic criteria have been revised several times to incorporate additional features including those by Marburg in 1909 (Uhtoff’s sign, absence of abdominal reflexes and pyramidal tract lesions), Broman in 1965 (oligoclonal bands in cerebrospinal fluid, dissemination in space and time), and recently by McDonald in 2001, 2005, 2010 and 2017 (Magnetic Resonance Imaging (MRI) evidence) (Gafson et al., 2012; Thompson et al., 2018).

**Clinical course**

MS is a heterogeneous disease and exhibits variability in disease course and rate of progression to disability from person to person (Weiner, 2009). Kurtzke developed the Expanded Disability Status Scale (EDSS) to quantitate symptoms by assigning weights to pyramidal, cerebellar, brain stem, sensory, bowel, bladder, cerebral, visual and other deficits in patients (Kurtzke, 1983). Using the EDSS system, the National Multiple Sclerosis Society Advisory Committee on Clinical Trials in MS (Lublin et al., 2014) has defined four clinical courses (Figure 1.1A): clinically isolated syndrome (CIS; single neurological episode lasting 24 hours or more), relapsing remitting MS (RRMS; periodic relapse and exacerbation with partial remission), primary progressive MS (PPMS; gradual worsening of symptoms with no relapse or remission) and secondary progressive MS (SPMS; transition from relapsing remitting to progressive disease) (Lublin et al., 2014).
Most people diagnosed with MS transition from one form to another as the disease progresses. The disease course of MS for Augustus d’Esté (Landtblom et al., 2010a) is a prime example of clinical heterogeneity. He started with CIS that transitioned to RRMS and later to SPMS until his death in 1848 (Figure 1.1B).

**Figure 1.1. MS disease course.**
(A). Four clinical courses are defined for MS. (B) The disease course of Augustus d’Esté
Epidemiology

Approximately 2.3 million people in the world have MS and the disease is usually found to be more prevalent in women than men (Feigin, 2017). Most people are diagnosed between the ages 20 and 50 years, although MS can also occur in young children (pediatric MS) and older adults (late onset MS). The prevalence of MS varies geographically, with high levels in North America, Northern Europe, Australia and New Zealand (>100/100,000 inhabitants) but low levels in Eastern Asia and Africa (2/100,000 inhabitants) (Wade, 2014). Several population-based studies have shown a steady increase in the prevalence and incidence of MS worldwide (Etemadifar et al., 2013; Kingwell et al., 2013). In US, prevalence estimates were calculated in 1976 (123,000), 1990 (180,000), 2008 (572,312), 2010 (727,344) and 2017 (913,925) and exhibit a similar trend (Wallin et al., 2019). Alonso et al., compared the mean incidence of MS cases among 38 population-based studies between 1966 and 2007 and found the rise is not uniform across all age groups and genders (Alonso et al., 2008). While the incidence has generally increased in both men and women over the years, there is a disproportional increase in incidence among women >50 years and could explain the increase in sex ratio from 2.4 ♀:1 ♂ to 3.5 ♀:1♂ over the last 50 years (Ramagopalan et al., 2010). Possible explanations for the increased incidence include improved access to medical facilities, better diagnostic tools and increase in life expectancy in general. However, this cannot explain the disproportionate increase in MS among women.

Migration studies in MS have shown that individuals who travel below the age of 15 or 16 from an area where the disease is common to an area where it is rarer show a
decrease in rate of disease (Dean et al., 1971; Kurtzke et al., 1985; Gale et al., 1995). The same is true for migration to high-risk areas. Dean et al., has reported that immigrants to England from India and Pakistan at an age younger than 15 years had a higher risk of developing MS than those immigrating after that age (Dean et al., 1997). This suggests that the years preceding adulthood can shape the course of disease late in life.

**Therapies available**

There is no cure for MS. Several therapies were tried during the 1800s and 1900s including injection of malarial parasites, arsenic, mercury, corticosteroids, cyclophosphamide, cyclosporin, methotrexate, and azathioprine with little to no success (Patwa, 2014). The first positive trial using interferon beta-1b (IFN-β-1b) was reported in 1993 (Duquette et al., 1993) and earned Food and Drug Administration (FDA) approval as a prescribed treatment to reduce disability and the number of relapses. Currently, several additional disease-modifying therapies (DMT) are available to decrease the frequency of relapses and/or to delay disease progression in MS (Table 1.1). These DMTs have varying mechanisms of action with immunosuppressive and immunomodulatory effects on leukocyte numbers (alemtuzumab, ocrelizumab, cladribine), proliferation (teriflunomide, mitoxantrone), trafficking (fingolimod, natalizumab) and cytokine production (interferon beta, glatiramer acetate). These treatments are not cheap and could range from $8,528 to $52,244 annually per patient depending on the clinical course and severity of disease (Marcus et al., 2013; Ernstsson et al., 2016; De Angelis et al., 2018).
Table 1.1. Disease modifying therapies approved for use in MS.

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Active ingredient</th>
<th>Target</th>
<th>FDA</th>
<th>MS type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Avonex®</td>
<td>Interferon beta-1a</td>
<td>Unknown; Immunoregulatory</td>
<td>1996</td>
<td>RRMS, CIS</td>
</tr>
<tr>
<td>2 Betaseron®</td>
<td>Interferon beta-1b</td>
<td>Unknown; Immunoregulatory</td>
<td>1993</td>
<td>RRMS, SPMS, CIS</td>
</tr>
<tr>
<td>3 Copaxone®</td>
<td>Glaterimer acetate</td>
<td>leukocyte differentiation</td>
<td>1996</td>
<td>RRMS, CIS</td>
</tr>
<tr>
<td>4 Extavia®</td>
<td>Interferon beta-1b</td>
<td>Unknown; Immunoregulatory</td>
<td>2009</td>
<td>RRMS</td>
</tr>
<tr>
<td>5 Mayzent</td>
<td>Siponimod</td>
<td>Binds sphingosine-1-phosphate receptor (S1P); reduce leukocyte migration to CNS</td>
<td>2019</td>
<td>RRMS, SPMS, CIS</td>
</tr>
<tr>
<td>6 Glatopa®</td>
<td>Glaterimer acetate</td>
<td>Leukocyte differentiation</td>
<td>2015</td>
<td>RRMS</td>
</tr>
<tr>
<td>7 Plegridy®</td>
<td>Peginterferon beta-1a</td>
<td>Unknown; Immunoregulatory</td>
<td>2014</td>
<td>RRMS</td>
</tr>
<tr>
<td>8 Rebif®</td>
<td>Interferon beta-1a</td>
<td>Unknown; Immunoregulatory</td>
<td>1998</td>
<td>RRMS, CIS</td>
</tr>
<tr>
<td>9 Aubagio®</td>
<td>Teriflunomide</td>
<td>Inhibits dihydroorotate dehydrogenase; reduce leukocyte proliferation</td>
<td>2012</td>
<td>RRMS</td>
</tr>
<tr>
<td>10 Gilenya®</td>
<td>Fingolimod</td>
<td>Binds S1P; reduce migration of leukocytes to CNS</td>
<td>2010</td>
<td>RRMS</td>
</tr>
<tr>
<td>11 Tecfidera®</td>
<td>Dimethyl fumarate</td>
<td>activate nuclear erythroid 2-related factor 2; neuroprotection</td>
<td>2013</td>
<td>RRMS</td>
</tr>
<tr>
<td>12 Lemtrada®</td>
<td>Alemtuzumab</td>
<td>Bind and depletes CD52 bearing mature lymphocytes</td>
<td>2014</td>
<td>RRMS</td>
</tr>
<tr>
<td>13 Novantrone®</td>
<td>Mitoxantrone</td>
<td>Type II topoisomerase inhibitor; inhibits leukocyte proliferation</td>
<td>2000</td>
<td>RRMS, SPMS</td>
</tr>
<tr>
<td>14 Ocrevus®</td>
<td>Ocrelizumab</td>
<td>Targets CD20; B cell depletion</td>
<td>2017</td>
<td>RRMS, PPMS</td>
</tr>
<tr>
<td>15 Tysabri®</td>
<td>Natalizumab</td>
<td>Binds α4-integrin; prevent migration of leukocytes into CNS</td>
<td>2006</td>
<td>RRMS</td>
</tr>
<tr>
<td>16 Mavenclad</td>
<td>Cladribine</td>
<td>Purine analogue; reduce leukocyte numbers</td>
<td>2019</td>
<td>RRMS</td>
</tr>
</tbody>
</table>

CIS: clinically isolated syndrome; RRMS: relapsing remitting MS; PPMS: primary progressive MS; SPMS: secondary progressive MS.
Pathogenesis

CNS lesions are a pathologic hallmark in all types of MS (Huang et al., 2017). These are areas of demyelination close to post capillary venules both in white and grey matter, along the ventricles, optic nerves, corpus callosum, cerebellar peduncles, long tracts and subpial region of the spinal cord and brain and are regions where blood-brain barrier (BBB) permeability is compromised (Compston et al., 2008). Myelin is a specialized membrane of lipids and proteins made by Schwann cells in the peripheral nervous system (PNS) and oligodendrocytes in CNS and wraps around axons allowing for rapid and efficient propagation of electrical impulses (Podbielska et al., 2013). The exact cause of demyelination in MS is unknown. However, genetic, pathological and animal studies have strongly suggested a role for the immune system in mediating immune attack against myelin antigens (Gay et al., 1997; Denic et al., 2011; Sawcer et al., 2011). The infiltrating population of immune cells express high levels of integrins including leukocyte function associated antigen-1 (LFA-1) and very late antigen-4 (VLA-4) (Claudio et al., 1995). LFA-1 and VLA-4 are ligands for intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1) respectively, which are expressed by endothelial cells that line the BBB in the CNS (van de Stolpe et al., 1996). This LFA-1/ICAM-1 and VLA-4/VCAM-1 interaction helps immune cells extravasate into the CNS.

The active lesions of MS are characterized by immune cell infiltrates composed of clonally expanded CD4+ (perivascular cuff), CD8+ (parenchyma) and γδ T cells (perivascular cuff), B cells (perivascular cuffs and parenchyma) and macrophages (perivascular cuffs and parenchyma), suggesting an antigen specific response (Traugott et
In this regard, several antigens are proposed as targets of autoreactive T cells and autoreactive antibodies in MS including proteins of myelin or oligodendrocytes (myelin basic protein [MBP], myelin oligodendrocyte glycoprotein [MOG], proteolipid protein [PLP], myelin-associated glycoprotein [MAG]), 2’,3’-cyclic nucleotide 3’-phosphodiesterase [CNPase], myelin-associated oligodendrocyte basic protein [MOBP], non-myelin proteins αB-crystallin [HspB5], S100β, transaldolase-H, contactin-2/TAG-1), and non-protein antigens (glycolipids) (Hohlfeld et al., 2016). Since autoreactive B and T cells are part of the normal lymphocyte repertoire (Wucherpfennig, 1994), it is not fully understood how, where and when they are activated, their selective recruitment into the CNS, and their reactivation and maintenance within the CNS during MS disease. One hypothesis is the initial damage in MS takes place within the CNS, which leads to the drainage of myelin antigens to the periphery where they get picked up by the immune system that unleashes an autoimmune response against the CNS (Caprariello et al., 2018). An alternate hypothesis is that the initial events in MS happen outside the CNS possibly as a result of microbial (Berer et al., 2017) or viral (Fujinami et al., 1985; Kakalacheva et al., 2011) trigger with cross-reactivity to myelin antigens.

Within the CNS, T cells encounter cognate myelin antigens on antigen presenting cells (APC) and are reactivated setting up an inflammatory cascade that triggers recruitment of more T cells, as well as B cells, dendritic cells (DC), microglia and natural killer (NK) cells. Several cells can serve as APC in the CNS, including microglia, DC, macrophages and B cells, and provide co-stimulatory signals to myelin-reactive T cells.
Activated macrophages/microglia can phagocytose myelin directly or, along with T cells and the CNS glial cells, can release cytotoxic products and mediators such as proteases, inflammatory cytokines (tumour necrosis factor-α, interferon-γ), soluble reactive oxygen and nitrogen intermediates, free radicals and glutamic acid (Selmaj et al., 1988; Cuzner et al., 1999; Smith et al., 1999; Neumann et al., 2002). Infiltrating B cells and complement further damage myelin. As a result of CNS damage, additional neoantigens are released into the periphery (epitope spreading) that prime further autoreactive B and T cells and subsequent invasion into the CNS (Tuohy et al., 1997).

In the earliest clinical presentation of MS, CIS, increased peripheral CD4+ T cell activation is linked to the occurrence of a second attack (Corvol et al., 2008). The central role of CD4+ T cells in MS has been further explored by immune and functional profiling in MS patients (Chitnis, 2007). Naïve CD4+ T cells differentiate into several types of T helper (Th) cell lineages including Th1, Th2, Th17, Follicular B helper cells (T\textsubscript{FH}), and regulatory T cells (Treg) depending on antigenic stimulation and cytokine milieu (Luckheeram et al., 2012). These Th lineages produce distinct cytokines (Th1: interferon gamma (IFN-γ), tumor necrosis factor alpha (TNF-α), and lymphotoxin (LT); Th2: interleukin 4 (IL-4), interleukin 5 (IL-5), and interleukin 13 (IL-13); Th17: interleukin 17A (IL-17A), interleukin 17F (IL-17F), interleukin 21 (IL-21), and interleukin 22 (IL-22); T\textsubscript{FH}: interleukin 2 (IL-2), IL-4; and Treg: interleukin 10 (IL-10) and tumor growth factor beta (TGF-β). Historically, Th1 cells were thought to be the main pathogenic T cells in MS: 1) increased levels of Th1 cytokines are evident during MS relapse, whereas Th2 cytokines
are found increased during remission in MS patients (Imam et al., 2007); 2) administration of IFN-γ in relapsing/remitting MS patients can exacerbate disease (Panitch et al., 1987b); 3) animal models of MS have demonstrated that mice lacking in the small subunit of interleukin-12 (IL-12) cytokine, IL-12p40<sup>−/−</sup>, or T-box transcription factor (T-bet), T-bet<sup>−/−</sup>, are resistant to CNS autoimmune disease (Gran et al., 2002; Bettelli et al., 2004), and 4) treatment with glatiramer acetate, a peptide of MBP, reduces relapses in MS due to shift of pathogenic Th1 cells to protective Th2 cells (Oreja-Guevara et al., 2012). Th1 cells may contribute to the development and progression of disease by secreting IFN-γ (activates macrophages to secrete nitric oxide and promote myelin destruction), TNF-α and TNF-β (direct damage to myelin sheath and oligodendrocytes) cytokines (Selmaj et al., 1988; Chu et al., 2018).

The Th1 driven model of MS disease was challenged by observations that IFN-γ and IFN-γ-receptor-deficient mice, as well as mice that lack other molecules involved in Th1 differentiation, such as IL-12p35, IL-12 receptor β2, were not protected from CNS autoimmune disease, but instead were more susceptible to the disease (Krakowski et al., 1996; Gran et al., 2002; Zhang et al., 2003). This became apparent with the discovery of interleukin 23 (IL-23), which is related structurally to IL-12 (Cua et al., 2003). IL-23 shares the p40 chain with IL-12, which is associated with either p19 (IL-23) or p35 (IL-12), respectively. IL-23 was shown to be necessary to drive the induction or expansion of Th17 (Langrish et al., 2005). In addition to IL-23, Th17 differentiation program require interleukin 6 (IL-6), IL-21, TGF-β, and receptor-related orphan receptor gamma-T (RORγt) as the master regulator (Yang et al., 2008). Analysis of peripheral blood
mononuclear cells (PBMCs), cerebrospinal fluid (CSF) and CNS lesions among MS patients found higher levels of IL-17, TGF-β and IL-6 transcripts (Lock et al., 2002). An increased frequency of Th17 cells is detected in PBMC and CSF of some RRMS and CIS patients compared with noninflammatory neurological diseases (Matusevicius et al., 1999; Tzartos et al., 2008). In vitro BBB migration assays have shown that Th17 cells transmigrate efficiently, highly express granzyme B and display enhanced cytotoxic activity against neurons compared with Th1 cells (Kebir et al., 2007). Th17 cell also exhibit impressive plasticity and lymphocytes expressing both IL-17 and IFN-γ have been found in MS brain tissue, suggesting that IL-17+ IFN-γ+ T cells may be involved in pathology (Kebir et al., 2009). Secukinumab is a monoclonal antibody against IL-17A, and a phase II clinical trial has shown to reduce MRI lesions among RRMS patients (Havrdová et al., 2012).

Other effector T cell subsets important in MS include, granulocyte macrophage colony stimulating factor (GM-CSF) expressing CD4+ (Th-gm), CD8+ T cells, gamma delta (γδ) T cells, regulatory T cells (nTreg, iTreg, Tr1, Tfr and iTr35) and B cells (Rasouli et al., 2015; Willing et al., 2018). Th-gm are elevated in the CSF of MS patients compared to the blood and can secrete proinflammatory cytokines including IFN-γ and GM-CSF upon exposure to IL-12 (Noster et al., 2014; Rasouli et al., 2015; Restorick et al., 2017). These cytokines can promote the activation, maturation, and differentiation of macrophages and of DC cells respectively (Sonderegger et al., 2008).
One of the major populations infiltrating CNS lesions among MS patients are CD8+ T cells (Booss et al., 1983). These cells express granzyme B, perforin, IFN-γ, and IL-17 (Tc17). They are enriched in CSF of RRMS patients (Salou et al., 2015), and are thought to contribute to direct injury of neurons (Kebir et al., 2007; Huber et al., 2013). Blockade of CD8+ T cell migration into the CNS reduces clinical disease in animal models suggesting potential role in the development of MS.

γδ T cells are a unique class of immune that can be activated with or without their cognate T-cell receptor (TCR) ligands and are an early effector T cell population during an immune response. These cells can produce IL-17 upon stimulation with interleukin 1 beta (IL-1β) and can amplify Th17 autoimmune responses (Sutton et al., 2009). γδ T cells are found in the CNS lesions of MS patients (Wucherpfennig et al., 1992). Animal models of MS strongly suggest contribution of γδ T cells in disease pathogenesis (Blink et al., 2014a).

The aberrant activation of several of these effector cells in MS is kept in check by regulatory T cells (iTreg and nTreg) expressing forkhead box protein P3 (FOXP3) transcription factor (Kitz et al., 2018). iTreg cells are FOXP3+CD4+CD25+ cells, which develop in peripheral lymphoid organs after antigen priming, in contrast to the natural Treg cells (nTreg) which are released from the thymus as a distinct lineage with FOXP3 already expressed (Chen et al., 2003). Activation of naïve T cells in the presence of TGF-β results in the expression of Foxp3 and the generation of iTregs. The activity of myelin reactive T cells, which are present in healthy individuals, is kept in check through several immunosuppressive mechanisms. These include secretion of IL-10, interleukin 35 (IL-35), TGF-β, granzyme B, adenosine, induction of immunosuppressive indoleamine 2,3-
dioxygenase (IDO), and inhibition of effector T cells by consumption of IL-2 or negative signaling by cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) (Roncarolo et al., 2006; Schmidt et al., 2012). These regulatory mechanisms are however impaired in MS patients where reduced numbers of circulating nTreg, decreased expression of FOXP3, loss of IL-10 secretion and impairment of regulatory capacity are observed (Astier et al., 2006; Venken et al., 2008; Frisullo et al., 2009; Dhaeze et al., 2015).

While earlier view about MS predominantly focused on T lymphocytes as the key cell type that mediates autoimmune damage, emerging evidence suggest that B lymphocytes can also contribute to disease. Recent successes of CD20+ B-cell targeting therapies (rituximab, ocrelizumab, and ofatumumab) have renewed interest in their role in MS pathogenesis (Sabatino et al., 2019). 95% of MS patients have IgG oligoclonal bands (OCB) in CSF but not in their blood, suggesting a compartmentalized B-cell response (Kabat et al., 1942; Link et al., 2006). These OCB are long lived consistent with in situ local antigen stimulation and Ig production. Some of these OCB are reactive against myelin (Genain et al., 1995; von Budingen et al., 2008). Recently, Brändle et al. reported that OCB in MS target ubiquitous intracellular proteins, which may suggest a protective role (Brandle et al., 2016). Other studies found OCB antibodies targeting against many virus and infectious agents like Epstein-Barr virus (EBV), measles, mumps, herpes simplex virus type-1 (HHV-1) and type-6 (HHV-6), varicella-zoster virus, cytomegalovirus, rotavirus, β-hemolytic streptococcus, Haemophilus influenzae type B, Escherichia coli, and enterococcus (Winger et al., 2016). Additional studies are needed to examine the specificity of OCB in MS and to determine if they participate in pathogenesis.
The initiator and effector functions of T and B lymphocytes are coordinated by
the innate immune cells including, DCs and microglial cells, and their contribution to MS
disease has become increasingly recognized (Gandhi et al., 2010). DCs are professional
APC and promote activation and differentiation of naïve T cells into Th1, Th2, Th17 or
iTregs. Two major DC subsets can be discriminated in blood: myeloid or conventional DCs
(cDCs) and plamacytoid DCs (pDCs). cDCs phenotypically express CD11c, human
leukocyte antigen (HLA)-DR, blood dendritic cell antigen (BDCA)-1, and absence of
CD123 (Merad et al., 2013). In addition, they express toll-like receptor (TLR) 2 and TLR4.
Activated cDCs secrete IL-12p70 that can help polarize Th cells into Th1 (Schulz et al.,
2000). In contrast, pDCs are identified by a CD11c− HLA− DR+ CD123high phenotype. They
express TLR7 and TLR9 and upon activation produce type I interferon (IFN-I) (Kadowaki
et al., 2001). Importantly, both cDCs and pDCs are found CNS lesions and CSF of MS
patients (Serafini et al., 2006). cDCs isolated from peripheral blood of RRMS and SPMS
patients have an activated phenotype, with increased expression of activation markers such
as cluster of differentiation (CD) 40 and CD80, decreased expression of immunoregulatory
molecule programmed death ligand-1 (PD-L1), and higher levels of Th1 and Th17
polarizing cytokines (IFNγ, IL-12, TNF-α, IL-6 and IL23) (Karni et al., 2006). pDCs
isolated from PBMC of MS patients exhibit decreased expression of activation markers
including CD86, CD83, CD40 and 4-IBBL (Stasiolek et al., 2006).

Microglia are the resident immune cells of the CNS and express HLA class II and
costimulatory molecules CD83 and CD40, which are essential for interaction with effector
T and B cells (Stasiolek et al., 2006). In addition, microglial cells express all nine TLRs
and expression of these receptors is necessary for the generation of CNS autoimmune responses (Jack et al., 2007). They secrete a variety of proinflammatory cytokines (IL-1β, IL-6, and TNF-α), are capable of phagocytosis, and can recruit peripheral cells to the site of injury (Lampron et al., 2013). Activated microglia are present during early and late MS stages and express proinflammatory markers involved in phagocytosis (CD68), in antigen presentation (HLA Class I and II, CD86), and production of reactive oxygen species (p22phox) (Woodroofe et al., 1986; Bogie et al., 2014). The recent discovery of transmembrane protein 19 (TMEM19) as a unique marker for microglia has significantly advanced our understanding of their role in MS. Analysis of TMEM19+ cells among different stages of CNS lesions in MS have shown that the initial tissue is associated with activated microglia, and these cells after phagocytosis of tissue debris transform into cells with a macrophage phenotype (Woodroofe et al., 1986).

Etiology

MS is a multifactorial disease with several risk factors, which are broadly grouped into environment, sex and genetics. These factors appear to act synergistically to modify the overall risk for disease.

Environmental risk. There is strong evidence that exposure to environmental factors during gestation and/or early in childhood can influence susceptibility to MS later in life (Goodin, 2009). Risk promoting factors include exposure to EBV (odds ratio 3.6), HHV-6 (odds ratio 1.6), active and passive smoking (odds ratio 1.3-1.6), vitamin D deficiency <50mM in serum (odds ratio 1.4), reduced exposure to UV radiation (odds ratio 2.0), month of birth (odds ratio 1.3), cigarette smoking (odds ratio 2.4), and adolescent
obesity (odds ratio 2.0) (Filippi et al., 2018). MS risk is reduced by exposure to cytomegalovirus (odds ratio 0.7), fasting mimicking or ketogenic diets (odds ratio 0.5) and fish consumption (odds ratio 0.55) (O’Gorman et al., 2012).

**Sex differences.** Increased incidence (Wallin et al., 2012), slower progression (Tremlett et al., 2006), earlier onset (Bergamaschi, 2007) and amelioration of symptoms during pregnancy (Airas et al., 2012) among women all support the important contribution of sex in MS susceptibility. These effects may be driven by differences in sex chromosomes, sex hormones or both. Animal models of MS have shown that presence of two X-chromosomes (Chr-X) increases susceptibility to CNS autoimmune disease independent of hormones (Smith-Bouvier et al., 2008). Chr-X inactivation and genomic imprinting effects can additionally skew expression of genes implicated in autoimmunity including Toll-like receptor 7 (TLR7), CD40 ligand (CD40L) and FoxP3 between men and women (Sellebjerg et al., 2012). Conversely, several genes on the Chr-Y are known to provide resistance to CNS autoimmune disease in animals models of MS (Case et al., 2015). Moreover, maternal transmission of MS is linked to mitochondrial effects (Ban et al., 2008).

The role of sex hormones in MS susceptibility is apparent in women during pregnancy, when high levels of estriol seem to ameliorate symptoms especially during the third trimester (Confavreux et al., 1998b). Symptoms worsen post-partum before returning to the pre-pregnancy rate (Confavreux et al., 1998a). Some studies have demonstrated that pregnancy could offer long term protection in MS by delaying progression to disability (Runmarker et al., 1995). Others suggest that the immune response during pregnancy is
modulated by sex hormones with a shift to Th2 phenotype and an increase in the levels of regulatory T cells (Garay et al., 2007; Tai et al., 2008; Sherer et al., 2017). Preclinical studies of MS show that estriol treatment has both anti-inflammatory and neuroprotective properties (Spence et al., 2012). However, human phase II trials with estriol have shown only modest benefits in reducing relapses among women with MS (Voskuhl et al., 2016).

Numerous studies have demonstrated that androgens are protective in MS with low levels associated with an increased risk of MS (Pakpoor et al., 2014). The four androgen hormones, 5α-dihydrotestosterone (DHT), testosterone, androstenedione, and dehydroepiandrosterone (DHEA), are all synthesized from cholesterol in the gonads and adrenal glands. DHT is more potent than testosterone, while androstenedione and DHEA have only 10 and 5% of the potency of testosterone, respectively (Marchetti et al., 2013). Testosterone can be converted to androstenedione (and vice versa) and both can be aromatized to estrogens by the enzyme aromatase (Remage-Healey et al., 2011). DHEA can be metabolized to testosterone and estrogens. Thus, studies in which testosterone, androstenedione and DHEA have been used for in vivo treatment can be difficult to interpret. DHT alone cannot be converted to estrogens and thus, studies utilizing DHT are most easily interpreted.

Among men with MS, low levels of testosterone correlate with worse EDSS scores (Bove et al., 2014). A phase II trial of testosterone treatment has shown benefits in RRMS patients by reducing brain atrophy in grey and white matter and improving cognition and spatial memory in participants (Sicotte et al., 2007). Further support comes from animal studies in which castration is linked to increased disease severity, whereas
exogenous supplementation with testosterone ameliorates disease by reducing inflammation, demyelination and axonal damage (Dalal et al., 1997b; Palaszynski et al., 2004b). Mechanistically, testosterone and its more biologically active form, DHT, can act through the androgen receptor (AR) to exert their biological actions (Davey et al., 2016). The expression of AR in various immune organs and multiple immune cells (T lymphocytes, B lymphocytes, macrophages, DC, and mast cells) provides some indication of the level at which androgens influence immunity.

Increased levels of androgens in males may protect from autoimmunity by maintaining higher levels of thymic autoimmune regulator (Aire) expression. Aire exerts T-cell self-tolerance by promoting expression of tissue-specific antigens (TSA) in medullary thymic epithelial cells. Developing thymocytes that recognize these TSAs with high affinity undergo negative selection, thus preventing the release of self-reactive T cells into the periphery and thus autoimmunity.

Genetic susceptibility. The contribution of genetic inheritance in MS is not new and is supported by several observations. These include higher MS incidence and prevalence in some ethnic populations (Caucasians, Sardinians, Parsis and Palestinians) while being uncommon in Uzbeks, Samis, Turkmen, Kyrgyzis, Kazakhs, native Siberians, North and South Amerindians, Japanese, Chinese, African blacks, and New Zealand Maori, which suggested genetics as an important component of disease (Rosati, 2001). Genetic relatedness and predisposition to MS was further supported by familial clustering of disease in which siblings of affected individuals had 15-20 fold higher risk (2-4%) of developing disease compared to the general population (0.2%) (Harirchian et al., 2018).
Classical twin studies in MS puts the age-adjusted risk at 35% for monozygotic twins, 6% for dizygotic twins and 3% in siblings reinforcing direct relationship between genetic relatedness and susceptibility (Filippi et al., 2018). In addition, it also suggests that the genetic architecture underlying susceptibility may involve several polymorphic loci and non-Mendelian inheritance of disease. Early attempts in the 1980s to uncover the genetic contribution through linkage studies among multiple-affected member MS families established a strong contribution of chromosome 6p21.3 (Jersild et al., 1972; Naito et al., 1972). This ~4Mb large interval spans 220 closely linked genes and includes six of the classical human leukocyte antigen (HLA) genes: the class I genes HLA-A, HLA-B, and HLA-C, and the class II genes HLA-DPB1, HLA-DQB1, and HLA-DRB1 consistent with the notion that MS is an autoimmune disease (Sawcer et al., 1996). Among the 24000 alleles describe to date for HLA genes (Robinson et al., 2015), the HLA-DRB1*15:01 allele, has the strongest effect on MS susceptibility in both Caucasians and non-Caucasians, with an odds ratio of 3.08 (Mosca et al., 2017). It remains to be determined how and why the selection of this allele influence disease risk.

Early success in the search for genetic loci in MS led to subsequent large and ambitious genome wide-microsatellite screens or candidate gene studies but they failed to identify additional linkages outside of the HLA region (Barcellos et al., 2003; Sawcer et al., 2005). These studies made clear that MS may not be caused by a small number of mutations having large effects that can be identified using traditional linkage analysis but is likely a result of several polymorphic loci each having a modest contribution to overall disease.
1.2. Genome wide association studies (GWAS) in MS.

The completion of the human genome sequencing and HapMap projects in early 2000 provided a wealth of information about common genetic variation across the genome (Collins et al., 2003). Subsequent advances in high throughout genotyping technologies made it possible to profile large numbers of samples in a cost-effective manner. These allowed screening of thousands of unrelated individuals in population-based studies to determine association of a genetic variant with particular trait in case-control subjects. The first GWAS was performed for myocardial infarction (MI) using 92,788 gene-based single-nucleotide polymorphism (SNP) markers in a Japanese cohort of 1133 affected individuals and 1006 controls (Ozaki et al., 2002). They discovered a functional SNP in the lymphotoxin-alpha gene that was associated with susceptibility to MI. This was followed by successful GWAS for osteoarthritis (Mototani et al., 2005), rheumatoid arthritis (Suzuki et al., 2003), and diabetic nephropathy (Tanaka et al., 2003).

Considering family-based linkage studies had failed to identify these risk loci in MS other than HLA, this led to the rationale for and drive to perform high-powered GWAS to screen SNPs in several thousand samples. The GWAS strategy in MS, like other complex disorders, relied on the common disease-common variant hypothesis (CDCV) (Wang et al., 2005). The CDCV hypothesis argues that common genetic variants with high frequency in the population, but relatively low penetrance, are the major contributors to genetic susceptibility to common diseases (McCarthy et al., 2008). Thus, by screening for difference in allele frequency of relatively common SNPs between cases and healthy controls, one can gain insight into the genetics of MS (Figure 1.2).
In 2007 the first MS GWAS was published and identified the first non-HLA regions with genome-wide significance ($p<1\times10^{-8}$) (Hafler et al., 2007). This study utilized 2322 MS cases and 789 control subjects and found significant associations in interleukin-2 receptor subunit alpha (IL-2Rα) and interleukin-7 receptor subunit alpha (IL-7Rα) (Lundmark et al., 2007; Cerosaletti et al., 2013). These were later supported by functional data where it was shown that polymorphism in IL-2Rα regulates GM-CSF production in Th cells (Hartmann et al., 2014). Similarly, polymorphic IL-7Rα was found to alter the pathogenicity of CD8+ T cells (Kreft et al., 2012).

From 2007 onwards, thirty-three MS-GWAS studies have been performed (Buniello et al., 2019) to uncover risk alleles (Table 1.2). These studies were remarkably successful in identifying significant associations in 200 autosomal variants outside the MHC, one on the Chr-X and 32 independent effects in the broader MHC locus (Alcina et al., 2019). Remarkably, the majority of these MS-associated loci encode immune related genes, strongly supporting the notion that MS is primarily an immune-mediated disease. Of interest, one third of the identified loci are associated with at least one other autoimmune disease, suggesting common disease mechanism(s) may underlie most, if not all autoimmune diseases (Cotsapas et al., 2011).
After more than 12 years of GWAS studies in MS and identification of 233+ genetic variants, several caveats remain to be explained. The total heritability explained is a mere 30% of total genetic variance (Beecham et al., 2013). The ‘missing heritability’ is thought to result from common variants with much smaller effect sizes that are below the detection limit in current GWAS studies. A recent and modified association study of MS patients discovered several genetic associations in rare variants (MAF <1%) that could explain 5% of the heritability in MS (Mitrović et al., 2018). The current design of GWAS assumes a single-gene model in which each locus acts independently of the others (Spencer et al., 2009). It is quite possible that a substantial portion of heritability lies in epitasis or genetic interactions among variants. This is true for MS where three types of epistasis have been documented in the MHC locus that can exert an effect on MS risk (Dyment et al., 2005).
The distinction between GWAS-associated variant and causal variant may be masked by linkage disequilibrium (LD) in the genome and may require additional fine mapping to exclude false positive associations. Unfortunately, the majority of MS-GWAS studies are performed in Caucasians where the pattern of LD structure as well as frequency of risk allele vary across diverse population groups (African Americans, Ashkenazi Jews) (Martin et al., 2018). Thus, the applicability of GWAS findings across diverse populations is an overestimation. In fact, it has been demonstrated that variants associated with diseases found in European ancestry populations do not always replicate in non-European populations (Simon-Sanchez et al., 2008; Weiss et al., 2009; Yamada et al., 2009; Haga, 2010).

An overwhelmingly large number of identified variants in MS GWAS—greater than 90%—are located outside of coding sequences (promoter, enhancer, repressor, intron, splice site, intergenic, non-coding RNA etc), making direct interpretation of their functional effects challenging (Schaub et al., 2012; Paraboschi et al., 2018). These can perturb transcription factor binding sites or chromatin structure in cis or trans and can have direct impact on gene expression (Gao et al., 2018). Understanding how non-coding SNPs alter MS susceptibility remains a significant challenge due to incomplete understanding of cis and trans effects of variants.

The genetic basis of heterogeneity in distinct clinical courses of MS is incomplete with few GWAS studies that predict suggestive associations with age of onset, clinical severity, brain atrophy, glutamate levels and MRI T2 lesion load with disease susceptibility
<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
<th>Population</th>
<th>XY data</th>
<th>SNP</th>
<th>Year</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,862</td>
<td>5,418</td>
<td>UK, US</td>
<td>yes</td>
<td>334,923</td>
<td>2007</td>
<td>(Hafler et al., 2007)</td>
</tr>
<tr>
<td>975</td>
<td>1,466</td>
<td>UK</td>
<td>no</td>
<td>12,374</td>
<td>2007</td>
<td>(Burton et al., 2007)</td>
</tr>
<tr>
<td>2,679</td>
<td>3,125</td>
<td>Dutch isolate</td>
<td>No</td>
<td>262,000</td>
<td>2008</td>
<td>(Aulchenko et al., 2008)</td>
</tr>
<tr>
<td>795</td>
<td>1,275</td>
<td>Spain, US</td>
<td>no</td>
<td>500,000</td>
<td>2008</td>
<td>(Comabella et al., 2008)</td>
</tr>
<tr>
<td>1,364</td>
<td>1,507</td>
<td>Dutch, Canadian</td>
<td>No</td>
<td>465,534</td>
<td>2008</td>
<td>(Hoppenbrouwers et al., 2008)</td>
</tr>
<tr>
<td>978</td>
<td>883</td>
<td>Netherlands, Switzerland, US</td>
<td>yes</td>
<td>551,642</td>
<td>2009</td>
<td>(Baranzini et al., 2009)</td>
</tr>
<tr>
<td>4,839</td>
<td>9,340</td>
<td>US, UK, Switzerland, Netherlands</td>
<td>No</td>
<td>709,690</td>
<td>2009</td>
<td>(De Jager et al., 2009)</td>
</tr>
<tr>
<td>3,874</td>
<td>5,723</td>
<td>Australia, New Zealand, UK, US</td>
<td>Yes</td>
<td>303,431</td>
<td>2009</td>
<td>(Booth; et al., 2009)</td>
</tr>
<tr>
<td>68</td>
<td>136</td>
<td>Finland</td>
<td>Yes</td>
<td>297,343</td>
<td>2010</td>
<td>(Jakkula et al., 2010)</td>
</tr>
<tr>
<td>1,913</td>
<td>-</td>
<td>France, Sweden</td>
<td>No</td>
<td>105,035</td>
<td>2010</td>
<td>(Brynedal et al., 2010a)</td>
</tr>
<tr>
<td>2,657</td>
<td>2,877</td>
<td>Sardinian (Italy)</td>
<td>Yes</td>
<td>6,607,266</td>
<td>2010</td>
<td>(Sanna et al., 2010)</td>
</tr>
<tr>
<td>592</td>
<td>825</td>
<td>Germany</td>
<td>yes</td>
<td>758,000</td>
<td>2010</td>
<td>(Nischwitz et al., 2010)</td>
</tr>
<tr>
<td>382</td>
<td>-</td>
<td>US</td>
<td>Yes</td>
<td>500,000</td>
<td>2010</td>
<td>(Baranzini et al., 2010b)</td>
</tr>
<tr>
<td>3,742</td>
<td>8,708</td>
<td>US, UK</td>
<td>No</td>
<td>2,560,000</td>
<td>2011</td>
<td>(Wang et al., 2011)</td>
</tr>
<tr>
<td>1,470</td>
<td>-</td>
<td>US, UK</td>
<td></td>
<td>2,110,417</td>
<td>2011</td>
<td>(Briggs et al., 2011)</td>
</tr>
<tr>
<td>9,772</td>
<td>17,376</td>
<td>15 countries</td>
<td>No</td>
<td>465,434</td>
<td>2011</td>
<td>(Sawcer et al., 2011)</td>
</tr>
<tr>
<td>5,545</td>
<td>12,153</td>
<td>US, UK, Australia, New Zealand, Netherlands, Switzerland</td>
<td>No</td>
<td>2,529,394</td>
<td>2011</td>
<td>(Patsopoulos et al., 2011)</td>
</tr>
<tr>
<td>576</td>
<td>632</td>
<td>Italy, Sweden, Australia,</td>
<td>No</td>
<td>277,866</td>
<td>2012</td>
<td>(Martinelli-Boneschi et al., 2012)</td>
</tr>
<tr>
<td>Country(s)</td>
<td>Sample Size</td>
<td>Year</td>
<td>Study Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway, Denmark US, UK, Netherlands, Switzerland, Spain, Italy European</td>
<td>4,912 7,498</td>
<td>2012</td>
<td>(Matesanz et al., 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden, Denmark, Norway European</td>
<td>284</td>
<td>2013</td>
<td>(Gourraud et al., 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy, Sweden, Belgium, Denmark, Norway 14 countries</td>
<td>4,396 534</td>
<td>2013</td>
<td>(Mero et al., 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 countries</td>
<td>2,197 249</td>
<td>2013</td>
<td>(Leone et al., 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 countries</td>
<td>29,300 50,794</td>
<td>2013</td>
<td>(Beecham et al., 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia, New Zealand, US, UK 14 countries</td>
<td>4,088 12,030</td>
<td>2014</td>
<td>(Goris et al., 2014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 countries</td>
<td>5,258 727</td>
<td>2015</td>
<td>(Goris et al., 2015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US, UK, France</td>
<td>420 296</td>
<td>2015</td>
<td>(Esposito et al., 2015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany, Italy</td>
<td>12,950 13,718</td>
<td>2016</td>
<td>(Andlauer et al., 2016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US, Australia</td>
<td>3,599</td>
<td>2016</td>
<td>(Zhou et al., 2016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>166 171</td>
<td>2017</td>
<td>(Clarelli et al., 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK, Sardinian (Italy)</td>
<td>9,113 8,192</td>
<td>2017</td>
<td>(Steri et al., 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>449</td>
<td>2017</td>
<td>(Zhou et al., 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 countries</td>
<td>14,802 26,703</td>
<td>2017</td>
<td>(Patsopoulos et al., 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Baranzini et al., 2009; Baranzini et al., 2010a; Brynedal et al., 2010b; Martinelli-Boneschi et al., 2012). A recent transcriptomics analysis of peripheral blood mononuclear cells from CIS, RRMS, PPMS and SPMS patients found unique transcriptional signatures in each clinical course that highlight selective dysregulation of MS susceptibility genes in distinct phases of disease (Srinivasan et al., 2017).
The high prevalence of MS in women, clinical heterogeneity of disease course between men and women, and distinct effects of sex-hormones all warrant an explanation. However, only a few GWAS studies in MS have predicted statistically significant associations with gender (Baranzini et al., 2009; De Jager et al., 2009) and sex chromosomes (Baranzini et al., 2009; Patsopoulo et al., 2017). A SNP with different disease associations in men and women can provide insight into the established sex difference in MS that may further guide the discovery of how disease mechanisms differ between sexes.

Lastly, GWAS studies are merely association studies and we may not be learning anything new about MS genetics unless we identify the causal genetic variant and its biology in pathogenesis of disease. Except for a few mechanistic studies conducted with MS associated risk loci (Maier et al., 2009; Gregory et al., 2012; Didonna et al., 2015), the biological function of most of the predicted candidates is unknown. Elucidating the functional pathways may require studying the effect on candidate gene expression at a tissue or cell level that may or may not be feasible in human studies. Human studies also present a challenge in functional follow up because the contribution of each GWAS gene (or non-coding element) in MS is small (often <1%), thus, both the importance of their roles in pathogenesis and the sex- and non-sex specific mechanisms by which different alleles act are difficult to test in human cohorts. Consequently, studies in experimental autoimmune encephalomyelitis (EAE), the principal autoimmune model of MS (Constantinescu et al., 2011), are useful for discerning the identities and sex-specific effects of predicted MS-GWAS variants in CNS autoimmune disease.
1.3. Experimental autoimmune encephalomyelitis (EAE)

Several animal models exist that can mimic clinical and pathological features of MS including immune mediated (EAE), virus-induced (Theiler’s murine encephalomyelitis virus), and toxin-induced (cuprizone) (Didonna, 2016). EAE is by far the most studied model to understand various aspects of autoimmunity in MS. Like MS, symptoms of EAE disease includes sensory loss, optic neuritis, difficulties with coordination and balance (ataxia), muscle weakness, spasms and progressive hind-limb paralysis (Burrows et al., 2019b). Several species including rodents, primates, cats, dogs and chickens exhibit EAE following immunizations with spinal cord homogenates or purified myelin peptides (active EAE) (Stromnes et al., 2006a). Alternately, EAE can be induced by adoptive transfer of encephalitogenic T cells (passive EAE) (Stromnes et al., 2006b). The histopathology and symptoms of EAE disease vary depending on the genetic background, antigenic material used and mode of immunization making parallels with the heterogeneity of MS disease (Simmons et al., 2013).

Active EAE

Active EAE is induced by subcutaneous injection of either spinal cord homogenate or purified myelin peptides such as myelin oligodendrocyte glycoprotein (MOG), myelin basic protein (MBP) or myelin proteolipid protein (PLP) together with an appropriate adjuvant (Stromnes et al., 2006a). These include complete Freund’s adjuvant (CFA) supplemented with Mycobacterium tuberculosis H37RA, and Bordetella pertussis toxin (PTX). CFA is known to promote proinflammatory Th1 responses and together with PTX, helps increase the permeability of the BBB in the CNS (Reiber et al., 1984). In mice,
which remain the most commonly used animal species in EAE studies, the disease follows a predictable clinical course with a 10-15-day induction period followed by an effector phase targeting mostly the spinal cord and characterized by ascending paralysis in the tail, hind limbs and forelimbs (classic EAE). In C57BL/6J (B6) strain, the disease exhibits a monophasic or chronic form of EAE while SJL/J (SJL) exhibit a relapsing remitting course (Li et al., 2011). NOD/Lt initially exhibits relapsing remitting course that transitions to a chronic form of EAE (Baker et al., 2019).

Daily clinical assessment of disease from day 12-30 are performed using a non-invasive grading system scaled from 0 to 5 (0, no clinical expression of disease; 1, flaccid tail without hind limb weakness; 2, hind limb weakness; 3, complete hind limb paralysis and floppy tail; 4, hind leg paralysis accompanied by a floppy tail and urinary or fecal incontinence; and 5, moribund state (Miller et al., 2007). In a small number of antigen-specific models, brain inflammation occurs and are characterized by proprioception defects, ataxia, spasticity and hyper-reflexivity (atypical EAE) (Stromnes et al., 2008).

**Passive EAE**

Instead of immunization with myelin-specific antigens, EAE can be induce passively by inoculating naïve syngeneic mice with activated myelin-specific T cells (Ben-Nun et al., 1981). Passive EAE allows one to evaluate the effector phase of disease in the absence of adjuvant. Symptoms typically appear earlier, are more severe and homogenous in passive EAE. The clinical features and assessment scores utilized are identical to those for active EAE. Passive EAE models offer the advantage that the myelin-specific T cells
can also be manipulated \textit{in vitro} to study the role of specific cytokines and other biological agents before transfer to the recipient.

Considering EAE requires prior immunization with antigen/adjuvant or injection of autoreactive CD4+ T cells while MS seem to develop naturally, spontaneous models of EAE were developed in which genetically susceptible strains expressed transgenic TCR for myelin peptides (Bettelli \textit{et al.}, 2003). These animals are born normal and develop a spontaneous EAE disease several months after birth. The spontaneous model allows one to study EAE disease without confounding factors such as immunizing agents and/or route of immunization.

**Immunology of EAE**

Autoreactive T cells are normal constituents of the peripheral T cell repertoire and are activated following immunization with myelin antigens (Anderson \textit{et al.}, 2000). These cells undergo maturation and clonal expansion in peripheral lymphoid organs (draining lymph nodes and spleen) under the guidance of APCs including DC and macrophages. Following differentiation and aided by their activation status and CFA/PTX pre-treatment, these effector T cells enter the circulation and make their way through the BBB into the CNS facilitated by expression of specific adhesion molecules, cytokines, chemokines and their receptors. Within the CNS, effector T cells become reactivated after they recognize myelin antigens on APC, such as microglia, macrophages and/or astrocytes. Reactivation leads to the expression of pro-inflammatory cytokines by effector T cells. This constant activation and recruitment (facilitated by chemokine production from effector T cells) bring in additional effector cells such as γδ T cells, CD8 T cells, monocytes, macrophages,
and neutrophils that lead to a vicious cycle of myelin destruction (Lees et al., 2008). Initially, the immune response is contained by recruitment of regulatory T cells that leads to resolution of some EAE symptoms (Korn et al., 2007). However, in later stages, the constant activation and recruitment overpowers the regulatory responses with the immune response spreading to other neuro-antigens presented by resident APCs (epitope spreading), thus failing to control autoimmune inflammation (McMahon et al., 2005).

*T lymphocytes.* Several flavors of T lymphocytes are important in EAE including CD4+ T cells (Th), cytotoxic (CD8+) cells, and regulatory T (Treg) cells. Each category is subdivided into unique subsets depending on cytokine environment, transcriptional program and epigenetic modifications. Th1 cells secreting TNF-α, IFN-γ and IL-2 were originally found to have a pathogenic role in the CNS (Sriram et al., 1982). These cells require IL-12 for their polarization from naïve T cells (Th0). Mice deficient in T-bet, key transcription factor for the development of Th1 cells, do not develop EAE (Bettelli et al., 2004; Park et al., 2014). Both TNF-α mice and administration of TNF antagonist reduce EAE disease (Körner et al., 1995; Riminton et al., 1998). Interestingly, there is evidence that suggest an anti-inflammatory role of this cytokine (Frei et al., 1997). This could explain the failed attempts in targeting TNF for the treatment of MS (Group, 1999). Th1 cells expressing IFN-γ are known to infiltrate CNS of EAE mice (Murphy et al., 2010). This is supported by clinical evidence showing that MS patients treated with recombinant IFN-γ in clinical trials develop severe inflammation (Panitch et al., 1987a). However, genetic knockouts of IFN-γ gene, its receptor and anti- IFN-γ treatments exacerbate EAE (Billiau et al., 1988; Ferber et al., 1996) suggesting a regulatory role.
Th2 cells are believed to exert a protective anti-inflammatory role in EAE. They secrete numerous cytokines including IL-4, IL-5, IL-10, and IL13 (Mosmann et al., 1996). IL-4−/− mice exhibit increased EAE severity, while overexpression of transcription factor for Th2, GATA-3, ameliorates disease (Fernando et al., 2014). Th17 are a proinflammatory population secreting IL-17 and are found in CNS infiltrates of EAE mice (Hofstetter et al., 2005). Mice lacking IL-17, its receptor, RORγt (transcription factor for the development of Th17) or IL-23 (cytokine needed for differentiation of Th17), have been reported to show an attenuation of CNS inflammation and suppression of EAE (Bettelli et al., 2004; Park et al., 2014) (Gonzalez-García et al., 2009). Autoreactive Th17 cells can induce EAE following adoptive transfer to naive mice, and this produces a more clinically severe form of the disease than Th1-mediated EAE (Jäger et al., 2009). Contrary to this, mice lacking IL-17A and IL-17F do not show any major alleviation of clinical disease (Haak et al., 2009) suggesting that these signature cytokines themselves are not required. It was later shown that GM-CSF secreted by Th17 cells is the main cytokine contributing to encephalitogenicity (El-Behi et al., 2011).

CD8+ T cells are divided into Tc1 (produce IFN-γ), Tc2 (produce IL-4, IL-5 and IL-10) and Tc17 (produce IL-17) (Gravano et al., 2013). Like CD4+, CD8+ T cells can migrate into the CNS and adoptive transfer of CD8+ T cells sensitized to MOG produces histologically more severe EAE (Sun et al., 2001). Complete depletion or genetic ablation of CD8+ T cells in Lewis rats or C57BL/6J mice protects them from EAE (Camara et al., 2013; Luo et al., 2014). While these studies suggest a pathogenic role for CD8+ T-cells in
EAE, there is conflicting evidence supporting a regulatory role (Najafian et al., 2003; York et al., 2010) suggesting involvement of specific effector subsets of CD8+ T cells.

Tregs expressing CD4+ CD25+ Foxp3+ and producing IL-10, IL-35 and TGF-β have a well-characterized role in promoting peripheral tolerance by regulating inflammatory responses. Absent or dysfunctional Tregs due to genetic defects in Foxp3 in mice leads to a lymphoproliferative disorder (Brunkow et al., 2001). With regards to CNS autoimmune disease, the frequency of Tregs are increased in the CNS during recovery from EAE which are thought to suppress the production of IFN-γ by MOG-sensitized T cells in coculture (Korn et al., 2007). Transfer of CD25+ Treg, or administration of indoleamine 2,3-dioxygenase (IDO) metabolite (increases Treg number) leads to significant amelioration of EAE symptoms (Kohm et al., 2002; Yan et al., 2010). In addition, non-specific deletion of natural Treg by anti-CD25 antibodies has been reported to exacerbate EAE. Possible mechanisms include soluble mediators like IL-10 (prevents co-stimulation of T cells), IL-35 (upregulates inhibitory molecules such as PD-1, TIM3 and LAP3 on T cells) and TGF-β (regulates antibody production; promotes Treg differentiation), cell-to-cell contact with autoreactive effector T cells (inhibitory molecules: CTLA-4, TIM3, LAP3, PD-1) or inhibiting APCs such as DCs (Danikowski et al., 2017) (Zhang et al., 2004). CD25+ CD4+ Treg from IL-10-deficient mice are unable to suppress active EAE (Zhang et al., 2004). Dysfunctional Tregs are also reported in EAE (Viglietta et al., 2004).

B lymphocytes. B cells are important regulators of immune system by secretion of antibodies, cytokines and antigen presentation. In EAE, B cells contribute to demyelination and CNS damage through the production of anti-myelin antibodies and administration of
anti-IgM ameliorates EAE (Willenborg et al., 1983). The role of antibodies secreted by B cells in disease pathogenesis is supported by studies where adoptive transfer of anti-MOG antibodies resulted in demyelination in Lewis rat and SJL mice (Linnington et al., 1987; Schluesener et al., 1987). In addition to antibodies, B cells primed by Th1 cells can secrete proinflammatory cytokines including IFN-γ, IL-12, and TNFα, whilst B cells primed by Th2 cells secrete anti-inflammatory cytokines, such as IL-4 and IL-13 all of which are important in EAE disease (Harris et al., 2000). Pathogenicity of B cells in EAE is linked to the secretion of IL-6 (Barr et al., 2012). In MOG protein-induced EAE, administration of anti-CD20–mediated B-cell depletion ameliorated EAE (Weber et al., 2010). In contrast, anti-CD-20 treatment in EAE induced by MOG35-55 peptide exacerbates disease (Matsushita et al., 2008). In support of a regulatory role, B-cell deficient mice (μMT) fail to recover from EAE disease compared with wild-type (WT) counter-parts (Wolf et al., 1996).

**Innate cells.** While the adaptive immune system is critical for the induction of EAE, innate immune cells comprising of macrophages, microglia, DCs and NK cells are involved in several stages of EAE. During EAE, activated macrophages are found in CNS lesions and correlate with progression to paralytic EAE (Ajami et al., 2011). They express several markers including calcium-binding adapter protein (IBA-1), major histocompatibility complex class II (MHC II), CXCR1, and CD11b (London et al., 2013). Subtypes of macrophage include the predominantly proinflammatory M1 cell (iNOS+) which secretes cytokines including IL-6, IL-12, TNF-α, IL-1, and IL-23 and IL-1β and their presence is associated with increased EAE severity (Liu et al., 2013). M2 cell
(Arg1+), on the other hand, are anti-inflammatory in nature secreting IL-10 and adoptive transfer of IL-4-activated M2 cells ameliorates clinical disease ameliorates EAE (Mikita et al., 2011).

Microglia bears several similarities with macrophages including shared surface markers (F4/80, CD11b) and activation states (M1/M2), and are scattered throughout the CNS, coming into close contact with important cells in EAE including neurons, astrocytes, and oligodendrocytes. Activation of microglia is a hallmark of EAE pathology and in vivo depletion of CD11b+ cells (microglia and macrophages) significantly repress EAE disease (Heppner et al., 2005).

DCs are professional APC residing in the peripheral and lymphoid organs. In the absence of inflammation, they exhibit low surface expression of major histocompatibility complex (MHC) and costimulatory molecules and are characterized by high endocytic capacity. Activation of DCs by microbes, proinflammatory cytokines, or CD40 ligand make them proficient at activating naïve T cells (Steinman et al., 2003). Distinct DC subsets in mouse include myeloid and plasmacytoid DC. Myeloid DC (mDC) express CD11c, CD11b, CD103, and ESAM surface marker while plasmacytoid DC (pDC) lack these antigens and instead express B220 and Siglec. Genetically engineered mice expressing MHC-II on CD11c+ DC cells are susceptible to EAE, suggesting that DCs alone are sufficient to mediate CNS inflammation and clinical disease development (Greter et al., 2005a). Depletion of CD11c+ DCs or exclusively pDC has shown to ameliorate EAE (Isaksson et al., 2009; Isaksson et al., 2012; Yogev et al., 2012).
\(\gamma\delta\) T cells are a small heterogenous population of unconventional innate like T cells and are defined by expression of TCR composed of \(\gamma\) and \(\delta\) chains. Activation of \(\gamma\delta\) does not require antigen processing and MHC and thus make them early effector cells. They have been identified clonally expanded in CNS infiltrates of EAE mice and are the first cells to respond to IL-23 upon EAE induction (Blink et al., 2014b). Early \(\gamma\delta17\) T cell-derived cytokine secretion (GM-CSF, IL-17, IL-21, and IL-22) has shown to enhance the pro-inflammatory activity of \(\alpha\beta\) Th17 cells (Sutton et al., 2009). Mice lacking \(\gamma\delta\) T cells exhibit less severe EAE when compared to wild-type mice (Blink et al., 2014b) while some studies suggesting a protective role (Ponomarev et al., 2004). These contradictory findings were resolved by the discovery of unique subsets of \(\gamma\delta\) and it was shown that anti-\(V\gamma4\) treatment exacerbates EAE disease whereas anti-\(V\gamma1\) treatment is protective (Blink et al., 2014b). The \(V\gamma4^+\) subset produces high levels of IL-17, and accounts for 15-20% of the IL-17 producing cells in the CNS during EAE. In contrast, the \(V\gamma1\) subset produces CCR5 ligands, which may promote regulatory T cell differentiation. The role of other \(\gamma\delta\) subsets (\(V\gamma5, V\gamma6, V\gamma7\)) in EAE remain to be determined.

NK cells are large granular lymphocytes that respond rapidly to a variety of insults with cytolytic activity and cytokine secretion and are protective in EAE (Hao et al., 2010). CNS-infiltrating NK cells secrete brain-derived neuroprotective factor (BDNF) (Hammarberg et al., 2000) that may reduce CNS damage. In vitro experiments indicate that the downregulation of EAE by NK cells may arise from their inhibitory effects on T cell proliferation (Zhang et al., 1997).
Overall, EAE pathology is mediated by the concerted actions of several innate and adaptive immune cells with marked heterogeneity and distinct sub-populations attributed to their functions. The discovery of unique markers distinguishing respective subsets will continue to improve our understanding EAE disease.

**Genetics and sex differences in EAE**

*Genetics.* EAE bears similarities to MS especially with the genetic control of susceptibility. Different strains of mice and rats vary in their susceptibility to EAE. Using mice congenic for the major histocompatibility complex (MHC), termed H-2 in the mouse, it is demonstrated that EAE disease was restricted by the MHC class II, notably H-2A, gene products, however there are exceptions such as B10.S (H-2s) and B 10.Q (H-2q) (Abdul-Majid *et al.*, 2000) Andersson *et al.*, 2004. Whilst MHC expression is critical for the generation of autoimmunity, there is variability in the background genes of the susceptible MHC haplotypes and this may mask the importance of this restriction (Levine *et al.*, 1973). In addition to MHC, several non-MHC loci are implicated with EAE incidence, onset, severity, and histopathology (Butterfield *et al.*, 1998; Butterfield *et al.*, 2000). QTLs controlling sub-phenotypes of EAE, such as histopathological lesion-severity in the brain or spinal cord, weight loss, *Bordetella pertussis* histamine sensitivity (Bphs) histamine sensitivity (Histh), demyelination and paralysis have been described (Encinas *et al.*, 1996; Butterfield *et al.*, 1998; Butterfield *et al.*, 1999; Teuscher *et al.*, 1999; Blankenhorn *et al.*, 2000b; Encinas *et al.*, 2001; Ma *et al.*, 2002; Mazon Pelaez *et al.*, 2005; Tyler *et al.*, 2019). These studies have indicated that EAE disease is a complex polygenic trait of dominant, additive (heterozygotes) and recessive alleles whose products may interact in an epistatic
fashion, where the detection of susceptibility loci depends on the presence of another interacting locus (Baker, 2005).

Several of these quantitative trait loci (QTL) exhibit sex-specificity mirroring gender differences in MS (Table 1.3) (Bearoff et al., 2015). The presence of unique QTL in males and females regulating the severity and characteristics of CNS lesions in EAE suggests that these QTL are responsive to sex hormones. Similarly, EAE susceptibility is influenced by the Y-chromosome and parent-of-origin effect (Teuscher et al., 2006; Spach et al., 2009a; Case et al., 2015). In addition, EAE model is prone to gene-environmental interactions since use of PTX can override EAE resistance of some mouse strains (Munoz et al., 1984; Blankenhorn et al., 2000a).

### Table 1.3. EAE loci in mouse models

<table>
<thead>
<tr>
<th>EAE locus</th>
<th>Location</th>
<th>sex specificity</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eae1</td>
<td>17</td>
<td>20</td>
<td>Incidence</td>
</tr>
<tr>
<td>Eae2</td>
<td>15</td>
<td>15</td>
<td>Incidence</td>
</tr>
<tr>
<td>Eae3</td>
<td>3</td>
<td>42</td>
<td>Incidence</td>
</tr>
<tr>
<td>Eae4</td>
<td>7</td>
<td>50</td>
<td>Incidence, spinal cord histopathology</td>
</tr>
<tr>
<td>Eae5</td>
<td>17</td>
<td>21</td>
<td>Incidence</td>
</tr>
<tr>
<td>Eae6a</td>
<td>11</td>
<td>7</td>
<td>Severity</td>
</tr>
<tr>
<td>Eae6b</td>
<td>11</td>
<td>24</td>
<td>Duration</td>
</tr>
<tr>
<td>Eae7</td>
<td>11</td>
<td>48</td>
<td>Severity, acute disease</td>
</tr>
<tr>
<td>Eae8</td>
<td>2</td>
<td>103</td>
<td>Incidence, severity</td>
</tr>
<tr>
<td>Eae9</td>
<td>9</td>
<td>35</td>
<td>Duration</td>
</tr>
<tr>
<td>Eae10</td>
<td>3</td>
<td>72</td>
<td>Onset</td>
</tr>
<tr>
<td>Eae11</td>
<td>16</td>
<td>41</td>
<td>Male Incidence, brain histopathology</td>
</tr>
<tr>
<td>Eae12</td>
<td>7</td>
<td>12</td>
<td>Female Remitting relapsing disease</td>
</tr>
<tr>
<td>Eae13</td>
<td>13</td>
<td>37</td>
<td>Male Monophasic disease</td>
</tr>
<tr>
<td>Eae14</td>
<td>8</td>
<td>21</td>
<td>Female Incidence</td>
</tr>
<tr>
<td>Eae15</td>
<td>10</td>
<td>16</td>
<td>Male Brain histopathology</td>
</tr>
<tr>
<td>Eae16</td>
<td>12</td>
<td>6</td>
<td>Spinal cord histopathology</td>
</tr>
<tr>
<td>Eae17</td>
<td>10</td>
<td>36</td>
<td>Female Severity, spinal cord demyelination</td>
</tr>
</tbody>
</table>
Sex differences. Sex differences in EAE were first reported in Lewis rats where female exhibit clinical relapses but males exhibit a monophasic disease course (Keith, 1978). Similarly, female SJL mice exhibit multiple relapses of EAE while males do not (Bebo Jr et al., 1996). Adoptive transfer of encephalitogenic T cells from female SJL exhibit greater incidence and severity of EAE compared with male SJL (Bebo Jr et al., 1998a). In addition to SJL, ASW, B10.PL, PL and NZW strains of mice exhibit significant differences in clinical disease between males and females. The severity of EAE is higher in females of the H-2^s strains (SJL/J and ASW) but lower in females of the H-2^u strains (B10.PL and PL/J) (Papenfuss et al., 2004b).

Administration of testosterone to female SJL mice results in less severe EAE, while castration of male SJL mice induces relapses (Dalal et al., 1997a). However, adoptive transfer of encephalitogenic T cells from female SJL donors into male recipients result in severe EAE suggesting testosterone effects may not be universal (Bebo Jr et al., 1998b). In agreement with this, there is no effect of orchiectomy in male B10.PL mice or testosterone treated C57BL/6 mice on EAE disease (Voskuhl, 2002). Similarly, no sex

<table>
<thead>
<tr>
<th>Eae18</th>
<th>18</th>
<th>-</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eae19</td>
<td>19</td>
<td>34</td>
<td>Male</td>
</tr>
<tr>
<td>Eae20</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Eae21</td>
<td>2</td>
<td>37</td>
<td>Female</td>
</tr>
<tr>
<td>Eae22</td>
<td>11</td>
<td>61</td>
<td>Female</td>
</tr>
<tr>
<td>Eae23</td>
<td>11</td>
<td>38</td>
<td>Male</td>
</tr>
<tr>
<td>Eae24</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Eae25</td>
<td>18</td>
<td>54</td>
<td>Male</td>
</tr>
<tr>
<td>Eae26</td>
<td>5</td>
<td>27</td>
<td>Male</td>
</tr>
<tr>
<td>Eae27</td>
<td>1</td>
<td>82</td>
<td>Female</td>
</tr>
</tbody>
</table>

1Chromosome.
2cM position of genetic marker correlated to linkage peak according to www.jax.org
differences are noted in the C57BL/6 or NOD strains suggesting genetics may play a more dominant role than sex hormones in influencing EAE. Intriguingly, administration of estriol (E2) protects SJL, B10.PL and B10.RIII mice from EAE while ovariectomy results in increased EAE severity (Kim et al., 1999).

1.4. The signaling lymphocytic activation molecule (SLAM) family of receptors

The signaling lymphocytic activation molecule (SLAM) family is comprised of nine distinct receptors SLAMF1 (SLAM or CD-150), SLAMF2 (CD48), SLAMF3 (Ly-9 or CD229), SLAMF4 (2B4 or CD244), SLAMF5 (CD84), SLAMF6 (Ly108 in mice, NTB-A or SF2000 in humans), SLAMF7 (CRACC, CD319 or CS1), SLAMF8 (CD353 or BLAME), and SLAMF9 (SF2001 or CD84H) and are expressed on hematopoietic cells (Wu et al., 2016). These receptors are part of the CD2 family which represents a subset of the immunoglobulin (Ig) superfamily of cell surface receptors. Family members are characterized by a N-terminal variable (V) domain lacking disulfide bonds and a membrane-proximal C2 domain containing two conserved disulfide bonds. An exception is SLAMF3, in which the V-CD2-like sequences have been duplicated, resulting in an extracellular domain containing four Ig-like folds. Furthermore, the genes encoding CD2 family members are clustered on mouse Chr 1 and 3 and on two syntenic pericentric loci on human Chr 1, suggesting that they arose through duplication of a common ancestral precursor.

Most of the SLAM family members function through homophilic binding and are important in regulating several immune responses including CD4+ T cell differentiation, CD8+ T cell proliferation, cytokine production by macrophages and DC, B cell activation,
antibody production, and NK cell functions (Veillette et al., 2003; Wang et al., 2004; Chatterjee et al., 2012; Chu et al., 2014).

Table 1.4. The SLAM family and signaling adaptors

<table>
<thead>
<tr>
<th>Name</th>
<th>Expression</th>
<th>Ligand(s)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLAMF1</td>
<td>Thymocytes, naïve B cells, memory T cells, in-vitro activated T and B cells, mature DC, platelets, hematopoietic stem cells</td>
<td>Self, measles virus</td>
<td>• IL-4 secretion by CD4+ T cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• IL-12, TNFα production by macrophages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Measles virus receptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Platelet aggregation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Promotes myeloid cell migration</td>
</tr>
<tr>
<td>SLAMF2</td>
<td>Monocyte, B and T cells, NK cells, eosinophils, γδ T cells</td>
<td>SLAMF4</td>
<td>• Effector functions of CD8+ T cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Associated with mast cell activation by <em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Promotes B cell activation and inhibits apoptosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Increases TNFα production from mast cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Promotes eosinophil activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Enhances macrophage phagocytic functions</td>
</tr>
<tr>
<td>SLAMF3</td>
<td>Thymocytes, T cells, B cells, NKT cells, NK cells, hematopoietic stem cells</td>
<td>Self</td>
<td>• Negative regulator of T cell receptor signaling and IFN-γ secretion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Splenocytes from Slamf3−/− mice display proliferation defects</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Negatively regulate NKT development</td>
</tr>
<tr>
<td>SLAMF4</td>
<td>NK cells, NKT cells, γδ T cells, memory CD8+ T cells, monocytes, basophils, eosinophils, DC</td>
<td>SLAMF2</td>
<td>• NK cell cytokine secretion, cytotoxicity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Immune synapse formation in CD8+ T cells signal through SLAMF4</td>
</tr>
<tr>
<td>Protein</td>
<td>Functions</td>
<td>Cell Types</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>SLAMF5</td>
<td>Regulates adhesion, chemotaxis and peroxidase production in eosinophils</td>
<td>Thymocytes, hematopoietic stem cells, B cell, T cells, NKT cells, mast cells, monocytes, macrophages, DC, neutrophils, basophils, eosinophils, platelets</td>
<td></td>
</tr>
<tr>
<td>SLAMF6</td>
<td>NK cell cytokine secretion, cytotoxicity</td>
<td>NK cells, T cells, NKT cells, B cell, eosinophils, plasmacytoid DC (pDC)</td>
<td></td>
</tr>
<tr>
<td>SLAMF7</td>
<td>IL-4 secretion by CD4+ T cells</td>
<td>NK cells, CD8+ T cells, CD4+ T cells, B cells, mature DC</td>
<td></td>
</tr>
<tr>
<td>SLAMF8</td>
<td>Increases reactive oxygen species, IL-6 and TNFα production from neutrophils, regulates germinal center (GC) response</td>
<td>CD8+ T cells, neutrophils, macrophages, monocytes and DC</td>
<td></td>
</tr>
<tr>
<td>SLAMF9</td>
<td>NK cell cytokine secretion, cytotoxicity</td>
<td>Macrophages, pDC</td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>NK cell activation</td>
<td>Thymocytes, T cells, NK cells, NKT cells, B cells, eosinophils, platelets</td>
<td></td>
</tr>
<tr>
<td>EAT-2</td>
<td>Migration of DCs</td>
<td>NK cells, DCs, T cells, B cells, macrophages, platelets</td>
<td></td>
</tr>
<tr>
<td>ERT</td>
<td>NK cell downregulation</td>
<td>NK cells</td>
<td></td>
</tr>
</tbody>
</table>

Several small SLAM family adaptor proteins are known to associate with the intracellular domains of SLAM family members and transduce activating or inhibitory signals.
signals. These include SLAM-associated protein (SAP), Ewing's sarcoma-associated transcript 2 (EAT-2) and EAT-2-related transducer (ERT). While SAP (SH2D1A) is located on chromosome X, EAT-2 (SH2D1B) and ERT (SH2D1B2) are on chromosome 1. In addition, several phosphatases can bind to the cytosolic side of the SLAM family receptors and compete with SAP family adaptors. All SLAM adaptor proteins are cytosolic and exhibit differential expression in immune cells allowing for a variety of co-stimulatory or inhibitory responses (Dragovich et al., 2018). Table 1.4 lists SLAM members and adaptors, their expression, ligands, and cellular functions.

Defects in SLAM family members and adaptor proteins have been implicated in immune diseases. For example, SAP undergoes a loss of function mutation in X-linked lymphoproliferative (XLP) disease. These patients have massive lymphoproliferation that fails to clear infected B cells leading to enlarged lymph nodes, liver, spleen, and hepatitis. Without treatment, most people with XLP survive only into childhood. Defects in humoral responses and lack of germinal center formation are also observed in in virally infected or immunized SAP-deficient mice (Booth et al., 2011) (Crotty et al., 2003). Several studies indicate that these mice have defects in Th2 response (IL-4 secretion), which are restored after reconstitution with WT CD4+ T cells, but not WT B cells (Cannons et al., 2006).

Interestingly, compared to severe immunodeficiencies in Sap−/− mice, single ablation of SLAMF receptors (Slamf1−/−, Slamf3−/−, Slamf5−/−, and Slamf6−/− mice) causes a mild phenotype (Wang et al., 2015). This suggests that functional redundancy exists among the SLAMF receptors.
The importance of SLAM/SAP signaling is also apparent in systemic lupus erythematosus (SLE), an autoimmune multi-organ disease typically associated with vasculopathy and autoantibody production that can lead to seriously disabling or even life-threatening complications, such as lupus nephritis and neuropsychiatric disorders (Harvey et al., 1954). Dysregulation of various SLAM family receptors (SLAMF1, SLAMF2, SLAMF4, SLAMF5, SLAMF6 and SLAMF7) are associated with SLE (Chatterjee et al., 2012; Karampetsou et al., 2017). In agreement with human data, loss of function mutations in Sh2d1a and polymorphism in Slamf6 are associated with increased severity of murine SLE (Wandstrat et al., 2004).

1.5. SLAM family receptor 1, 2 and 7 in MS-GWAS studies

SLAMF1, 2, and 7 have been identified as MS-GWAS candidates (Beecham et al., 2013; Patsopoulos et al., 2017; Madireddy et al., 2019). In addition to SLE and MS, the SLAM locus is implicated in other autoimmune diseases including Graves’ disease, Crohn's disease, ankylosing spondylitis, psoriasis, primary sclerosing cholangitis, and ulcerative colitis (Zhao et al., 2013; Ellinghaus et al., 2016; Langefeld et al., 2017). SLAM/Slam locus is highly conserved across several species including chimpanzee, dog, cow, rat, and mouse, which suggests an important immunological function.

Considering animal models are critical in validation of GWAS candidates, we searched the literature to find ways to test the effect of Slamf1, Slamf2 and Slamf7 polymorphisms in EAE. Wandstrat et al., reported sequencing the entire Slam locus (Slamf1-Slamf9) among 34 inbred laboratory strains of mice (Wandstrat et al., 2004). Interestingly, two stable, divergent haplotypes (Haplotypes 1 and 2) were observed. The
natural genetic variation allows us to assess the effect of the entire Slam locus provided the relevant locus is captured in a congenic mouse strain. Coincidently, Wandstrat et al., described establishing a congenic line (B6.129c1) that captured the Slam interval of haplotype-2 (129/SvJ) on the haplotype-1 (C57BL/6J) background (Wandstrat et al., 2004). This congenic line exhibits SLE-like disease with antinuclear autoantibodies at 9 months of age exclusively in female animals.

Like case/control studies in MS, we therefore sought to utilize the natural genetic variation in the Slam locus between 129 and B6 strains to test the hypothesis that SLAMF1/Slamf1, SLAMF2/Slamf2, and/or SLAMF7/Slamf7, and not a disease related gene in linkage disequilibrium with the SLAM/Slam locus, are the true GWAS-genes contributing to CNS autoimmune disease susceptibility.
1.8. Literature Review References


Alcina, A., Fedetz, M., & Matesanz, F. (2019). Genetics of Multiple Sclerosis. In J. Martín & F. D. Carmona (Eds.), *Genetics of Rare Autoimmune Diseases* (pp. 183-202). Cham: Springer International Publishing.


and the Clinical Course of Relapsing Experimental Autoimmune Encephalomyelitis. J. Neuroimmunol., 84(2), 122-130.


Spontaneous Autoimmune Encephalomyelitis in Mice. Proc Natl Acad Sci U S A, 114(40), 10719-10724.


48


51


CHAPTER 2: SEX-SPECIFIC REGULATION OF CNS AUTOIMMUNITY BY SIGNALING LYMPHOCYTIC ACTIVATION MOLECULE (SLAM) LOCUS.

Abbas Raza¹, Laure K. Case Ph.D.⁶, Erin E. Stebbins¹, Dimitry N. Krementsov Ph.D.², Qian Fian¹, Jonathan E. Boyson, Ph.D.³ and Cory Teuscher Ph.D.¹,⁴

¹Department of Medicine, ²Department of Biomedical & Health Sciences, ³Department of Surgery, ⁴Department of Pathology and Laboratory Medicine, University of Vermont, Burlington, VT 05405 and ⁶The Jackson Laboratory, Bar Harbor, ME 04609

Address correspondence to: Cory Teuscher, 89 Beaumont Avenue, C-329A Given Building, Larner College of Medicine, University of Vermont, Burlington, VT 05405; c.teuscher@uvm.edu

Abbreviations: Slam/SLAM= Signaling lymphocytic activation molecule;
T=Testosterone; DHT=5α-dihydrotestosterone
2.1. Abstract

Multiple Sclerosis (MS) is the most common neurodegenerative disease affecting young adults and is characterized by myelin loss, varying degrees of axonal damage, and progressive neurological dysfunction. The disease has high prevalence in women (3 ♀:1 ♂), although men display a more aggressive course of disease. Risk factors for MS susceptibility are multifactorial and include genetic and environmental influences. The gene discovery efforts in MS using genome-wide association studies (GWAS) have identified at least 233 candidates. However, the contribution of each GWAS gene (or non-coding element) is small (often <1%), thus both the importance of their roles in pathogenesis and the sex- and non-sex specific mechanisms by which different alleles act are difficult to test. Consequently, studies in experimental autoimmune encephalomyelitis (EAE), the principal autoimmune model of MS, are useful for discerning the identities and sex-specific effects of MS-GWAS in CNS autoimmune disease. The signaling lymphocytic activation molecule family members 1 (SLAMF1), 2 (SLAMF), and 7 (SLAMF7) have been identified as potential MS-GWAS candidates. Since the SLAM locus is highly conserved between humans and mice, with natural polymorphic variation in each Slamf gene segregating between 129/SvJ and C57BL/6J, we utilized congenic mapping to test the hypothesis that SLAMF1, SLAMF2, and/or SLAMF7, and not a disease-related gene in linkage disequilibrium with the SLAM locus, are the true GWAS-genes contributing to CNS autoimmune disease susceptibility. Herein, we show evidence that the SLAM candidate gene variants identified through MS-GWAS are not the causal variants in CNS autoimmune disease. Instead, a locus distal to Slamf1, Slamf2, and Slamf7 defines disease
susceptibility. This locus exhibits male-specific resistance that segregates with the frequency of CD11b+ cells and TCRαβ+ CD4+ Foxp3+ T cells in the draining lymph nodes and TCRαβ+ CD4+ Foxp3+ and TCRαβ+ CD8+ IL-17+ T cells in the CNS during EAE. Orchiectomy enhanced the severity of EAE disease among congenic male mice while supplementation of testosterone (T) or 5α-dihydrotestosterone (DHT) did not ameliorate disease. Using a bioinformatics approach, we identified Slamf8 and 9 along with other differentially expressed genes whose expression is testis-dependent, but not directly regulated by T or DHT, as potential positional candidates supporting the existence of novel gene-by-testis interactions controlling susceptibility to CNS autoimmune disease. Further refinement of this locus is required to identify the causal gene(s) that may be targeted for prevention and/or treatment of MS in men.
2.2. Introduction

Multiple sclerosis (MS), a chronic inflammatory disease of the central nervous system (CNS), is characterized by myelin loss, varying degrees of axonal damage, and progressive neurological dysfunction (Noseworthy et al., 2000). It is the most common disabling neurologic disease of young adults and adolescents, affecting 2.3 million individuals worldwide (Chen et al., 2017; Kim et al., 2019). MS prevalence is ~3 fold higher in women, although the disease course in men typically presents as a more rapid severe progressive disease (Whitacre et al., 1999).

The etiology of MS involves both genetic and environmental factors (Gourraud et al., 2012). The heritability of MS is estimated to be ~30%, largely associated with the inheritance of susceptible HLA haplotypes that are present in up to 70% of MS cases (Schmidt et al., 2007). Polymorphisms in multiple non-MHC genes have also been associated with disease in linkage and genome-wide association studies (GWAS) (Hauser et al., 2013). The latest and most extensive GWAS of MS has identified 233 independent associations with MS susceptibility (Sawcer et al., 2019). The contribution of each GWAS gene (or non-coding element (Kumar et al., 2014)) is small (often <1%), and both the importance of their roles in pathogenesis and the mechanisms by which different alleles act are difficult to test. Thus, animal models of MS like experimental autoimmune encephalomyelitis (EAE) (Constantinescu et al., 2011) are useful for discerning the identities and effects of MS gene candidates and to understand their genetic role in CNS autoimmune disease from an evolutionary perspective (Spach et al., 2009b; Blankenhorn et al., 2011; Bearoff et al., 2015).
GWAS studies have shown that: (1) most SNPs (single nucleotide polymorphisms) associated with MS in large populations of people are located close or inside genes encoding immune related molecules (Farh \textit{et al.}, 2015; Axisa \textit{et al.}, 2016); (2) the promoters, enhancers, intergenic regions, and ncRNAs in which these SNPs lie may control much of the phenotypic variation (an estimated $\sim$90\% of associated variants are non-coding (Farh \textit{et al.}, 2015; Ransohoff \textit{et al.}, 2015)). The inclusion of gene functional information and network-based pathway analysis into a GWAS meta-analysis is the best approach to finding 'true' MS genes (Consortium, 2013).

The signaling lymphocytic activation molecule family members 1 (\textit{SLAMF1}), 2 (\textit{SLAMF}), and 7 (\textit{SLAMF7}) have been identified as potential MS-GWAS candidates ($p = 9 \times 10^{-6}$ for all three genes) (Beecham \textit{et al.}, 2013; Patsopoulos \textit{et al.}, 2017; Madireddy \textit{et al.}, 2019). The \textit{SLAM} locus (\textit{SLAMF1-SLAMF9}) is highly conserved between humans and mice (Limaye \textit{et al.}, 2008). A comparison between two strains of mice, 129/SvJ (129) and C57BL/6J (B6), revealed a high level of genetic variation in each SLAM ortholog (Limaye \textit{et al.}, 2008). Therefore, we utilized B6 mice carrying overlapping congenic intervals [B6.129c1 (c1) and B6.129c2 (c2)] on chromosome 1 (Chr 1) derived from the 129 strain to test the hypothesis that \textit{SLAMF1}, \textit{SLAMF2}, and/or \textit{SLAMF7}, and not a gene(s) in linkage disequilibrium with the \textit{SLAM} locus, are the causal GWAS-gene contributing to CNS autoimmune disease susceptibility (Wandstrat \textit{et al.}, 2004; DeVault \textit{et al.}, 2019). We found that the male c1 congenic mice develop less severe EAE than B6 or c2 congenic mice. In contrast, the female mice exhibit no difference in disease susceptibility between B6 or c1 congenic mice. Splenocytes from immunized c1 male mice produce significantly more
interleukin 10 (IL-10) and a higher frequency of CD11b+, TCRαβ+ CD4+ Foxp3+ T cells in draining lymph nodes (DLN). Furthermore, infiltrating immune cells in CNS exhibit higher proportion of TCRαβ+ CD4+ Foxp3+ and TCRαβ+ CD8+ IL-17+ T cells that segregates with resistance to disease in c1 male mice. The effect of orchiectomy and hormone replacement by testosterone (T) or dihydrotestosterone (DHT) on EAE disease course of male c1 congenic mice was investigated. To physically map causal loci distal to MS-GWAS candidates, we generated a panel of interval specific recombinant congenic lines (ISCL) encompassing the candidate genetic interval on Chr 1. The congenic mapping studies reported herein demonstrate that a ~1.2Mb interval distal to Slamf1, Slamf2 and Slamf7 regulates susceptibility to EAE. Collectively these results demonstrate the need to functionally validate candidate genes identified in MS-GWAS to identify the causal gene(s) that may be targeted in therapeutic and/or diagnostic interventions in MS.
2.3. Results

2.3.1. EAE susceptibility in male mice is controlled by a locus in linkage disequilibrium with Slamf1, Slamf2, and Slamf7

The SLAM locus on chromosome 1 includes nine family members (SLAMF1-SLAMF9) that are highly conserved between human and mice (Table 2.1). Three of the family members (SLAMF1, SLAMF2, and SLAMF7) have been predicted as MS-GWAS candidates (Beecham et al., 2013; Patsopoulos et al., 2017; Madireddy et al., 2019). Interestingly, there is natural genetic variation in the Slam locus among inbred strains of mice where two stable and divergent haplotypes are described (Wandstrat et al., 2004). Two congenic lines have been described (B6.129c1 and B6.129c2) that introgress Slam intervals belonging to haplotype-2 (129) onto haplotype-1 (B6) (Figure 2.1) (Wandstrat et al., 2004; DeVault et al., 2019). The B6.129c1 (c1) line carries a 6.6Mb region of 129 that includes all of Slamf1-Slamf9 whereas B6.129c2 (c2) mice carries a small 1Mb interval of 129 that includes Slamf1-Slamf7.

We studied the susceptibility of c1, c2 and background B6 mice to MOG35-55 + CFA immunization (Figure 2.2). Given the sexual dimorphism in EAE (Bearoff et al., 2015), the data for female and males were analyzed separately. A significant difference in the severity of disease course was not detected among the female mice (interaction $p > 0.05$; strain $p > 0.05$; time $p < 0.0001$). However, the severity of the disease course of male c1 mice was significantly reduced compared to both B6 and c2 mice (interaction, $p < 0.0001$; strain, $p < 0.0001$; time, $p < 0.0001$; B6 = c2 > c1. An analysis of EAE-associated clinical quantitative trait variables (Supplementary Table 2.1) revealed that male c1 exhibited
significantly less cumulative diseases scores (CDS), delayed mean day of onset (DO), increased peak score (PS) and lesser days affected (DA) compared with B6 and c2. These data demonstrate that the c2 line, which harbors polymorphic orthologues of *SLAMF1*, *SLAMF2* and *SLAMF7* did not exhibit reduced differences in severity of disease compared with B6. This suggests that EAE susceptibility in male c1 mice is controlled by a locus in linkage disequilibrium to *Slamf1*, *Slamf2*, and *Slamf7*.

### 2.3.2. Immune profiling of c1 and B6 mice

Although the exact pathogenic mechanisms underlying EAE are not known, several immune cell subsets and cytokines are important during the initiation of immune response in secondary lymphoid organs and subsequent trafficking to CNS during later stages of disease (Barthelmes *et al.*, 2016). Therefore, to elucidate the mechanisms associated with differential EAE susceptibility observed in c1 males compared with others, we first examined the frequency of immune cells in the DLN on day 10 (D0) post MOG<sub>35-55</sub> + CFA immunization. There was no significant difference in the frequency of TCRαβ+, CD19+, TCRαβ+ CD4+, TCRαβ+ CD4+ IFN-γ+, TCRαβ+ CD8+, TCRαβ+ CD8+ IFN-γ+, TCRγδ+ IFN-γ+ cells between c1 and B6 (Table 2.2). Increased frequency of TCRαβ+ CD4+ IL17+ in both c1 male and female mice and TCRγδ+ in female c1 was observed. Importantly, increased frequency of CD11b+ cells and TCRαβ+CD4+Foxp3+ co-segregated with EAE resistance in c1 males.

The reduced severity of EAE in c1 males was probed further during the peak stage of disease on day 22 (D22) post MOG<sub>35-55</sub> + CFA immunization by profiling immune cells infiltrating the CNS. In agreement with DLN results, we found a significant increase in the
frequency of TCRαβ+CD4+Foxp3+ and TCRαβ+CD8+IL-17+ T cells that co-segregated with EAE resistance in c1 males (Figure 2.2).

2.3.3. Increased IL-10 by male c1 T cells co-segregates with resistance to EAE

EAE induction is highly dependent on CD4+ T cells capable of producing several cytokines including IFN-γ, IL-17, GM-CSF, and IL-10 (Ponomarev et al., 2007; Herndl-Brandstetter et al., 2014). DLN cells from c1 mice restimulated ex vivo with MOG35-55 produced significantly more IL-10 compared to B6 mice (Figure 2.3). Measurement of other cytokines including IFN-γ, IL-17, and GM-CSF did not differ between c1 and B6 (data not shown). These results suggest that the differences in EAE susceptibility observed in c1 mice may, in part, be attributed to increased regulatory T-cell responses from DLN in the form of IL-10 cytokine.

2.3.4. Gene-by-testis specific interactions modify EAE disease in c1 male mice

Genetic studies have shown that distinct genetic loci govern development of MOG-induced EAE in C57BL/6J mice and are exhibited in a particular sex (Bearoff et al., 2015). In addition, sex hormones including testosterone (T) and estrogen (E2) are known to regulate autoreactive Th1/Th17 responses in EAE disease (Dalal et al., 1997b; Haghmorad et al., 2016). We, therefore, aimed to determine whether endogenous male sex hormones might be regulating disease through the 129 derived quantitative trait locus (QTL). Male c1 mice were orchiectomized (ORX) or sham-operated at 4 weeks of age and supplemented with testosterone (c1 ORX + T) along with placebo control pellets (c1 ORX + CNT) prior to immunization with MOG35-55 + CFA (see material and methods). Testosterone can exert its effect either directly by stimulation of the androgen receptor.
(AR) or via aromatization in target tissues to estradiol (E2), the ligand for the estrogen receptors (ERs) α and β (Ishikawa et al., 2006). Thus, we included supplementation with non-hydrolysable 5α-dihydrotestosterone (c1 ORX + DHT) as a control for testosterone dependent biological effects. Unmanipulated (unoperated) c1 males served as an additional control.

A significant difference in the severity of disease course was not detected between the sham and unmanipulated mice (interaction p = 0.95; strain p = 0.58; time p >0.9999; c1=c1 Sham) (Figure 2.4A). The severity of the disease course of c1 ORX + CNT mice was significantly increased compared to Sham and c1 while neither of the male androgen (T or DHT) replacements ameliorated disease susceptibility (interaction p < 0.0001; strain p < 0.0001; time p < 0.0001; c1 ORX + CNT= c1 ORX + DHT = c1 ORX + T > c1 Sham or c1). Analysis of EAE-associated clinical quantitative trait variables revealed that the CDS and incidence was significantly different among the various groups (Figure 2.4B) with CDS of c1 ORX + CNT= c1 ORX + DHT = c1 ORX + T > c1 Sham or c1 (p < 0.0001) while incidence of c1 ORX + CNT= c1 ORX + DHT = c1 ORX + T > c1 Sham or c1 (p < 0.004). Serum measurements of T or DHT (Figure 2.4C) confirm the results that EAE resistance of male c1 mice is dependent on testis but not testosterone or 5α-dihydrotestosterone.
2.3.5. Lack of evidence for Slamf1, 2 and 7 as the true GWAS candidates regulating CNS autoimmune disease

To physically map the causal EAE locus in c1 congenic male mice, we generated a panel of interval specific recombinant congenic lines (ISRCL) encompassing the candidate genetic interval on Chr 1 and studied them in a step-wise fashion using MOG55-55 + CFA immunization. The severity of the disease course was significantly different among the congenic lines (interaction, \( p < 0.0001 \); strain, \( p < 0.0001 \); time, \( p < 0.0001 \)) with B6 = c8 = c10 > c1 = c7 (Figure 2.5A) and correlated with significant difference in CDS (Figure 2.5B). This demonstrates that the c7 congenic interval is sufficient to capture the full c1 phenotype. Importantly, these results confirm that the causal variant in the c1 congenic line is not Slamf1, Slamf2 or Slamf7 since c7 line does not include the 129-derived polymorphic interval encoding these genes. The c8 line, carrying the distal portion of c7 interval, did not differ in EAE susceptibility with B6. Therefore, the causal gene (s) influencing CNS autoimmune disease resides within a ~1.2Mb interval (172,282,646-173,408,539) (Figure 2.5C).

2.3.6. Polymorphic candidate genes within the refined locus whose expression is testis-dependent are novel candidates regulating CNS autoimmunity.

All genetic elements (protein coding genes, microRNA, long non-coding RNA, pseudogenes) underlying the ~1.2Mb interval were retrieved using genomic coordinates as input in Mouse Genomics Informatics (MGI) database at Jackson Laboratory (http://www.informatics.jax.org/). This yielded thirty-nine genes and may include the
causal gene (s) regulating EAE susceptibility (Figure 2.6). Given that the locus is polymorphic between B6 and 129 strains of mice, we first filtered the gene list to focus on candidates that exhibit SNPs (both coding and non-coding) between 129 and B6 mouse strains. Secondly, we included gene expression data (Accession # GSE54945, GSE5901, and GSE66873) available in various mouse tissues including CD4+ T cells in spleen, cortical thymic epithelial cells and prostrate tumor comparing castrated (ORX) and sham control mice in public databases such as Gene Expression Omnibus (GEO) of National Center of Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/geo/) to delimit potential candidate regulating EAE disease (see material and methods). The prioritized list includes thirteen genes (Atp1a2, Apcs, Cadm3, Dsp23, Fcer1a, Kcnj9, Kcnj10, Olfr16, Pigm, Slamf8, Slamf9, Tagln2, Vsig8) that are novel candidates not previously associated with any MS-GWAS. Several of them are associated with important immune functions and require mechanistic validation.
2.4. Discussion

Many risk-modifying loci have been identified through GWAS studies in MS (Cotsapas et al., 2018). While the number of loci associated with susceptibility have grown over time due to greater statistical power of studies conducted, approaches to determine causal association with MS disease have lagged. In this study, we attempted to understand the role of a MS-GWAS predicted genetic locus called SLAM/Slam using the natural genetic variation in B6 and 129 strains of mice and an integrated genetics and functional approach to test association of Slamf1, Slamf2 and Slamf7 on Chr 1 with the development of CNS autoimmune disease. The results of our study demonstrate that male c1 and c7 congenic mice develop significantly less severe EAE compared with B6 and c2, c8, and c10 congenic mice. This male-specific resistance segregates with the frequency of CD11b+ cells and TCRαβ+ CD4+ Foxp3+ T cells in the draining lymph nodes and TCRαβ+ CD4+ Foxp3+ and TCRαβ+ CD8+ IL-17+ T cells in the CNS during EAE. We have also shown that orchiectomy of male c1 congenic mice enhanced the severity of EAE disease while supplementation of T or DHT did not ameliorate disease.

Sex differences are apparent in several inbred strains including SJL, NZW, ASW, B10.PL and PL where female exhibits a different disease course than male mice following EAE induction (Papenfuss et al., 2004a). Generally, B6, 129 and NOD do not exhibit sex differences in EAE although there are exceptions (Papenfuss et al., 2004a; Rahn et al., 2014). The gender difference seen in our male c1 congenic mice could either be due to differences in sex hormones, differences in sex chromosomes, or both. Since our congenic
lines harbor the same genetic background of B6 except for the congenic interval on Chr 1, we can exclude the contribution of polymorphic loci on sex chromosome.

Studies looking at the effect of orchiectomy and hormone replacement in EAE disease exhibit variable outcomes depending on the mouse strain. Administration of male androgens including T and its non-hydrolyzable form, DHT, in female SJL mice ameliorates EAE, while castration of male SJLs induces relapse (Dalal et al., 1997a). In contrast, there is no effect of orchiectomy in male B10.PL mice or B6 mice (Palaszynski et al., 2004a). Intriguingly, administration of testosterone or DHT did not restore resistance in male c1 congenic mice in line with the idea that endogenous androgens are not protective in all genetic backgrounds (Voskuhl, 2002). Moreover, several immunomodulatory molecules are synthesized by testis in addition to male androgens including inhibins and activins however it is unknown how these molecules can module the immune response in the periphery and CNS (Aleman-Muench et al., 2012; Setchell, 1974; Uhlen et al., 2015; Ahn et al., 2017).

Immune profiling data demonstrate that the polymorphic locus in c1 modulates the frequency of CD11b+, TCR\(\alpha\)\(\beta\)+ CD4+ Foxp3+ and TCR\(\alpha\)\(\beta\)+ CD8+ IL-17+ T cells that co-segregates with resistance to EAE disease. CD11b is expressed on the surface of many leukocytes important in EAE pathogenesis including monocytes, dendritic cells, neutrophils, natural killer cells, granulocytes and macrophages. Thus, it remains to be determined which CD11b+ cell type is upregulated in male c1 congenic mice. Tregs expressing CD4+ CD25+ Foxp3+ and producing IL-10 cytokine has a well-characterized role in EAE by promoting peripheral tolerance and are known to suppress the production
of IFN-γ and IL-17 by MOG-sensitized T cells in coculture (Korn et al., 2007); (Chaudhry et al., 2011). We found an increased secretion of IL-10 by DLN cells in ex vivo recall assays. These results, together with the reports of other groups, indicate that induction of Tregs and secretion of IL-10 is an important step in the tolerance to EAE disease.

Our results with interval specific recombinant congenic (ISRC) lines established that the causal gene lies distal to MS GWAS candidates Slamf1, Slamf2 and Slamf7 in a ~1.2 Mb interval. This interval encodes several genetic elements including protein coding genes, microRNA, long non-coding RNA and pseudogenes all of which are candidates for this male specific phenotype. Given that the locus is polymorphic between 129 and B6, we excluded genetic elements that exhibit no SNP between the two strains. As a second filter, we excluded genes that do not exhibit change in gene expression following orchiectomy. The resulting prioritized list includes thirteen genes: Atp1a2, Apcs, Cadm3, Dsp23, Fcer1a, Kcnj9, Kcnj10, Olfr16, Pigm, Slamf8, Slamf9, Tagln2, Vsig8. The list excludes several of the lncRNA and pseudogenes that are not annotated in gene expression datasets and could be ‘true’ candidates regulating CNS autoimmunity in male c1 congenic mice. Among the annotated genes, several exhibit immune related functions. Apcs encodes serum amyloid P component that belongs to the pentraxin family of proteins and are endowed with Ca2+-dependent lectin-like binding activity. Importantly, Apcs is associated with CNS autoimmunity (Ji et al., 2012). Transgenic mice expressing Apcs have unexpectedly attenuated EAE due to impaired encephalitogenic responses. However, genetic knockouts of Apcs exhibit severe EAE. In vitro studies demonstrate that Apcs could change the affinity between α4-integrin and T cells thereby influencing the trafficking to CNS.
Slamf9 encodes the ninth member of the Slam family and is expressed by myeloid cells with abundant expression in plasmotoid dendritic cells (pDC). Genetic knockouts of Slamf9 leads to defective differentiation and activation of pDCs with reduced capacity to secrete IL-6, IFN-α and TNF-α cytokines suggesting a proinflammatory role of Slamf9 (Sever et al., 2019a). Slamf9<sup>−/−</sup> pDCs express higher levels of chemokine receptor 5 (CCR5), which could increase their mobilization to the periphery. Importantly, Slamf9<sup>−/−</sup> exhibit amelioration of EAE disease (Sever et al., 2019a). Given the important role of Slamf9 in pDC activation and differentiation, immune profiling of male c1 congenic mice should be performed to assess whether they are defective between c1 and B6. Furthermore, the expression levels of all thirteen prioritized gene candidates need to be validated using quantitative PCR before and after orchiectomy in male c1 or c7 congenic mice.

Given that functional redundancy exists among the different Slamf receptors (Slamf1, Slamf3, Slamf5 and Slamf6), it is quite possible that mouse Slamf9 might carry out similar functions to Slamf1, Slamf2 and Slamf7 in humans. Supporting this hypothesis, the genes coding from SLAM-family receptors are located within a ∼400 kilobase (kb)-cluster on chromosome 1, in humans and mice (Morra et al. 2001). This observation, coupled with the conserved exon-intron structure of SLAM-related genes, implies that the SLAM family was generated by sequential duplication of a single ancestor gene and thus may exhibit overlapping functions. Lastly, since several of the candidate genes in the ∼1.2Mb interval have genetic knockouts available, one can assess their contribution to CNS autoimmune disease using F<sub>1</sub> cross with male c1 or c7 congenic line.
As a follow up to orchiectomy dataset, one can exclude genes whose expression is unaffected by testosterone supplementation and in this regard several datasets are available in GEO database (GSE95692, GSE93726, GSE102143) although the cell type may be irrelevant using this approach. Confirmation of differential expression following orchiectomy and testosterone supplementation among c7 congenic mice would be needed to predict the causal gene(s). In summary, we demonstrate the need to functionally validate candidate genes identified in MS-GWAS. Our results suggest lack of evidence that polymorphic Slamf1, Slamf2 and Slamf7 genes do not regulate EAE disease, but a locus distal to these genes is important in CNS autoimmune disease development in mice. Since the SNPs identified in human SLAM locus are markedly different from the Slam locus in mice, this however does not preclude the possibility that the identified GWAS loci may be causal in MS. This locus is male-specific and supports the existence of novel gene-by-testis interactions controlling susceptibility to disease that may be targeted for prevention and/or treatment of MS in men.
2.5. Material and Methods

2.5.1. Animals

C57BL6/J (B6) were purchased from the Jackson Laboratory (Bar Harbor, Maine). B6.129c1 (c1) and B6.129c2 (c2) congenic mice have been previously described (Wandstrat et al., 2004; DeVault et al., 2019). B6.129c7 (c7), B6.129c8 (c8), B6.129c10 (c10), and B6.129c11 (c11) strains were generated by backcrossing c1 mice to B6 mice and intercrossing the heterozygous progeny. Offspring in which recombinants were identified were backcrossed to B6 to allow for the generation of homozygous congenic lines. All mice were housed in the specific pathogen-free barrier facility at the University of Vermont. All procedures involving animals were approved by the University of Vermont Institutional Animal Care and Use Committee.

2.5.2. DNA extraction and Genotyping

DNA was isolated from mouse tail clippings as previously described (Sudweeks et al., 1993). Briefly, individual tail clippings were incubated with cell lysis buffer (125 mg/ml proteinase K, 100 mM NaCl, 10mM Tris-HCl (pH 8.3), 10 mM EDTA, 100 mM KCl, 0.50% SDS, 300 ml) overnight at 55°C. The next day, 6M NaCl (150 ml) was added followed by centrifugation for 10 min. at 4°C. The supernatant layer was transferred to a fresh tube containing 300 µl isopropanol. After centrifuging for 2 min, the supernatant was discarded, and the pellet washed with 70% ethanol. After a final 2 min. centrifugation, the supernatant was discarded, and DNA was air-dried and resuspended in TE. The congenic interval boundaries were determined using primer sets designed to amplify across informative insertions/deletions and SNPs (Supplementary Table 2.2).
Thermocycling was carried out for a 15 µL reaction mix with 2mM MgCl2, 200µM dNTPs, 0.2µM primers, 1 unit of Taq polymerase and ~50ng of genomic DNA together with an initial 2-min 97°C denaturation followed by 35 cycles of 97°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec. The final extension was for 5 min at 72°C. Amplicons were subjected to 2.5% agarose gel electrophoresis and visualized by ethidium bromide and UV light.

2.5.3. Induction and evaluation of actively induced EAE

Eight-ten week old male and female mice were injected subcutaneously in the posterior right and left flank with a sonicated emulsion containing 100 µg of MOG35-55 and an equal volume of complete Freund's adjuvant (CFA; Sigma-Aldrich, St. Louis, MO, USA) supplemented with 200 µg of Mycobacterium tuberculosis H37Ra (Difco Laboratories, Detroit, MI, USA). Then 1 week later, all mice received an identical injection of MOG35-55 + CFA.

Mice were scored daily for clinical signs of EAE beginning on D10 after injection as follows: 0, no clinical expression of disease; 1, flaccid tail without hind-limb weakness; 2, hind limb weakness; 3, complete hind-limb paralysis and floppy tail; 4, hind-limb paralysis accompanied by a floppy tail and urinary or fecal incontinence; and 5, moribund. Clinical quantitative trait variables, including disease incidence, mean day of onset, cumulative disease score, number of days affected, overall severity index, and peak score, were generated as previously described (Fillmore et al., 2004). The incidence of EAE was recorded as positive for any mouse with clinical signs of EAE for 1 or more days. Susceptibility was analyzed as a quantitative trait, using a disease index generated
by averaging the clinical scores for each animal over the course of the experiment. The severity of disease among affected animals was analyzed using a severity index generated by averaging the clinical scores for each animal over the number of days that it exhibited clinical symptoms. Severity was assessed only in affected animals. Days affected was calculated as the number of days an animal displayed a clinical score of 1, and onset was the day clinical signs were first observed.

2.5.4. Orchiectomy and hormone administration

Male mice at 4 weeks of age were anesthetized using ketamine (80 mg/kg) and xylazine (10mg/kg) as described (Fillmore *et al.*, 2004). The area around testes was shaved using a trimmer and sterilized using surgical scrub, alcohol and 5% povidone-iodine (Betadine®; SmartPak Equine LLC, Plymouth, MA, USA). A single <1cm incision was made in the scrotal sac, the vas deferens was clipped and the testes were removed. The incision was then sealed with gut (dissolving) sutures to close the wound. Buprenorphine hydrochloride was injected at 0.05mg/kg subcutaneously immediately after surgery and again 4 hours later for analgesia. After the surgeries, mice were placed on a paper towel in the cages to prevent inhalation of the bedding and warmed with a heat lamp until they begin moving around the cage. They were allowed rest for 2 weeks then implanted with hormone pellets.

For hormone replacement, mice were anesthetized by inhalation of isoflurane (0.2ml) in a transparent cylindrical induction chamber and a small incision was made dorsally and slightly lateral to the nape approximately 3mm caudal to the ear. Sixty-day release testosterone (12.5mg total; 0.21mg/day), 5α-dihydrotestosterone (5mg total;
0.08mg/day) or control pellets (Innovative Research, Sarasota, FL) were subcutaneously implanted. The incision was sealed, and mice injected with Buprenorphine hydrochloride as before. The animals were rested an additional 2 weeks to allow recovery and then immunized with EAE.

2.5.5. Cytokine and proliferation assays

For ex vivo cytokine assays, mice were immunized using the EAE immunization protocol described earlier, spleens and draining lymph nodes were harvested on D10, and single-cell suspensions were prepared (1 x 10⁶ cells/ml) in RPMI 1640 (10% FBS) and restimulated with 50 µg/ml MOG₃⁵–₅₅ peptide. Cell culture supernatants were recovered after 72 h and IL-10 was measured using a commercial ELISA kit (R&D Systems, Minneapolis, MN, USA), according to the manufacturer's instructions.

2.5.6. CNS-infiltrating mononuclear cell isolation

At D15 post-immunization, animals were perfused with saline, and brains and spinal cords were removed. A single-cell suspension was obtained and passed through a 70-mm strainer. Mononuclear cells were obtained by Percoll gradient (37%/70%) centrifugation and collected from the interphase. Cells were washed and stimulated for 4 h with PMA + ionomycin in the presence of Brefeldin A (GolgiPlug; BD Biosciences). Cells were labeled with LIVE/DEAD ultraviolet-blue dye (Invitrogen) followed by surface staining (CD45 from Invitrogen and CD4, CD8, TCR-γδ, CD11b, and TCR-β from BD Biosciences). Afterward, cells were fixed, permeabilized, and stained for intracellular IL-17A (BD Biosciences) and IFN-γ (Invitrogen).
2.5.7. Antibodies and flow cytometric analysis

Single-cell suspensions of lymph node cells and splenocytes were prepared and the red blood cells were lysed with ammonium chloride. Total numbers of cells were counted using the Advia 120 hematology analyzer (Bayer/Siemens, Tarrytown, NY). For flow cytometric analysis, the cells were washed twice and incubated for 30 min on ice with the desired fluorochrome-conjugated mAbs or isotype control Ig at 0.5 μg/10^6 cells. For the identification and phenotypic analysis of Treg cells (CD4+ CD8– TCRβ+ Foxp3+), the following surface anti-mouse mAb were used: anti-CD4 (MCD0417, Caltag); anti-CD8, and anti-CD25 (53-6.7, PC61; Biolegend); anti-TCRβ, and anti-Foxp3 staining set (H57-5987 and FJK-16s; eBioscience), according to the manufacturer’s instructions. Viable cells were selected for flow cytometric analysis (LSR II, BD) based on forward and side scatter properties and analysis was performed using FlowJo software (TreeStar Software, Inc).

2.5.8. In Silico analysis and Data Mining

Murine homologs of all SLAMF members were identified using Mouse Genome Informatics (MGI; http://www.informatics.jax.org/). Using the SNP search tool in MGI and default settings of 2kb upstream and downstream of the gene, each Slamf gene was run to identify polymorphism between 129 and B6 mouse strains. The genes underlying c7-c8 interval were retrieved using Genes and Markers Query Form in MGI and genomic coordinates as input. This list was filtered through SNP search tool in MGI to focus on those genes that are polymorphic between 129 and B6 mouse strains.
Public high-throughput RNA-sequencing and microarray data were acquired from Gene Expression Omnibus (GEO) of National Center of Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/geo/) to query genes within c7-c8 interval whose expression was modulated by orchiectomy (GSE54945, GSE5901, and GSE66873). The keywords searched in GEO were (“gene name” AND “castration”; “gene name” AND “orchiectomy”; “gene name” AND “androgens”, and “gene name” and “gonadectomy” In the GEO dataset search, GEO series, which met the following criteria were selected for our study: (1) Samples contained castration/orchiectomized and sham control group. (2) Samples were from humans (primary cell lines) or mouse (any tissue). (3) Series detected expression profiling by RT-qPCR/array/high-throughput sequencing data. The resulting final gene list was used to search primary literature in the public library of medicine (PubMed) of NCBI with the following key words: “gene name” AND “autoimmunity”; “gene name” AND “CNS disease”; “gene name” AND “inflammation”; “gene name” AND “disease”; “gene name” AND “neuroinflammation”, and “gene name” AND “EAE disease”.

98
2.5.9. Statistics

Statistical analyses as indicated in the figure legends were performed using GraphPad Prism 7 software (GraphPad Software Inc).
2.6. Acknowledgements

We thank Roxana del Río Guerra for help with flow cytometry. AR, DNK, and CT were supported by grants from the NIH and the National Multiple Sclerosis Society (NMSS). DNK was supported by NIH grants from the National Institute of Neurological Disease and Stroke (R01 NS097596), National Institute of Allergy and Infectious Disease (R21 AI145306), and the NMSS (RR-1602-07780).
## 2.7. References


Consortium, I. M. S. G. (2013). Network-Based Multiple Sclerosis Pathway Analysis with Gwas Data from 15,000 Cases and 30,000 Controls. *Am J Hum Genet*, 92(6), 854-865.


106
2.8. Figures

Figure 2.1. B6 mice with overlapping intervals encompassing the 129-derived Slam locus used in this study.

The congenic C57BL/6J.129c1 (c1) and C57BL/6J.129c2 (c2) lines have been described in literature (Wandstrat et al., 2004; DeVault et al., 2019). Black denote B6 derived regions while white denotes 129 derived regions. The coordinates of 129-derived interval in c1 and c2 congenic mice are listed together with the number of predicted protein coding genes. MS-GWAS candidates Slamf1, Slamf2 and Slamf7 reside between 171.6-171.9Mb.
Figure 2.2. Male c1 mice exhibit increased resistance to EAE.

Male and female B6, c1 and c2 mice at 8-12 weeks of age were immunized with 100µg of MOG\textsubscript{35-55} + CFA on D0 and D7. The clinical scores following immunization were recorded, and the significance of differences between clinical courses was calculated by regression analysis with the
best-fit curve shown, and two-way ANOVA followed by Holm-Sidak’s multiple comparison test. Error bars indicate standard error of mean deviations within the groups. The number in parenthesis indicate number of mice used. (A) A significant difference in the severity of disease course was not detected among the female mice (interaction \( p > 0.05 \); strain \( p > 0.05 \); time \( p < 0.0001 \)). (B) The severity of the disease course of male c1 mice was significantly reduced compared to both B6 and c2 mice (interaction, \( p < 0.0001 \); strain, \( p < 0.0001 \); time, \( p < 0.0001 \); B6 = c2 > c1).
Figure 2.3. Ex-vivo production of IL-10 by MOG35-55 specific T cells.

Draining lymph node (DLN) cells were isolated on D10 post-immunization, and restimulated with 5μg/ml MOG35-55 for 3 days. Production of IL-10 was determined by ELISA. Significance of the observed differences was determined by two-way ANOVA followed by Holm-Sidak’s multiple comparisons test (** p = 0.0019). Error bars indicate standard error of mean (SEM). Each group includes 3-5 animals.
Figure 2.4. EAE severity was enhanced in orchiectomized (ORX) c1 mice compared to sham mice while neither Testosterone (T) nor 5α-dihydrotestosterone (DHT) replacement ameliorate disease susceptibility.

Male c1 mice were ORX or received sham surgeries at 4-6 weeks of age. The mice were rested 2 weeks, and then underwent hormone replacement with 90-day release pellets containing either 5mg DHT, 12.5 mg T, or placebo control pellets (CNT) (Innovative Research of America, FL). Pellets were implanted 7–10 days prior to EAE induction. The clinical scores following immunization were recorded, and the significance of differences between clinical courses was calculated by regression analysis with the best-fit curve shown. Error bars indicate SEM. The number in parenthesis indicate number of mice used. (A) The severity of the disease course different significantly among the groups (interaction, \( p < 0.0001 \); strain, \( p < 0.0001 \); time, \( p < 0.0001 \)) with \( c1 \) ORX + CNT= \( c1 \) ORX + DHT = \( c1 \) ORX + T > \( c1 \) Sham or \( c1 \) as analyzed by two-way ANOVA followed by Holm-Sidak’s multiple comparison test. (B) The cumulative disease score (CDS: summation of daily scores) metric among treatment groups was analyzed by one-way ANOVA followed by Holm-Sidak’s multiple comparison test; \( p < 0.0001 \). The numbers within the columns is incidence (incidence: number of animals with clinical signs ≥1 for two or more days). The
significance of the observed difference in incidence was assess by Chi-square analysis; \( p = 0.004 \).

(C) Blood was obtained by intracardiac puncture from Sham, ORX and ORX + T supplemented or ORX and ORX + DHT treated c1 mice on D30 post EAE immunization. Serum levels of T and DHT were determined by ELISA in duplicate according to the manufacturer’s directions and the significance of observed differences calculated (T, \( p = 0.003 \) determined by Kruskal Wallis one-way ANOVA followed by Dunn’s multiple comparison test; DHT, \( p = 0.09 \) determined by the Mann-Whitney test). *\( p \leq 0.05 \), **\( p \leq 0.01 \), ***\( p \leq 0.001 \), ****\( p < 0.0001 \).
Figure 2.5. Generation and EAE testing of interval specific recombinant congenic (ISRC) lines across c1 interval confirms that EAE susceptibility in male mice is controlled by a locus in linkage disequilibrium to *Slamf1*, *Slamf2* and *Slamf7*.

Overlapping interval specific recombinant congenic (ISRC) lines were generated by crossing c1 to B6 mice. F2 hybrids were genotyped using tail snip DNA and PCR with Chr 1 microsatellite markers and deletion primers (Supplementary Table 2.2) discriminating 129 and B6 alleles across the c1 candidate interval. Mice carrying variable regions of c1 interval were selected and backcrossed an additional two generations to b6 mice. The lines were fixed by brother-sister mating to generate the c7, c8, and c10. (A) Male B6, c1, c2, c7, c8, and c10 mice at 8-12 weeks of age were immunized with 100µg of MOG35-55 + CFA on D0 and D7. The clinical scores following immunization were recorded, and the significance of differences between clinical courses was calculated by regression analysis with the best-fit curve shown. Error bars indicate standard error of mean deviations within the groups. The number in parenthesis indicate number of mice used. The clinical course of EAE differed significantly among the congenic lines as determined by repeated measures two-way ANOVA followed by Holm-Sidak’s multiple comparison test (interaction, $p < 0.0001$; strain, $p < 0.0001$; time, $p < 0.0001$) with B6 = c8 = c10 > c1 = c7. **** $p < 0.0001$. (B) The CDS metric among ISRC lines was analyzed by one-way ANOVA followed by Holm-Sidak’s multiple comparison test; $p < 0.0001$. (C) Congenic intervals on Chr-6. Each row represents a congenic line that denotes the approximate position (bp) of the congenic region (x-axis). The black portions of each row represent the regions homozygous for the B6 allele (the background strain) while the white portions represent the 129 regions (donor alleles). Region outlined in blue depicts the location of causal gene(s) regulating CNS autoimmunity.
Figure 2.6. Polymorphic candidate genes within the c1 locus whose expression is modulated by orchiectomy.

The list of thirty-nine genes residing within the ~1.2Mb c7-c8 interval controlling EAE susceptibility in male mice were filtered through the following criteria: 1) the gene is polymorphic between B6 and 129 strains of mice as assessed using Mouse Genomic Informatics (MGI) SNP database (http://www.informatics.jax.org/snp); and 2) the gene exhibit transcriptional change in expression following castration (ORX) as assessed by mining publicly available gene expression datasets in Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/). The final list of 13 candidates are searched in PubMed for gene function (See Material and Methods).

aGEO accession for cortical thymic epithelial cells #GSE66873, spleen CD4+ T cells # GSE54945, and prostrate xenograft tumor GSE33316. b Filtering was done using all coding and non-coding SNP available in MGI. cList of literature references (Bickerstaff et al., 1999; Griffin et al., 2009; Choubey et al., 2010; Na et al., 2015; Chen et al., 2016; Kinoshita et al., 2016; Zeng et al., 2018; Sever et al., 2019b).
2.9. Tables

Table 2.1. *SLAM* locus is highly conserved between human and mouse with natural genetic variation segregating between B6 and 129 mice.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Human Location</th>
<th>Mouse Location</th>
<th>#SNPa</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>SLAMF1</em></td>
<td>1:160608106-160647295 → Slamf1</td>
<td>1:171767127-171801184</td>
<td>137</td>
</tr>
<tr>
<td><em>SLAMF2</em></td>
<td>1:160678746-160711831 → Slamf2</td>
<td>1:171682009-171705258</td>
<td>96</td>
</tr>
<tr>
<td><em>SLAMF3</em></td>
<td>1:160796074-160828261 → Slamf3</td>
<td>1:171588624-171607410</td>
<td>114</td>
</tr>
<tr>
<td><em>SLAMF4</em></td>
<td>1:160830160-160862887 → Slamf4</td>
<td>1:171559193-171609746</td>
<td>260</td>
</tr>
<tr>
<td><em>SLAMF5</em></td>
<td>1:160541095-160579516 → Slamf5</td>
<td>1:171839697-171890718</td>
<td>100</td>
</tr>
<tr>
<td><em>SLAMF6</em></td>
<td>1:160485030-160523262 → Slamf6</td>
<td>1:171917515-171953170</td>
<td>101</td>
</tr>
<tr>
<td><em>SLAMF7</em></td>
<td>1:160739057-160754821 → Slamf7</td>
<td>1:171632403-171653035</td>
<td>105</td>
</tr>
<tr>
<td><em>SLAMF8</em></td>
<td>1:159826811-159837492 → Slamf8</td>
<td>1:172581758-172590568</td>
<td>73</td>
</tr>
<tr>
<td><em>SLAMF9</em></td>
<td>1:159951492-159954237 → Slamf9</td>
<td>1:172475358-172478575</td>
<td>99</td>
</tr>
</tbody>
</table>

*SLAM* genes highlighted in red are MS-GWAS candidates. The location of genes is based on GRCh37 (Human) and GRCm38 (Mouse). *Number of SNPs distinguishing B6 and 129 mice were determined using comparative sequence data across each Slam gene.
Table 2.2. Immune profile of c1 and B6 mice

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Sex</th>
<th>DLN</th>
<th>CNS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B6</td>
<td>N</td>
</tr>
<tr>
<td>CD45+</td>
<td>M</td>
<td>33.1 ± 5.3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>71.0 ± 5.3</td>
<td>4</td>
</tr>
<tr>
<td>CD11b+</td>
<td>M</td>
<td>1.8 ± 0.1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>8.8 ± 1.5</td>
<td>4</td>
</tr>
<tr>
<td>CD11c+</td>
<td>M</td>
<td>31.1 ± 2.4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>31.3 ± 3.2</td>
<td>4</td>
</tr>
<tr>
<td>CD19+</td>
<td>M</td>
<td>63.2 ± 3.0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>63.9 ± 3.2</td>
<td>4</td>
</tr>
<tr>
<td>TCRαβ+</td>
<td>M</td>
<td>51.9 ± 0.6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>51.8 ± 0.9</td>
<td>9</td>
</tr>
<tr>
<td>IL-17+</td>
<td>M</td>
<td>0.9 ± 0.2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.8 ± 0.1</td>
<td>4</td>
</tr>
<tr>
<td>IFN-γ+</td>
<td>M</td>
<td>8.8 ± 0.1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>12.3 ± 0.4</td>
<td>4</td>
</tr>
<tr>
<td>Foxp3+</td>
<td>M</td>
<td>12.3 ± 0.4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>10.1 ± 0.7</td>
<td>5</td>
</tr>
<tr>
<td>CD8+</td>
<td>M</td>
<td>36.6 ± 1.0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>36.3 ± 1.3</td>
<td>4</td>
</tr>
<tr>
<td>IL-17+</td>
<td>M</td>
<td>0.5 ± 0.1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.0 ± 0.6</td>
<td>4</td>
</tr>
<tr>
<td>IFN-γ+</td>
<td>M</td>
<td>3.4 ± 0.3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>2.9 ± 0.4</td>
<td>3</td>
</tr>
<tr>
<td>TCRγδ+</td>
<td>M</td>
<td>1.7 ± 0.1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.7 ± 0.2</td>
<td>4</td>
</tr>
<tr>
<td>IL-17+</td>
<td>M</td>
<td>9.2 ± 0.7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>8.4 ± 0.6</td>
<td>4</td>
</tr>
<tr>
<td>IFN-γ+</td>
<td>M</td>
<td>12.8 ± 2.0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>12.0 ± 2.3</td>
<td>4</td>
</tr>
</tbody>
</table>

Mononuclear cells were isolated from the draining lymph nodes and central nervous system of EAE immunized mice on D10 and D22 respectively post immunization, stimulated with PMA/ionomycin for 4 h in the presence of Brefeldin A, stained, and analyzed by flow cytometry. Representative bar graph data of the frequency of immune cells from draining lymph nodes and CNS is shown as mean ± SEM of n = 5-10 mice per strain. Significance of differences observed in the percentage of immune cells was determined by two-way ANOVA followed by Holmes Sidak’s post hoc test.
### Supplementary Table 2.1. Clinical disease metrics of c1, c2 and B6 mice.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sex</th>
<th>N</th>
<th>Incidence (%)</th>
<th>CDS</th>
<th>DO</th>
<th>PS</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>F</td>
<td>33</td>
<td>22/33 (67%)</td>
<td>19.8 ± 2.9</td>
<td>17.5 ± 0.5</td>
<td>1.8 ± 0.3</td>
<td>8.8 ± 1.2</td>
</tr>
<tr>
<td>c1</td>
<td>F</td>
<td>20</td>
<td>16/20 (80%)</td>
<td>20.9 ± 3.1</td>
<td>18.1 ± 0.5</td>
<td>2.1 ± 0.3</td>
<td>9.6 ± 1.3</td>
</tr>
<tr>
<td>c2</td>
<td>F</td>
<td>29</td>
<td>22/29 (76%)</td>
<td>24.0 ± 3.1</td>
<td>17.5 ± 0.5</td>
<td>2.1 ± 0.3</td>
<td>10.3 ± 1.0</td>
</tr>
</tbody>
</table>

\[ x^2 = 1.3 \]

\[ p = 0.5 \]

\[ B6 = c1 = c2 \]

B6 = c1 = c2

\[ F = 0.6 \]

\[ F = 0.4 \]

\[ F = 0.5 \]

\[ F = 0.4 \]

\[ p = 0.6 \]

\[ p = 0.7 \]

\[ p = 0.6 \]

\[ p = 0.6 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

\[ B6 = c1 = c2 \]

B6 = c1 = c2

B6 = c1 = c2

B6 = c1 = c2

B6 = c1 = c2

B6 = c1 = c2

Male and female B6, c1 and c2 mice at 8-12 weeks of age were immunized with 100µg of MOG<sub>35</sub>-55 + CFA on D0 and D7. Animals were scored daily starting on D10 post injection till D30. Clinical quantitative trait variables were generated as previously described (Butterfield et al., 1998). Mice were considered positive for incidence if they showed any clinical signs ≥1 for two or more days. Cumulative disease score (CDS) is the summation of daily scores, day of onset (DO) was the first day at which an animal presents with a score of ≥1, peak score (PS) is the highest score attained, days affected (DA) is the total days in which a mouse was at a score of ≥1. Significance of differences in incidence was determined by chi-square. The significance of the observed differences in other quantitative trait variables was determined by two-way ANOVA followed by Holm-Sidak’s multiple comparisons test when a significant overall effect was detected.
Supplementary Table 2.2. Primers used in genotyping B6.129 congenic lines.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Location</th>
<th>Sequence</th>
<th>B6</th>
<th>129</th>
</tr>
</thead>
<tbody>
<tr>
<td>169.71</td>
<td>169,704,331</td>
<td>F: 5'TCTCCTCACCCCCCAGTCTTA3'</td>
<td>588</td>
<td>433</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TGGCTCCCTTTGATTGACTC3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>169.90</td>
<td>169,906,770</td>
<td>F: 5'GAACAAAGTCTGCCCCCTTCTG3'</td>
<td>583</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GAAGTCCAGGGGGATCTGAC3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171.00</td>
<td>171,044,494</td>
<td>F: 5'CCAGTTTTCCAGGGCAAGATA3'</td>
<td>532</td>
<td>686</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GGACTGCCCCTCAACACACTA3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1Mit113</td>
<td>171,804,480</td>
<td>F: 5'CCTCAAATCAGGATTTAAGGG3'</td>
<td>206</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'ACATGGGGTTGGACTTGTGAT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171.83</td>
<td>171,835,566</td>
<td>F: 5'CTCTCCCTGAACCTGAGAC3'</td>
<td>766</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'CGCTCCAGATGTCCCCCAT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171.87</td>
<td>171,873,666</td>
<td>F: 5'GGCATTTCTGTTGGTGTT3'</td>
<td>690</td>
<td>445</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GGCTTGACCCCAGTGAC3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171.91</td>
<td>171,913,430</td>
<td>F: 5'GAAGGTCCATTGCTGCTT3'</td>
<td>559</td>
<td>353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TATAGAGGTGTTGGCGCAAGG3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.03</td>
<td>172,031,928</td>
<td>F: 5'GTAGCTGCCAATGTTGAT3'</td>
<td>569</td>
<td>354</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'CCCCTTTTGTCATTGTGTG3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.08</td>
<td>172,087,845</td>
<td>F: 5'ACACAAATGGGGTCTACAGT3'</td>
<td>641</td>
<td>364</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'CCAACACCTGGCCTACTAA3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs245610856</td>
<td>172,100,795</td>
<td>F: 5'TAGCCTGAGCAACAGCAAAAA3'</td>
<td>645</td>
<td>618</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TGACCTGCCCCATAGTGTT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.13</td>
<td>172,131,470</td>
<td>F: 5'CAAGGGATTTTGTGGTGGT3'</td>
<td>531</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'ATGTCTGTGCGCCCTCCCATAG3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.18</td>
<td>172,178,303</td>
<td>F: 5'AGGTTTTGTAGAGGCTCA3'</td>
<td>578</td>
<td>383</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GGACCTGAGCTGGATTACA3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.27</td>
<td>172,270,866</td>
<td>F: 5'TGGGTATACCTGGGACTGA3'</td>
<td>600</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GGGCTATTGGTACAGTATG3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.28</td>
<td>172,282,646</td>
<td>F: 5'CAGGCTGAATCCTCCATCAG3'</td>
<td>566</td>
<td>413</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'ATGAGGGTCAGAGAGAGAGA3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.32</td>
<td>172,324,995</td>
<td>F: 5'CCAGGTTCCTAGCTAGCAG3'</td>
<td>542</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TGACTGTGCTGCTGTCGAT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1Mit206</td>
<td>173,004,198</td>
<td>F: 5'TGAGGCACCTTGTATTCAG3'</td>
<td>123</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'CCAGATGTTCTTTGAAACCTTTC3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1Mit456</td>
<td>172,559,415</td>
<td>F: 5'TGGCTTCCAGAAGGATAG3'</td>
<td>113</td>
<td>227</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'GCCAGTGAGATGCACAGAC3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1MiT149</td>
<td>172,699,229</td>
<td>F: 5'AAAGAGAATCTGACTTACCCATAGG3'</td>
<td>100</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TGTGAGGGAGAAGATTATGTCTT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.70</td>
<td>172,701,841</td>
<td>F: 5'TGAGGTTGGTGCTCCGTAG3'</td>
<td>536</td>
<td>671</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'TCCCTTGGGTTGCTAACCTG3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>173.00</td>
<td>173,408,539</td>
<td>F: 5'AGGTGGGGAGCAATGAGAT3'</td>
<td>602</td>
<td>364</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: 5'AAACCCTGAAGCCAGTCGCA3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Size</td>
<td>F:</td>
<td>R:</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>174.60</td>
<td>174,603,762</td>
<td>5'TGCCATGAACATTGGGTAAA'</td>
<td>5'TGGCTAGCTCTTGGAGAGGA3'</td>
<td></td>
</tr>
<tr>
<td>403.00</td>
<td>175,641,251</td>
<td>5'TATTGAGGGGTGTTTTTATTTCTC3'</td>
<td>5'TCCCACGGGTCCCTGTATTCC3'</td>
<td></td>
</tr>
<tr>
<td>175.97</td>
<td>175,969,658</td>
<td>5'ATGCCTGGCAAAACATCTTC3'</td>
<td>5'GGTTGCTTTGTCCCATCCAGT3'</td>
<td></td>
</tr>
<tr>
<td>176.40</td>
<td>176,415,010</td>
<td>5'CTCTGGGAATTTGGAACAA3'</td>
<td>5'GATTGGCATGCTCTCTCTTC3'</td>
<td></td>
</tr>
<tr>
<td>D1MIT115</td>
<td>177,680,220</td>
<td>5'AAGGGAATGGGAATTAGGGTCA3'</td>
<td>5'TAACGGACACCCATTTTTAAACA3'</td>
<td></td>
</tr>
</tbody>
</table>

The location of primers is based on GRCm38 (Mouse). The size of PCR amplicon for each primer pair is listed under B6 and 129 strains are listed.
CHAPTER 3: CONCLUSION AND FUTURE DIRECTIONS.

Large scale quantitative trait loci (QTL) linkage mapping and genome-wide association studies (GWAS) have uncovered hundreds of thousands of genetic loci associated with human diseases. Despite generating promising results in complex disorders, both methods have limitations. For example, QTL studies are limited by resolution of causal loci, often resulting in large regions spanning entire chromosomes. Moreover, large QTLs when interrogated by high-resolution congenic mapping often reveals that the overall QTL effect is due to multiple linked genes within the QTL rather than a single gene (Yazbek et al., 2011; Parker et al., 2013). GWAS approaches, on the other hand, do define narrow regions of association. However, the amount of variance explained by GWAS is often small. Moreover, the resolution and interpretation of causal loci in disease from QTL linkage mapping or GWAS findings require functional testing.

In this work, we complement genetic predictions for several human diseases including multiple sclerosis (MS) and systemic capillary leakage syndrome (SCLS) with genetic and functional data in model organisms to associate loci with phenotypes and diseases.

In Chapter 2, we describe testing a genetic locus identified by several GWAS studies among MS patients to be associated with disease. MS is a complex autoimmune disease of CNS affecting young adults with no effective cure (Charcot, 1868; Lubetzki, 2018). The definitive cause for MS is unknown. However, a multipart etiology is suggested whereby genetics x environment x sex interactions define the overall outcome of disease. A familial nature of MS was suggested back in 1988 whereby siblings of affected individuals were found to exhibit a 30 to 50 times higher risk to develop disease (Burrows
et al., 2019a). Later studies confirmed this among monozygotic MS twins where it was found that if one twin had MS, 30-35% of the time the other twin also developed MS (Willer et al., 2003). Thus, began studies to identify the genetic basis of MS with the hope to discover highly penetrant loci (Xu et al., 2001). Except for association with HLA, most of these studies failed to identify loci associated with MS risk. This led to a major shift in the understanding of complex disorders like MS where it was hypothesized that disease-causing alleles or variants are common in the population and have a very small additive or multiplicative effect to phenotype (Collins et al., 1997; Pritchard et al., 2002).

Following the completion of human genome sequencing project in 2003 and subsequent information on common genetic variants with the HapMap project in 2005, interest started to mount to screen for common genetic variants in MS and other complex disorders (Manolio et al., 2009). This led the rationale and drive to perform large GWAS that have so far uncovered 233+ genetic variants that are associated with MS susceptibility (Baranzini et al., 2017). Considering these are merely associated variants, the next obvious step is to translate the GWAS findings into functional insights (causal variants) that can allow one to predict targets for MS therapies. Except for a handful of functional studies with IL7R, DDX39B, TNFRF1A and EVI5, a vast majority of MS GWAS candidates have not been validated (Gregory et al., 2007; Gregory et al., 2012; Didonna et al., 2015; Galarza-Munoz et al., 2017). The same trend is apparent in all the GWAS studies of complex diseases in which functional follow-up lags the discovery efforts (Figure 3.1).
Several questions are raised in Chapter 2 about the utility of GWAS studies in MS and if we are learning anything important by continuing to expand the list of associated variants. First, most of the significant SNPs in any GWAS lie in non-coding regions of the genome (>90%) and are annotated to the closest gene. This excludes the possibility that the variant could have effects on a gene far away (*trans*-acting variant) and could be important in the biology of disease. Understanding how non-coding SNPs alter MS susceptibility remains a significant challenge, due to incomplete understanding of cis and trans effects of variants (Reuveni et al., 2018). Second, the general design of GWAS excludes contributions from rare variants (<5% mean allele frequency in population) that are not genotyped (common variant-common disease model). In fact, rare variants could...
explain a substantial fraction of heritance in MS (Mitrovič et al., 2018). Third, GWAS assumes a single-gene model in which each locus acts independently of the others (Spencer et al., 2009). This excludes effects of epistasis or genetic interactions between variants that have been shown to exist in MS (Dyment et al., 2005). Fourth, linkage disequilibrium (LD) is an inherent characteristic of the human genome, thus the distinction between GWAS associated variant and causal variant may be masked and may require additional fine mapping to exclude false-positive associations. Moreover, LD structures vary across different ethnicity that may muddle the overall picture. Fifth, a clear majority of MS-GWAS studies are performed in Caucasians, thus, finding of genetic variants may not be truly applicable across other population groups (African Americans, Ashkenazi Jews) (Martin et al., 2018). In fact, it has been demonstrated that variants associated with diseases found in European ancestry populations do not always replicate in non-European populations (Simon-Sanchez et al., 2008; Weiss et al., 2009; Yamada et al., 2009; Haga, 2010). Sixth, knowledge about the genetic basis of heterogeneity in distinct clinical courses of MS is incomplete with few GWAS studies that predict suggestive associations with age of onset, clinical severity, brain atrophy, glutamate levels and MRI T2 lesion load with disease susceptibility (Baranzini et al., 2009; Baranzini et al., 2010a; Brynedal et al., 2010b; Martinelli-Boneschi et al., 2012). Seventh, the high prevalence of MS in women, clinical heterogeneity of disease course between men and women and distinct effects of sex-hormones all warrant an explanation from GWAS studies. However, very few studies have tested genes on sex chromosomes or segregated genetic data by sex (Baranzini et al., 2009; Patsopoulos et al., 2017). A SNP with different disease associations in men and
women can provide insight into the established sex difference in MS that may further guide the discovery of how disease mechanisms differ between sexes. Thus, it is imperative to validate these genetic variants using appropriate models to determine causality with disease.

Several animal models have been developed over the years to understand the various aspects of human MS. Among them, EAE is the most extensively studied and allows one to dissect the contribution of specific risk factors (Constantinescu et al., 2011). Although these animal models cannot fully replicate the MS disease course, they help perform mechanistic studies addressing disease pathogenesis that cannot be readily performed in MS patients. In fact, several of the approved disease-modifying drugs in MS came from EAE studies (Burrows et al., 2019a).

We chose to model the signaling lymphocytic activation molecule (SLAM) locus for functional testing in this study for the following reasons. First, this locus encodes a family of nine proteins that are expressed on hematopoietic cells and exert important immunoregulatory functions including CD4+ T cell differentiation, CD8+ T cell proliferation, cytokine production by macrophages and dendritic cells, B cell activation, antibody production, NK and natural-killer T cells (NKT) cell functions (Veillette et al., 2003; Wang et al., 2004; Aktan et al., 2010; Chatterjee et al., 2012; Chu et al., 2014). Second, three of the SLAM members (SLAMF1, 2, and 7) came up as susceptibility loci in numerous MS-GWAS studies (Beecham et al., 2013; Patsopoulos et al., 2017; Madireddy et al., 2019). Third, SLAM is an important susceptibility locus in other autoimmune diseases including Graves’ disease, systemic lupus erythematosus, Crohn's disease,
ankylosing spondylitis, psoriasis, primary sclerosing cholangitis and ulcerative colitis (Zhao et al., 2013; Ellinghaus et al., 2016; Langefeld et al., 2017). Considering there is sharing of genetic loci between autoimmune diseases, an understanding of SLAM genes in MS would help guide functional efforts in other autoimmune diseases (Márquez et al., 2018). Fourth, the SLAM locus is highly conserved across several species including chimpanzee, dog, cow, rat, and mouse suggesting an important immunological function. Fifth, the Slam (SLAM) locus exhibits natural genetic variation in mouse that exhibits as two stable and divergent haplotypes (Haplotypes 1 and 2) (Wandstrat et al., 2004). Sixth, two congenic lines have been described in the literature that capture various intervals of the Slam locus (Wandstrat et al., 2004) (DeVault et al., 2019). We, therefore, thought to leverage these mouse congenic lines and some additional lines that we generated with natural genetic variation in the Slam locus between 129 and B6 strains to test the hypothesis that SLAMF1/Slamf1, SLAMF2/Slamf2, and/or SLAMF7/Slamf7, and not a disease-related gene in linkage disequilibrium with the SLAM/Slam locus, are the true GWAS-genes contributing to CNS autoimmune disease susceptibility.

Our EAE experiments with c1, c2, c7, c8 and c10 congenic lines established that the causal variant in the c1 congenic line is not Slamf1, Slamf2 or Slamf7 but is predicted to lie in a small ~1.2Mb interval between 172.28-173.41Mb, which does include Slamf8 and Slamf9 (Figure 2.5). Considering several immunoregulatory cell types are associated with susceptibility and resistance to EAE and MS disease, we dissected the decreased severity of disease in c1 males through immune profiling experiments (Cheng et al., 2017). We found an increased frequency of CD11b+ cells and TCRαβ+ CD4+ Foxp3+ in DLN
and TCRαβ+ CD4+ Foxp3+ and TCRαβ+ CD8+ IL-17+ T cells in the CNS that co-segregate with EAE resistance in c1 male mice. In addition, we found increased production of IL-10 cytokine by CD4+ T cells following stimulation with MOG35-55 peptide that could explain reduced severity of autoimmune disease in c1 males. It remains to be determined whether the causal variant that regulates disease in c7 congenic line exhibits the same immune profile. Genetic cross between IL-10−/− mice and c7 congenic line followed by EAE and immune profiling of F1 would help to elucidate the mechanism of resistance. In addition, since CD4+ T cells play a major pathogenic role in EAE disease development, it will be helpful to include measurement of CD4+ T cell proliferation using variable doses of MOG35-55 peptide. Moreover, as a control for the frequency of immune cells in naïve state (no EAE induction), immune profiling of all congenic lines needs to be examined.

While MS is generally viewed as a female-centric disease, many studies have demonstrated that male sex is independently associated with rapid accumulation of disability (Finkelsztejn et al., 2011), poorer recovery from initial relapse (Cossburn et al., 2012) and display a more malignant form of MS (Gholipour et al., 2011), suggesting that the effects of sex in MS are much more complex. The resistance to EAE seen in the c1 congenic line is sex-specific with only the male mice showing amelioration of disease whereas females are as susceptible as B6. We confirmed this through orchiectomy experiments in male c1 mice that led to reversal of EAE resistance while neither of the male androgens tested (T or DHT) restored resistance of c1 congenic mice. This is a unique finding and it is possible that additional immunomodulatory molecules synthesized by testis in addition to male androgens e.g. inhibins and activins may regulate disease. It is
however unknown how these molecules can module the immune response in the periphery and CNS (Setchell, 1974; Uhlen et al., 2015; Ahn et al., 2017).

We have extensively utilized publicly available databases of gene expression such as gene expression omnibus (GEO) in this work to retrieve information from other researchers on gene expression following orchiectomy. Fine mapping of causal variant on Chr 1 was performed using an in-silico approach by first sorting candidates within the ~1.2Mb locus that were polymorphic between the two haplotypes (129 vs B6) and then using data mining in GEO to identify those candidates whose expression was known to be modulated by orchiectomy. Using this approach, we were able to identify thirteen genes that were polymorphic, including Atp1a2, Apcs, Cadm3, Dsp23, Fcer1a, Kcnj9, Kcnj10, Olfr16, Pgm, Slamf8, Slamf9, Tagln2, Vsig8. Interestingly, several of these candidates have important autoimmune functions (Figure 2.5) and it will be critical to validate these using quantitative PCR in several cell types including CD11b+ cells, TCRαβ+ CD4+ Foxp3+, and TCRαβ+ CD8+ IL-17+ T cells. Additionally, since several of the candidate genes in the ~1.2Mb interval have genetic knockouts available, one can assess their contribution to CNS autoimmune disease using F1 cross with male c1 or c7 congenic line.

In Appendix A, we describe an extremely rare disorder of unknown etiology called systemic capillary leakage syndrome (SCLS), which is characterized by recurrent episodes of vascular leakage. There are currently fewer than 200 cases with a confirmed diagnosis worldwide although its prevalence is on the rise, likely due to increased awareness among physicians and the public (Druey et al., 2017). Complications of acute SCLS include shock, compartment syndrome, and multi-organ dysfunction (Druey et al., 2017).
The pathogenic mechanisms underlying SCLS are unknown, and consequently, treatments have been developed primarily by trial-and-error. SCLS attacks are diagnosed based on the clinical triad of hypotension, elevated hematocrit, and hypoalbuminemia. SCLS flares are frequently preceded by respiratory viral and other infections, suggesting a role for inflammation in the induction of acute vascular leak (Eo et al., 2018).

During SCLS flares, transient spikes in circulating angiogenic proteins known to trigger vascular hyperpermeability [e.g., angiopoietin 2 (Angpt2), and vascular endothelial growth factor (VEGFA)] have been detected (Xie et al., 2012; Xie et al., 2014a). Additionally, sera from SCLS patients during episodes have been shown to impair microvascular endothelial cell (EC) barrier function, whereas convalescent sera from these same patients are functionally benign (Xie et al., 2012; Xie et al., 2014b). These results suggest that humoral factors present during disease flares are responsible for promoting vascular leak and systemic pathology.

Patients with SCLS routinely develop symptoms in mid-life, and they lack a family history of this disorder, both findings that suggest that the genetic basis of disease is multifactorial and complex. Early studies of our initial SCLS patient cohort resulted in the identification of a small genetic interval, 3p25.3, as the highest-ranking candidate susceptibility locus ($p \sim 10^{-6}$) with an odds ratio of ~41 (Xie et al., 2013). Whole exome sequencing (WES) of a single patient with fatal SCLS revealed a potentially pathogenic loss of function mutation in the gene ARHGAP5, which encodes a known of a regulator of endothelial permeability (p190BRhoGAP) (Pierce et al., 2017). Notably, this mutation has not been detected in any other subjects with SCLS (Pierce et al., 2018). These results
suggest that SCLS may be genetically heterogeneous, which is yet another significant obstacle to a more definitive analysis of this rare disorder. An appropriate animal model could not only help delineate the role of genetic factors in SCLS, but also would serve as a pivotal tool for modeling gene-environment interactions in numerous, often life-threatening, disorders and diseases in which vascular hyperpermeability has a central pathogenic function (e.g. systemic anaphylaxis, sepsis, Ebola virus, and dengue) (Escudero-Perez et al., 2014; Mikelis et al., 2015; Wang et al., 2019).

Using publicly available mouse phenotype data, we identified a mouse strain, SJL/J (SJL), that uniquely and spontaneously displays the clinical features of SCLS—hypoalbuminemia, elevated hematocrit, and hypotension (Bogue et al., 2018). Here, we investigated the feasibility of using SJL mice as a model to interrogate pathophysiological mechanisms of SCLS. We discovered that SJL exhibit susceptibility to histamine- and infection-triggered vascular leak. This trait “Histamine hypersensitivity” (Histh/Histh) was mapped to a region on Chr 6 in SJL mouse. Remarkably, Histh is syntenic to the genomic locus most strongly associated with SCLS in humans (3p25.3). Subsequent studies found that the Histh locus is not unique to SJL but additional mouse strains also exhibit Histh phenotype. Considering GWAS studies in SCLS are limited by the small number of patients, we utilized interval-specific SNP based association testing among Histh phenotyped mouse strains to predict Histh candidates. Furthermore, to dissect the complexity of Histh QTL, we developed network-based functional prediction methods to rank genes in this locus by predicting functional association with multiple Histh-related
processes. The top-ranked genes include *Cxcl12, Ret, Cacna1c*, and *Cntn3*, all of which have strong functional associations and are proximal to SNPs segregating with Histh.

In Appendix B, we developed a computationally technique to prioritize positional candidates based on computationally inferred gene function. Our method uses machine learning with functional genomic networks, whose links encode functional associations among genes, to identify network-based signatures of functional association to a trait of interest. We demonstrate the method by functionally ranking positional candidates in Histh locus on Chr 6 (45.9 Mb to 127.8 Mb) associated with histamine hypersensitivity (Histh). To dissect its complexity, we ranked genes in the Histh locus by predicting functional association with multiple Histh-related processes. We integrated these predictions with single nucleotide polymorphism (SNP) association data derived from a survey of 23 inbred mouse strains. The top-ranked genes included *Cxcl12, Ret, Cacna1c*, and *Cntn3*, all of which had strong functional associations and were proximal to SNPs segregating with Histh. These results demonstrate the power of network-based computational methods to nominate highly plausible quantitative trait genes even in highly challenging cases involving large QTLs and extreme trait complexity.
In Appendix C, we utilize the power of integrating genetic and functional approaches to understand susceptibility to *Bordetella pertussis* and pertussis toxin (PTX) induced histamine sensitization (Bphs/Bphs), a sub-phenotype with an established role in autoimmunity. Susceptibility to Bphs is controlled by histamine H₁ receptor (*Hrh1/H₁R*) alleles, with the susceptible (Bphs<sup>S</sup>/H₁<sup>R</sup>) and resistant (Bphs<sup>R</sup>/H₁<sup>R</sup>) alleles differing by three amino acids (P263L, V313M, and P331S) within the third intracellular loop associated with signal transduction, protein folding, and trafficking. Functionally, the two alleles equally activate Gα<sub>q/11</sub>, the G protein family members that couple H₁R signaling to second messenger signaling backways, indicating that susceptibility and resistance is not due to differential activation of Gα<sub>q/11</sub>. In contrast, the two alleles exhibit differential cell surface expression and altered intracellular trafficking, with the H₁<sup>R</sup> allele preferentially retained within the endoplasmic reticulum (ER), and all three residues comprising the H₁<sup>R</sup> haplotype required for altered expression. Given the importance of H₁R signaling in health and disease we phenotyped and sequenced the third intracellular loop of H₁R among a large panel of inbred laboratory and wild-derived mouse strains to identify potential allelic co-adaptations capable of complementing Bphs<sup>R</sup> in mice with a H₁<sup>R</sup> allele. Taking this approach, we identified eight inbred mouse strains with a H₁<sup>R</sup> allele that are Bphs<sup>S</sup> (Bphs<sup>S</sup>/H₁<sup>R</sup>). Genetic analyses mapped the locus complementing Bphs<sup>R</sup> to mouse Chr 6, in linkage disequilibrium with *Hrh1*; designated Bphs-enhancer (*Bphse*). Similar to the approaches used for *Histh*, we utilized interval-specific SNP based association testing and network-based functional enrichment to predict nine candidate loci for *Bphse* including
Atp2b2, Atg7, Pparg, Syn2, Ifi122, Raf1, Mkrn2, Timp4, and Gt(ROSA)26Sor. These findings reveal an additional mechanism by which genetic factors control Bphs^S.

To conclude, this work integrates several genetic (linkage analysis, SNP based association testing, congenic mapping) and functional (animal models of disease, immune profiling, gene expression databases, functional network of tissues in mouse) approaches to understand complex pathologies in humans (multiple sclerosis, systemic capillary leakage syndrome) and mouse (histamine hypersensitivity) that help enhance our understanding of the genetic basis of disease. Further genetic and functional validation of prioritized genes will provide mechanistic insights into disease pathogenesis and may facilitate to provide potential targets for future therapeutics and diagnostics.
COMPREHENSIVE BIBLIOGRAPHY


Alcina, A., Fedetz, M., & Matesanz, F. (2019). Genetics of Multiple Sclerosis. In J. Martín & F. D. Carmona (Eds.), *Genetics of Rare Autoimmune Diseases* (pp. 183-202). Cham: Springer International Publishing.


Chu, C., Wang, Y., Zhang, X., Ni, X., Cao, J., Xu, W., Dong, Z., Yuan, P., Wei, W., Ma,
Ligation-Dependent and -Independent Ly108–Cd3ζ Interactions. J Immunol,
193(8), 3860-3871.

Macrophages and Microglia in Multiple Sclerosis and Experimental Autoimmune

Clarelli, F., Liberatore, G., Sorosina, M., Osiceanu, A. M., Esposito, F., Mascia, E.,
Interferon-Beta Treatment in Multiple Sclerosis. Pharmacogenomics J, 17(1), 84-91.

Barrier Abnormalities in Chronic-Progressive Multiple Sclerosis. Acta
Neuropathol, 90(3), 228-238.

Human DNA Sequence Variation. Science, 278(5343), 1580-1581.

from Large-Scale Biology. Science, 300(5617), 286-290.

Comabella, M., Craig, D. W., Camina-Tato, M., Morcillo, C., Lopez, C., Navarro, A., Rio,
Multiple Sclerosis at 13q31.3 by a Pooled Genome-Wide Scan of 500,000 Single
Nucleotide Polymorphisms. PLoS One, 3(10), e3490.

372(9648), 1502-1517.

Confavreux, C., Hutchinson, M., Hours, M. M., Cortinovis-Tourniaire, P., & Moreau, T.

Confavreux, C., Hutchinson, M., Hours, M. M., Cortinovis-Tourniaire, P., & Moreau, T.
(1998b). Rate of Pregnancy-Related Relapse in Multiple Sclerosis. Pregnancy in

Consortium, I. M. S. G. (2013). Network-Based Multiple Sclerosis Pathway Analysis with
Gwas Data from 15,000 Cases and 30,000 Controls. Am J Hum Genet, 92(6), 854-
865.

Constantinescu, C. S., Farooqi, N., O'Brien, K., & Gran, B. (2011). Experimental
Autoimmune Encephalomyelitis (Eae) as a Model for Multiple Sclerosis (Ms). Br
J Pharmacol, 164(4), 1079-1106.

Corvol, J. C., Pelletier, D., Henry, R. G., Caillier, S. J., Wang, J., Pappas, D., Casazza, S.,
Sclerosis after the Initial Neurological Event. Proc Natl Acad Sci USA, 105(33),
11839-11844.


168


APPENDIX A: A NATURAL MOUSE MODEL REVEALS GENETIC DETERMINANTS OF SYSTEMIC CAPILLARY LEAK SYNDROME (CLARKSON DISEASE)

Abbas Raza¹, Eunice C. Chan³, Wei-Sheng Chen³, Zhihui Xie³, Linda M. Scott³, A. Robin Eisch³, Dimitry N. Krementsov², Helene F. Rosenberg⁴, Samir M. Parikh⁵, Elizabeth P. Blankenhorn⁶, Cory Teuscher¹, Kirk M. Druey³

¹Departments of Medicine and Pathology, ²Department of Biomedical and Health Sciences, University of Vermont Burlington, VT 05405; ³Lung and Vascular Inflammation Section, ⁴Inflammation Immunobiology Section, Laboratory of Allergic Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892. ⁵Division of Nephrology and Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. ⁶Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, 19129, USA.

Address correspondence to: Kirk M. Druey, 10 Center Drive, Bethesda, MD; kdruey@niaid.nih.gov. Cory Teuscher: 89 Beaumont Avenue Given C Bldg, Rm 329A, University of Vermont Larner School of Medicine, Burlington, VT 05405; c.teuscher@uvm.edu
A.1. Abstract

The Systemic Capillary Leak Syndrome (SCLS, Clarkson disease) is a disorder of unknown etiology characterized by recurrent episodes of vascular leakage of proteins and fluids into peripheral tissues, resulting in massive whole-body edema and hypotensive shock. The pathologic mechanisms and genetic basis for SCLS remain elusive. Here we identify an inbred mouse strain, SJL, that recapitulates cardinal features of SCLS, including susceptibility to histamine- and infection-triggered vascular leak. We named this trait “Histamine hypersensitivity” (Histh/Histh) and mapped it to a region on Chr 6. Remarkably, Histh is syntenic to the genomic locus most strongly associated with SCLS in humans (3p25.3). These studies reveal that the predisposition to develop vascular hyperpermeability has a strong genetic component conserved between humans and mice and provide a naturally occurring animal model for SCLS. Thus, genetic analysis of the Histh locus has the potential to reveal and functionally validate orthologous candidate genes that contribute not only to SCLS but also to normal and dysregulated mechanisms underlying vascular barrier function more generally.
A.2. Introduction

The Systemic Capillary Leak Syndrome (SCLS, Clarkson disease) is a rare disease. There are currently fewer than 200 cases with a confirmed diagnosis worldwide although its prevalence is on the rise, likely due to increased awareness among physicians and the public (Druey et al., 2017). SCLS is characterized by transient but potentially lethal episodes of diffuse vascular leakage. Complications of acute SCLS include shock, compartment syndrome, and multi-organ dysfunction (Druey et al., 2010). The pathogenic mechanisms underlying SCLS are unknown, and consequently treatments have been developed primarily by trial-and-error. SCLS attacks are diagnosed based on the clinical triad of hypotension, elevated hematocrit, and hypoalbuminemia. SCLS flares are frequently preceded by respiratory viral and other infections, suggesting a role for inflammation in the induction of acute vascular leak (Eo et al., 2018).

During SCLS flares, transient spikes in circulating angiogenic proteins known to trigger vascular hyperpermeability [e.g., angiopoietin 2 (Angpt2), and vascular endothelial growth factor (VEGFA)] have been detected (Xie et al., 2012; Xie et al., 2014b). Additionally, sera from SCLS patients during episodes have been shown to impair microvascular endothelial cell (EC) barrier function, whereas convalescent sera from these same patients are functionally benign (Xie et al., 2012; Xie et al., 2014a). These results suggest that humoral factors present during disease flares are responsible for promoting vascular leak and systemic pathology.
Patients with SCLS routinely develop symptoms in mid-life, and they lack a family history of this disorder, both findings that suggest that the genetic basis of disease is multifactorial and complex. Early studies of our initial SCLS patient cohort resulted in the identification of a small genetic interval, 3p25.3, as the highest-ranking candidate susceptibility locus \( p \approx 10^{-6} \) with an odds ratio of \( \sim 41 \) (Xie et al., 2013). Whole exome sequencing (WES) of a single patient with fatal SCLS revealed a potentially pathogenic loss of function mutation in the gene \( ARHGAP5 \), which encodes a known regulator of endothelial permeability (p190BRhoGAP) (Pierce et al., 2017). Notably, this mutation has not been detected in any other subjects with SCLS (Pierce et al., 2018). These results suggest that SCLS may be genetically heterogenous, which is yet another significant obstacle to a more definitive analysis of this rare disorder. An appropriate animal model could not only help delineate the role of genetic factors in SCLS, but also would serve as a pivotal tool for modeling gene-environment interactions in numerous, often life-threatening, disorders and diseases in which vascular hyperpermeability has a central pathogenic function (e.g. systemic anaphylaxis, sepsis, Ebola virus, and dengue) (Escudero-Perez et al., 2014; Mikelis et al., 2015; Wang et al., 2019).

Using publicly available mouse phenotype data, we identified a strain of mice, SJL/J (SJL), that uniquely and spontaneously displays the clinical features of SCLS—hypoalbuminemia, elevated hematocrit, and hypotension (Bogue et al., 2018). Here, we investigated the feasibility of using SJL mice as a model to interrogate pathophysiological mechanisms of SCLS. Previous studies suggested that the SJL strain of mice is susceptible to systemic histamine, a canonical mediator of vascular hyperpermeability (Linthicum et
al., 1982). We report herein that mortality of SJL mice in response to administration of histamine was highly correlated with evidence of increased vascular leakage in a pattern similar to that reported in SCLS patients. Classical linkage studies revealed that a recessive locus in SJL mice controlling histamine-induced mortality mapped to a region on mouse Chr 6, which we designated Histh (histamine hypersensitivity). Strikingly, Histh is syntenic with human 3p25.3, the highest ranking SCLS susceptibility locus. Considering the similarity of the Histh-mediated phenotype to SCLS, the results suggest that humans and mice share genetic traits that predispose both species to stress-induced vascular dysregulation.
A.3. Results

A.3.1. Dermal vasculature of SCLS patients exhibit hyper-responsive to leak provocateurs.

*In vitro* studies of endothelial cells isolated from skin of an SCLS patient demonstrated exaggerated responses to inflammatory mediators, suggesting that primary endothelial dysfunction contributes to the clinical symptoms of SCLS (Pierce *et al.*, 2017). To test this hypothesis directly *in situ*, we injected histamine or morphine intradermally in patients with SCLS and healthy controls and measured the area of drug-induced skin wheals caused by fluid extravasation. Histamine evokes vascular leakage by acting directly on the endothelium whereas morphine functions indirectly through mast cell degranulation and release of various permeability-inducing mediators including histamine, leukotrienes, and prostaglandins; both agents have been used safely in a prior human study of cutaneous vascular responsiveness (Keffer *et al.*, 1989; Mikelis *et al.*, 2015; Nakamura *et al.*, 2018). We observed significantly larger wheal sizes in SCLS patients compared to healthy controls in response to a range of concentrations of either histamine or morphine (Figure A.1A-B). Thus, with two unrelated stimuli provoking exaggerated vascular leakage in SCLS patients—and doing so in a dose-proportional fashion—the results were most suggestive of a generalized vascular hyper-responsiveness in SCLS.

A.3.2. SJL mice exhibit traits that phenocopy human SCLS.

Seeking an *in vivo* model of vascular hyper-responsiveness, we first identified mouse strains with a constellation of traits resembling SCLS by searching the Mouse
Phenome Database (https://phenome.jax.org) for data on systolic blood pressure, hematocrit, and plasma albumin. Compared to the mean trait variables for all strains, SJL mice were unique in that they exhibited significantly lower systolic blood pressure in conjunction with an increased hematocrit and hypoalbuminemia ($p<0.01$, 2-way ANOVA, SJL v. all other strains (Figure A.2A-C). As a comparator strain for functional studies, we selected the B10.S/SgMcJ (B10.S) mouse, which, like SJL, carries the $H2^S$ haplotype at the $H2$ (MHC) locus, but has been reported to be insensitive to histamine (Linthicum et al., 1982). We have previously used various SJL/B10.S crosses in genetic studies of susceptibility to autoimmune neuroinflammation (Butterfield et al., 1998), a disease that may be regulated in part by histamine-mediated effects on vascular or immune systems, and controlling for MHC-regulated effects in studies of immune/inflammatory disorders is paramount (Ma et al., 2002; Lu et al., 2010). Given that neither B10.S nor C57BL/10SgSnJ (the background recipient sub-strain used to generate B10.S) mice have been phenotyped for systolic blood pressure, hematocrit and serum albumin levels, we utilized the data for each of these parameters for genetically highly-related strains (C57BL/10, C57BL/6, C57BLKSC57BR, C57L, and C57BR (Petkov et al., 2004). We used only data that were obtained simultaneously with SJL/J mice and closely controlled for age, sex, and time of day for specimen collection. Blood pressure, albumin, and hematocrit in C57BL/10SgSnJ-related strains are significantly different from SJL/J mice (Figure A.2A-C). Compared to B10.S mice, SJL mice also exhibited increased circulating levels of SCLS-related angiogenic proteins Angpt2 and VEGFA (Figure A.2D-E). We therefore hypothesized that
these traits may be due to an underlying susceptibility to vascular leakage and that SJL mice may be useful as an *in vivo* model for SCLS.

**A.3.3. SJL mice exhibit age- and inflammation-dependent hypersensitivity to systemic administration of histamine.**

The SJL strain has been extensively studied, most notably to investigate immune dysregulation (Moriguchi *et al.*, 2018). To our knowledge, however, vascular function in SJL mice has not been queried in detail. To determine whether SJL mice were more susceptible to vascular leakage than B10.S mice, we administered histamine to two distinct SJL sub-strains, SJL/J and SJL/NCr mice, with B10.S mice as a control. Both SJL sub-strains exhibited histamine hypersensitivity and died within 30 minutes of histamine administration, even at the lowest intravenous doses, whereas B10.S mice all survived (Table A.1A). We have designated this phenotype “histamine hypersensitivity (Histh)”.

Similar to the emergence of SCLS symptoms in middle age (Druey *et al.*, 2017), we found that the Histh phenotype in SJL mice was age-dependent; mice greater than 6 months of age all succumbed to histamine administration whereas the younger, 8 week old mice did not (Table A.1A).

In a recent survey of SCLS patients, infectious triggers were identified in 35-50% of disease exacerbations (Eo *et al.*, 2018). We therefore tested whether an inflammatory stimulus potentiates a lethal response to histamine challenge in SJL mice. To study this, we administered complete Freund’s adjuvant (CFA), a complex mixture of antigens and oil widely used to augment immune responses. CFA-induced inflammation had no apparent impact on the responses of aged (>6 months) mice; both primed and un-primed
mice succumbed to intravenous histamine doses at 25 mg/kg and higher, but not to the lowermost dose (12.5 mg/kg). By contrast, the younger (8-week old) SJL mice, which were fully resistant to intravenous histamine alone, exhibited 100% mortality in response to histamine (25 mg/kg and higher doses) if first primed with CFA. CFA-primed B10.S mice of both age groups remained resistant throughout. Furthermore, (B10.S × SJL) F1 hybrid mice phenocopied B10.S mice, demonstrating that Histh is a recessive trait (Table A.1-B). Taken together, these findings suggest that genetically-controlled histamine hypersensitivity can be spontaneous and/or exacerbated by inflammatory stimuli.

A.3.4. Histh, the locus controlling susceptibility to vascular hypersensitivity to histamine, exhibits maximal linkage to mouse chromosome 6

To map the gene or genes controlling Histh, we treated ~478 (B10.S × SJL) F2 mice with histamine at 30 days after priming with CFA and performed genetic association analysis using pre-established genomic markers (Rhodes et al., 1998) (Table A.2). A genome scan using microsatellite markers that distinguish Histh-resistant B10.S and Histh-susceptible SJL mice identified a quantitative trait locus (QTL) on Chr 6 within an approximately ~100Mb region between D6Mit74 (48.72Mb) to D6Mit372 (148.45Mb) \((p=5.73 \times 10^{-5})\). In addition, there were minor linkages to Chr 8 \((p=2.80 \times 10^{-2})\) and Chr 15 \((p=9.74 \times 10^{-4}, \text{data not shown})\). We have designated this locus on Chr 6 as Histh (histamine hypersensitivity) with accession #6360897 in Mouse Genome Informatics (http://www.informatics.jax.org).
A.3.5. Congenic mapping of Histh

We then confirmed the existence and location of Histh on Chr 6 by congenic mapping (Table A.3). We used marker-assisted selection to introgress the Histh interval \((D6Mit74 \ (48.72Mb) \ through \ D6Mit254 \ (125.36Mb))\) from SJL onto the B10.S background. These mice were backcrossed for 12 generations and fixed as a homozygous interval-specific recombinant congenic line (ISCL) hereafter referred to as B10.S-Histh\textsuperscript{SJL}. The Histh phenotype was confirmed by testing susceptibility to histamine challenge 30 days after priming with CFA as above. Indeed, lethality due to Histh differed significantly among the strains \((\chi^2 = 51.61, \ df=1, \ p<0.0001)\); SJL and B10.S-Histh\textsuperscript{SJL} mice were significantly more susceptible to CFA/histamine than were B10.S mice \((\chi^2 =55.24, \ df=1, \ p<0.0001 \text{ for both strains})\), but their responses did not differ significantly from each other. Moreover, \((B10.S \times B10.S-\text{Histh}\textsuperscript{SJL})\) \(F_1\) hybrids were Histh-resistant, confirming the observation made earlier (see findings in Table A.1) regarding Histh as a recessive trait. Thus, we have physically mapped Histh to Chr 6:48-125 Mb and demonstrated that this locus is sufficient to provide full penetrance of the Histh phenotype.

A.3.6. Dermal vasculature of mice harboring an Histh susceptibility allele is hyperresponsive to histamine.

We hypothesized that Histh in mice is due to a genetic predisposition of endothelial cells to exaggerated barrier breakdown in response to permeability mediators. To evaluate histamine-mediated vascular hyperpermeability \textit{in vivo}, we used the well-established Miles assay (Radu \textit{et al.}, 2013), which measures plasma extravasation from cutaneous microvasculature through quantification of Evans blue (EB), an albumin-
binding dye, in skin. We first injected EB intravenously into older (>6 months) B10.S or SJL mice, followed by intradermal injection of histamine and quantification of extravasated EB in skin biopsies. Dermal EB extravasation increased significantly in skin biopsies of histamine-treated SJL v. B10.S mice, and extravasation in both strains appeared to be more extensive than that detected in response to PBS alone (Figure A.3A-B). To determine if the *Histh* locus is associated with histamine-induced vascular leakage, we performed the Miles assay in older B10.S-*Histh*<sup>SJL</sup> congenic mice. Compared with the responses of B10.S mice, B10.S-*Histh*<sup>SJL</sup> congenic mice exhibited a significant increase in dermal EB vascular leakage (Figure A.3C).

Finally, to determine if the increased susceptibility to histamine-induced cutaneous vascular leak in SJL and B10.S-*Histh*<sup>SJL</sup> mice is age-dependent and/or inflammation dependent, we performed Miles assays in 8-week-old mice that were primed with CFA prior to histamine challenge. In the absence of CFA priming, we observed no significant response to histamine in these younger mice (Figure A.3D). By contrast, CFA priming potentiated the vascular hyperpermeability response in the younger 8-week-old B10.S-*Histh*<sup>SJL</sup> congenic mice. B10.S controls exhibited no increase in hypersensitivity to histamine following CFA administration. Together, these results strongly suggest that the *Histh* locus plays a critical role in regulating histamine-induced vascular hyperpermeability, and that this phenotype is affected by both age and pre-existing systemic inflammation.
A.3.7. Susceptibility to histamine-induced systemic vascular leak is genetically controlled by *Histh*.

For unknown reasons, vascular leak in SCLS patients manifests prominently in skin and skeletal muscle, less frequently in gastrointestinal tract and myocardium (Pineton de Chambrun *et al.*, 2017; Druery *et al.*, 2018; Pineton de Chambrun *et al.*, 2018), and rarely in other internal organs including lungs, kidneys, and central nervous system (Druery *et al.*, 2017; Pineton de Chambrun *et al.*, 2017). To determine the extent of vascular leak in individual internal organs in response to histamine, we challenged young CFA-primed SJL, B10.S-*Histh*<sup>SJL</sup> and B10.S mice intravenously with EB followed by systemic (intravenous) administration of histamine or diluent control; EB content was quantified in various organs after 30 minutes. Histamine-mediated vascular leak was detected in skin and skeletal muscle of both SJL and B10.S-*Histh*<sup>SJL</sup> but not B10.S mice compared to PBS-treated counterparts (Figure A.4-A). We detected no dye extravasation in lungs, heart, or gut. We also observed a similar pattern of vascular leakage among older mice (greater than 6 months) following systemic administration of EB and histamine in the absence of CFA priming (Figure A.4-B). These data indicate that the *Histh* locus controls susceptibility to histamine-mediated vascular hyperpermeability with impact in a whole animal model. Moreover, the pattern of vascular leakage is highly reminiscent of that observed in SCLS patients, where skin edema is profound and frequently complicated by extensive rhabdomyolysis requiring fasciotomies (Druery *et al.*, 2017; Pineton de Chambrun *et al.*, 2017).

Given the prominent link between viral upper respiratory tract or other infections and acute SCLS flares, we determined whether acute virus infection, as a common link and physiologic inflammatory stimulus, also elicits vascular leakage in SJL mice. We inoculated SJL and B10.S mice with influenza virus A (H3N2) and assessed systemic vascular leak in correlation with systemic symptoms (i.e. weight loss). A pronounced, 15-20% weight loss was apparent in both strains after 7 days of infection indicating comparable susceptibility to H3N2 (Figure A.5A). However, compared to uninfected controls at day 7 after infection, vascular leakage was increased in H3N2-infected SJL mice but not in B10.S mice (Figure A.5B). In line with the histamine-challenge results, EB extravasation was most prominent in skin, similar to the distribution of fluid extravasation in SCLS. These results demonstrate that a clinically relevant infectious trigger can exacerbate genetically-controlled vascular hyperpermeability and suggest that the SJL mouse recapitulates multiple aspects of SCLS susceptibility, providing a useful and tractable animal model.


The extreme rarity of SCLS has greatly limited our understanding of the complex genetic factors that contribute to disease development. The one published genome-wide association study of SCLS patients has identified 3 SNPs on Chr 3p25.3 ($p \sim 10^{-6}$), with an odds ratio of ~41, as the highest-ranking candidate susceptibility locus (Xie et al., 2013). Using the list of published potential SCLS candidate loci, we sought syntenic mouse
homologous genes that map to the Histh interval (Figure A.6A). We identified 9 human genes that were significant hits in the SCLS GWAS and were also captured in this locus in B10.S-Histh<sup>SJL</sup> congenic mice. Most notably, human 3p25.3 is syntenic with the center of the mouse Histh locus on Chr 6, a region that also demonstrates strong linkage to the Histh phenotype (Table A.1). The shared genetic alignment of Histh with several SCLS-linked genes has provided strong focus on numerous gene candidates that may be involved in the pathogenesis of both Histh and SCLS. We generated protein functional interaction networks (Kramer <i>et al.</i>, 2014) using the list of syntenic genes to define SCLS or its sub-phenotypes and to interrogate potential mechanistic links between predicted SCLS gene candidates and disease (Figure A.6B). This approach identified several genes that are associated with aging (<i>ATP2B2, CAV3, CNTN3, CTNNA2, GRID2</i>), inflammation (<i>ATP2B2, CAV3, RAD18, KBTBD8</i>), vascular permeability (<i>SFXN5, RAD18</i>) and anaphylaxis (<i>CAV3, RAD18, CTNNA2, ATP2B2 and GRID2</i>). In summary, these results suggest that SJL mice and human subjects with SCLS share a similar genetic basis for increased susceptibility to vascular hyperpermeability.
A.4. Discussion

SCLS is a unique, relapsing-remitting disease that can have devastating consequences. Although disorders with features of SCLS have recently emerged in children (Hsu et al., 2015), most patients present in mid-life and lack any family history of the disease. Not unexpectedly, whole exome sequencing (WES) performed on DNA samples from several children with SCLS, their families, and unrelated adults did not uncover any shared single nucleotide variants that could readily explain the phenotype (Pierce et al., 2018). Thus, multiple genetic abnormalities may contribute to SCLS, indicating that our alternative approach of synteny studies may be more appropriate.

Our discovery of a shared susceptibility locus for vascular hyperpermeability in mice has led to unexpectedly strong conclusions about human 3p25 increasing the risk of SCLS in a mechanistic fashion. Specifically, we have characterized the vascular phenotype of the inbred SJL mouse strain, which shares genetic and phenotypic similarities with human patients with SCLS. Because SJL mice recapitulate cardinal features of SCLS, this mouse model may serve to advance our understanding of disease mechanisms. Just as patients with SCLS are typically asymptomatic between episodes, SJL mice do not exhibit overt symptoms of vascular leak at baseline, although deeper investigation revealed that these mice maintain the prototypic SCLS triad of high hematocrit, low serum albumin and hypotension at homeostasis (Figure A.1). Histamine-induced mortality in SJL mice correlated with vascular leakage in skin and skeletal muscle, which are the most prominent sites of pathology in SCLS patients. Both SCLS patients and SJL mice are uniquely
susceptible to inflammation-associated vascular leakage, particularly that induced by systemic infection, such as that resulting from infection with influenza A.

Although we note that SJL mice have monoclonal gamma globulins in serum as do more than 80% of SCLS patients (Druey et al., 2017), this trait is controlled by the mammary tumor virus locus 29 (Mtv29), which encodes for an endogenous superantigen (vSAg29) (Tsiagbe et al., 1990). Our genetic mapping studies exclude its role in increased histamine susceptibility as it falls outside of the Histh locus. Accordingly, no pathogenic role for SCLS paraproteins has been demonstrated thus far in humans (Zhang et al., 1993; Xie et al., 2012). Based on the histamine-induced mortality and vascular leakage findings, we can conclude that the Histh locus controls histamine sensitivity and vascular permeability as a function of age and various inflammatory stimuli (CFA, viral infection). These patterns reflect the pathogenesis of SCLS, notably reflecting the fact that most patients who have spontaneous episodes are middle-aged. Interestingly, SCLS crises in children are nearly always preceded by infection (Hsu et al., 2015).

Most important, the synteny map of the Histh locus and SCLS GWAS has provided strong focus on several gene candidates that may be involved in the pathogenesis of both Histh and SCLS in that they have demonstrated roles in processes involved in vascular endothelial barrier integrity. We further characterized potential functions of several of these genes in SCLS by searching PubMed and the International Mouse Phenotyping Consortium (http://www.mousephenotype.org) for associations with processes involved in vascular barrier integrity. Among these, ATP2B2/Atp2b2 encodes ATPase plasma membrane Ca\(^{2+}\) transporting 2 protein, which plays a critical role in
intracellular calcium homeostasis and endothelial cell responses to histamine (Worthen et al., 2000). Atp2b2 also regulates endothelial nitric oxide (NO) synthase (eNOS) phosphorylation in endothelial cells (Holton et al., 2010), a critical step in histamine- and VEGF-induced vascular permeability previously implicated in SCLS-associated vascular dysregulation (Di Lorenzo et al., 2013; Umbrello et al., 2014). Notably, Atp2b2−/− mice have significantly reduced serum albumin levels at homeostasis compared with WT controls (Dickinson et al., 2016). Cav3 encodes caveolin 3, a protein also implicated in eNOS regulation (Sun et al., 2015). Although Cav3 knockout mice develop heart failure due to myocardial fibrosis and dilated cardiomyopathy (Bryant et al., 2018), functions of Cav3 in the peripheral vasculature have not been studied in mice. CTNNA2/Ctnna2 encodes α-catenin 2, which functions as a linker between cadherin adhesion receptors and the cytoskeleton, and thereby regulates cell-cell adhesion dynamically in response to histamine (Kugelmann et al., 2018). However, since genome-wide knockout of Ctnna2 in mice is associated with neonatal lethality, with most homozygotes dying with 24 hours after birth (Abe et al., 2004), vascular-specific deletion of Ctnna2 may be necessary to interrogate its potential role in Histh and SCLS. RAD18/Rad18 is a E3 ubiquitin-protein ligase involved in post replication repair of UV-damaged DNA. Notably, Rad18 knockout mice exhibit higher Hct at homeostasis (Tateishi et al., 2000). Finally, CNTN3/Cntn3 encodes contactin 3, an activator of the small GTPase Arf6 that has been linked to inflammation-triggered vascular permeability (Zhu et al., 2012). Cntn3−/− mice exhibit normal physiological levels of serum albumin and reduced Hct compared to WT controls (Dickinson et al., 2016).
Because several of the genes captured in Histh map to chromosomal locations that are outside of the top region of synteny on Chr 3 in humans (Figure A.6B), there may be additional causal loci for SCLS. Indeed, because there were 134 separate loci significantly associated with SCLS in our original GWAS (Xie et al., 2013), we focused on the strongest genetic signal from the SCLS GWAS that allowed us to interrogate candidates that overlap the Histh locus. Even within Histh, the true causal variants for SCLS may lie in strong linkage with genotyped SNP. In future studies, it will be critical to further dissect the Histh QTL by generating interval specific recombinant sub-congenic lines (Bennett et al., 2002) and/or by using mouse GWAS (Flint et al., 2012) in order to fine map and validate the true causative genes. Since the rarity of SCLS in humans poses a large barrier to mapping the causative gene(s), our new mouse genetic model provides an important alternative approach.

The present results advance the field by showing that SCLS patients are hyper-responsive to mediators of vascular permeability, suggesting that aberrant endothelial function contributes directly to clinical symptoms. Although SCLS patients are typically asymptomatic between episodes, dermal microvascular endothelial cells isolated from a patient with fatal SCLS were persistently hyper-responsive to inflammatory mediators including LPS, TNFα, and IL-1β in vitro (Pierce et al., 2017). Previous histological studies of skin and muscle of SCLS patients have failed to uncover gross structural or ultrastructural abnormalities within the microvasculature that could account for this phenotype (Druey et al., 2010). Thus, our results demonstrating that the cutaneous vasculature of SCLS patients is hyper-responsive to inducers of permeability supports the
hypothesis that the acute manifestations of SCLS result from the exaggerated functional responses of a susceptible host to otherwise common inflammatory triggers in a fashion attributable to underlying genetic defects within the endothelium, resulting in an accelerated breakdown of vascular barrier function. As such, future studies including assessment of expression and sequence of top Histh candidate genes in SCLS patients and mice and their role in endothelial responses to inflammation will be essential to determine their contribution to these phenotypes.
A.5. Material and Methods

A.5.1. Patients and skin testing.

Patients were seen at the Clinical Center of the National Institutes of Health under an IRB-approved study protocol (I-0184) after informed consent. Histamine phosphate or morphine sulfate were injected intradermally at separate sites along the dorsal aspect of the upper arm. After 15-20 minutes, the size of wheals was determined manually and analyzed using ImageJ.

A.5.2. Animals.

B10.S-Histh\textsuperscript{SJL}, (B10.S × SJL/J)\textsubscript{F\textsuperscript{1}} and (B10.S × SJL/J) × B10.S F\textsubscript{2} mice were generated and maintained under specific pathogen free conditions in the vivarium of the Given Medical Building at the University of Vermont according to National Institutes of Health guidelines. All animal studies were approved by the Institutional Animal Care and Use Committee of the University of Vermont or the NIAID/NIH (animal study protocol LAD3E).

A.5.3. Histh Phenotyping.

Cohorts of 4 male and 4 female mice were used in each histamine challenge. Each mouse was sensitized (D0 and D7) by subcutaneous injection with a 50/50 mix of complete Freund’s adjuvant (CFA) and PBS or left unmanipulated. 30 days later histamine hypersensitivity was determined by intravenous (i.v.) injection of 100, 50, 25 and 6.25 mg/kg histamine (dry weight free base) diluted in PBS. Deaths were recorded at 30 minutes post injection and the data reported as the number of dead mice over the number of mice.
in the study. 6-month old/aged animals did not receive any CFA priming prior to histamine challenge.

**A.5.4. DNA extraction and genotyping.**

DNA was isolated from mouse tail clippings as previously described (Sudweeks et al., 1993). Briefly, individual tail clippings were incubated with cell lysis buffer (125 µg/ml proteinase K, 100 mM NaCl, 10 mM Tris-HCl (pH 8.3), 10 mM EDTA, 100 mM KCl, 0.50% SDS, 300 µl) overnight at 55°C. The next day, 6M NaCl (150 µl) was added followed by centrifugation for 10 min. at 4°C. The supernatant layer was transferred to a fresh tube containing 300 µl isopropanol. After centrifuging for 2 minutes, the supernatant was discarded and the pellet washed with 70% ethanol. After a final 2 min. centrifugation, the supernatant was discarded, and DNA was air dried and resuspended in TE. Genotyping was performed by using established microsatellite markers (Butterfield et al., 1998). Polymorphic microsatellites were selected to have a minimum polymorphism of 8 bp for optimal identification by agarose gel electrophoresis. Briefly, primers were synthesized by IDT-DNA (Coralville, IA) and diluted to a concentration of 10 µM. PCR amplification was performed using Promega GoTaq according standard methods and amplicons were subjected to 2% agarose gel electrophoresis and visualized by ethidium bromide and UV light.

**A.5.5. Data resources in the Mouse Phenome Database.**

Phenotype data for systolic blood pressure (MPD#23602), hematocrit (MPD#31825) and albumin (MPD#46001) were queried using the Mouse Phenome Database ([https://phenome.jax.org/](https://phenome.jax.org/)) for laboratory inbred strains. The significance of the
observed differences for each trait was determined using the Mann-Whitney U test comparing SJL/J against the mean trait variables for all strains studied.

A.5.6. Miles assay.

To assess histamine-induced vascular leak in mice, we used the Miles assay as described previously (Radu et al., 2013). Briefly, mice were injected intraperitoneally with pyrilamine maleate (4 mg per kg body weight, Sigma) 30 minutes prior to injection with Evans blue dye to reduce background permeability during handling. Mice were then injected with 100 µl of 0.5% Evans blue dye in PBS (Sigma) via retro-orbital injection, followed by intradermal injections of histamine or saline (50 µl total volume). 30 minutes after the intradermal injection, the dorsal skin was collected with a 12-mm biopsy punch, and Evans blue (EB) dye was extracted with formamide (Sigma; 56°C for 48 hrs). The amount of EB in each sample was determined by measuring the absorbance at 620 nm, and results were expressed as EB dye amount (ng) per 100 mm² of the skin, with quantification against a standard curve.

A.5.7. Influenza virus infection.

Mice were anesthetized with isoflurane and influenza A/HK/1/68 (H3N2) virus was administered intranasally in a total volume 3 µl in each nare (6 × 10⁴ cfu/ml). Weight loss was monitored, and mice were sacrificed on D8 post-inoculation for analysis of vascular leakage.
A.5.8. Systemic vascular leak analysis.

To assess influenza-mediated vascular leak in various tissues/organs, we injected mice intravenously with 100 µl of 2% EB in PBS retro-orbitally. Fifteen minutes post-injection, the mice were deeply anesthetized by isoflurane inhalation and perfused with 5 ml of heparinized PBS through the left ventricle of the heart to remove the EB remaining in the vascular space. Tissues were heated at 95°C for one hour to obtain dry weights. The amounts of Evans blue dye (ng) were quantified as described above and normalized by dry weights of individual tissues (mg). Results were expressed as fold change compared to corresponding controls (PBS inoculated mice). To analyze histamine-mediated systemic vascular leak, same procedures were performed except that mice were injected with Evans blue dye immediately following intraperitoneal injection with 100 µl of histamine in PBS (12.5 mg per kg body weight).

A.5.9. Linkage analysis and generation of Histh congenic mice.

Segregation of genotype frequency differences with susceptibility and resistance to Histh in (B10.S × SJL/J) × B10.S were tested by Chi-square (χ2) analysis in 2 × 2 contingency tables. B10.S-HisthSJL congenic mice were derived by marker-assisted selection of SJL/J derived alleles and successive backcrossing to B10.S mice.

A.5.10. Synteny mapping between Histh locus and SCLS GWAS candidates.

The list of significant SNPs from SCLS GWAS (Xie et al., 2013) were retrieved and annotated to the corresponding human genes using SNP Nexus tool (Dayem Ullah et al., 2018). The orthologous mouse genes and their genomic coordinates were retrieved.
using MGI-gene database at Jackson Labs (http://www.informatics.jax.org/batch). Genes that mapped outside the Histh interval were excluded from subsequent analysis.
A.6. Acknowledgements

Funding & Disclaimer: Funding for this study was provided in part by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health (project numbers Z01-AI-000746; Z01-AI-000943). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.
A.7. References


198


A.8. Figures

Figure A.1. SCLS patients are hyper-responsive to histamine and morphine.

(A-B) SCLS patients (n=16) or healthy controls (n=7) were injected intradermally with the indicated concentrations of histamine (A) or morphine (B). Wheal sizes were determined using ImageJ. ****p<0.00001, 2-way ANOVA.
Figure A.2. SJL mice phenocopy SCLS.

(A-C) Phenotype data was obtained from mouse phenome database (https://phenome.jax.org/) for systolic blood pressure (A, MPD#23602), hematocrit (B, MPD#31825) and albumin (C, MPD#46001). Values for SJL mice (bold) were compared each mean trait variable for all strains (***p<0.001, ****p< 0.0001, Mann-Whitney U test). (D-E) Serum levels of VEGFA (D) and Angpt2 (E) in SJL and B10.S mice; (****p<0.0001, Mann-Whitney).
Figure A.3. Dermal vasculature of mice containing the identified Histh allele is hyper-responsive to histamine.

(A-D) SJL, B10.S or B10.S-HisthSJL congenic mice were injected with Evans Blue (EB) dye i.v. followed by intradermal challenge with histamine for 30 min. (A) Skin biopsies from aged (>6 months) mice after intradermal treatment with histamine (625 ng/mouse) or saline. (B-C) Quantification of EB extravasation in skin biopsies of B10.S, SJL, or B10.S-HisthSJL congenic mice. Each symbol represents one mouse; **p<0.003; ****p=0.00005; Holm-Sidak corrected t test. (D) Young (8-week-old) mice were primed with CFA or left untreated (ctrl) prior to histamine challenge. ***p=0.0006; ****p<0.0001, 2-way ANOVA, Tukey multiple comparisons v. B10.S mice (n=3-5 mice/group); ns= not significant.
Figure A.4. Systemic administration of histamine induces vascular leak in SJL and B10.S-HisthSJL but not B10.S mice.

(A-B) EB was injected i.v. followed by i.p. injection of histamine (12.5 mg/kg) in either young (8-week-old) mice primed with CFA (A) or aged mice (>6 months of age) (B). Extravasated dye was normalized to dry weight of the tissue/organ and expressed as fold change compared with controls (PBS). Each symbol represents an individual mouse; 2 independent experiments were performed; *p<0.03, **p=0.001; ***p=0.0004, Holm-Sidak corrected t test.
Figure A.5. Influenza-associated vascular leak in SJL mice as a means to model SCLS.

SJL and B10.S mice were infected with influenza virus A/H3N2. (A) Weights were determined D0 and D7 post-infection (*p=0.01, ****p<0.00001, Holm-Sidak corrected t test. (B) EB dye extravasation was evaluated at D8 post infection; symbol represents an individual animal; 2-4 separate experiments were performed. ***p=0.0002, Holm-Sidak corrected t test.
A Rare diseases (Austin, Tex.), 1(1), e27445. doi:10.4181/rdis.27445
Figure A.6. Shared genetic and phenotypic alignment between Histh and SCLS.

(A) Of the 134 genetic loci implicated in SCLS (7), 9 (shown in red) were found to overlap with Histh locus on mouse Chr 6 (shown in orange). (B) Protein functional interaction networks for aging, inflammation, vascular permeability and anaphylaxis were generated using Ingenuity Pathway Analysis. These networks were used to assess biological interactions with 9 predicted SCLS candidates. Dotted lines represent an interaction. Each sub-phenotype and associated loci are color coded. Aging=pink; Inflammation=green; Vascular permeability=yellow and Anaphylaxis=orange
A.9 Tables

Table A.1. SJL mice exhibit age and/or inflammation-dependent histamine hypersensitivity (Histh).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Histamine (mg/kg)</th>
<th>aged</th>
<th>8 week</th>
<th>Strain</th>
<th>Histamine (mg/kg)</th>
<th>CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SJL/J</td>
<td>100</td>
<td>4/4</td>
<td>0/4</td>
<td>SJL/J</td>
<td>100</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4/4</td>
<td>0/4</td>
<td></td>
<td>50</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>2/4</td>
<td>0/4</td>
<td></td>
<td>25</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>12.5</td>
<td>0/4</td>
</tr>
<tr>
<td>SJL/NCr</td>
<td>100</td>
<td>4/4</td>
<td>0/4</td>
<td>SJL/NCr</td>
<td>100</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4/4</td>
<td>0/4</td>
<td></td>
<td>50</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>2/4</td>
<td>0/4</td>
<td></td>
<td>25</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>12.5</td>
<td>0/4</td>
</tr>
<tr>
<td>B10.S/SgMcDJ</td>
<td>100</td>
<td>0/4</td>
<td>0/4</td>
<td>B10.S/SgMcDJ</td>
<td>100</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>50</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>25</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>12.5</td>
<td>0/4</td>
</tr>
<tr>
<td>(B10.S × SJL) F₁</td>
<td>100</td>
<td>0/4</td>
<td>0/4</td>
<td>(B10.S × SJL) F₁</td>
<td>100</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>50</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>25</td>
<td>0/4</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>0/4</td>
<td>0/4</td>
<td></td>
<td>12.5</td>
<td>0/4</td>
</tr>
</tbody>
</table>

Young (8-10 week-old) or aged (>6 months) mice were left untreated or pretreated with CFA by intraperitoneal (i.p.) injection and challenged 30 days later with the indicated doses of histamine (mg/kg) by i.v. injection. Deaths were recorded at t=30 min. Results are expressed as the # of deaths/total mice. CFA=complete Freund's adjuvant.
Table A.2. Histamine sensitivity (Histh) maps to mouse chromosome 6.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Location</th>
<th>Dead</th>
<th>Alive</th>
<th>$\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B10.S Ho</td>
<td>B10.S het</td>
<td>SJL Ho</td>
<td>SJL het</td>
</tr>
<tr>
<td>D6Mit116</td>
<td>25150229</td>
<td>29</td>
<td>36</td>
<td>53</td>
<td>90</td>
</tr>
<tr>
<td>D6Mit74</td>
<td>48726556</td>
<td>21</td>
<td>47</td>
<td>48</td>
<td>111</td>
</tr>
<tr>
<td>D6Mit17</td>
<td>71119218</td>
<td>16</td>
<td>56</td>
<td>45</td>
<td>116</td>
</tr>
<tr>
<td>D6Mit8</td>
<td>83713869</td>
<td>12</td>
<td>34</td>
<td>54</td>
<td>91</td>
</tr>
<tr>
<td>D6Mit178</td>
<td>94225829</td>
<td>13</td>
<td>31</td>
<td>56</td>
<td>94</td>
</tr>
<tr>
<td>D6Mit36</td>
<td>104503360</td>
<td>8</td>
<td>62</td>
<td>51</td>
<td>124</td>
</tr>
<tr>
<td>D6Mit54</td>
<td>112269957</td>
<td>5</td>
<td>38</td>
<td>55</td>
<td>89</td>
</tr>
<tr>
<td>D6Mit366</td>
<td>115242853</td>
<td>6</td>
<td>38</td>
<td>56</td>
<td>88</td>
</tr>
<tr>
<td>D6Mit216</td>
<td>121115242</td>
<td>9</td>
<td>65</td>
<td>41</td>
<td>120</td>
</tr>
<tr>
<td>D6Mit254</td>
<td>125356646</td>
<td>11</td>
<td>65</td>
<td>42</td>
<td>127</td>
</tr>
<tr>
<td>D6Mit59</td>
<td>138976326</td>
<td>10</td>
<td>58</td>
<td>49</td>
<td>124</td>
</tr>
<tr>
<td>D6Mit372</td>
<td>148450482</td>
<td>8</td>
<td>62</td>
<td>47</td>
<td>115</td>
</tr>
</tbody>
</table>

(B10.S x SJL) F$_2$ mice were genotyped using microsatellite markers, and phenotyped for Histh: histamine sensitivity was determined by i.v. injection of 1.0mg of histamine free base in 0.2 ml of PBS 30 days post CFA injection. Deaths were recorded t=30 min. after histamine challenge.
Table A.3. Congenic mapping confirms the existence and location of the *Histh* locus.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Marker/Location(bp)</th>
<th>Histh</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJL/J</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit74, 4872656-48726705</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit177, 71119218-71119467</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit517, 94225829-94225955</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit54, 112269957-112270141</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit178, 125356646-125356785</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit254, 148450482-148450593</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D6Mit132, 148450482-148450593</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CFA</td>
<td>aged</td>
</tr>
<tr>
<td>SJL/J</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>B10.S</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>B10.S-<em>Histh</em>&lt;sub&gt;SJL&lt;/sub&gt;</td>
<td>B</td>
<td>S</td>
</tr>
<tr>
<td>(B10.S x B10.S-<em>Histh</em>&lt;sub&gt;SJL&lt;/sub&gt;)&lt;sub&gt;F&lt;sub&gt;1&lt;/sub&gt;&lt;/sub&gt;</td>
<td>B</td>
<td>B/S</td>
</tr>
</tbody>
</table>
| Cohorts of young (8 to 10-week old) mice pre-conditioned with CFA, or aged mice (>6 months) left untreated were challenged 30 days later with histamine (50-100 mg/kg) by i.v. injection. Deaths were recorded at t=30 min. Results in the two right columns indicate the # of animals dead/total mice. ND= not done; CFA=complete Freund’s adjuvant.
APPENDIX B: NETWORK-BASED FUNCTIONAL PREDICTION AUGMENTS GENETIC ASSOCIATION TO PREDICT CANDIDATE GENES FOR HISTAMINE HYPERSENSITIVITY IN MICE.

Anna L. Tyler¹, Abbas Raza², Dimitry N. Krementsov⁴, Laure K. Case¹, Elizabeth P. Blankenhorn⁷, Cory Teuscher²,³, Runlin Ma⁸ and J. Matthew Mahoney⁵,⁶

¹The Jackson Laboratory, Bar Harbor, ME, 04609, USA; ²Department of Medicine, ³Department of Pathology, ⁴Department of Biomedical and Health Sciences, ⁵Department of Neurol Sci, ⁶Department of Computer Science, University of Vermont, Burlington, VT 05405, USA; ⁷Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, 19129, USA; ⁸Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China.

Address correspondence to: John M. Mahoney, 95 Carrigan Drive, Stafford 118, University of Vermont, Burlington, VT 05405; John.M.Mahoney@uvm.edu
B.1. Abstract

Genetic mapping is a primary tool of genetics in model organisms; however, many quantitative trait loci (QTL) contain tens or hundreds of positional candidate genes. Prioritizing these genes for validation is often ad hoc and biased by previous findings. Here we present a technique for computationally prioritizing positional candidates based on computationally inferred gene function. Our method uses machine learning with functional genomic networks, whose links encode functional associations among genes, to identify network-based signatures of functional association to a trait of interest. We demonstrate the method by functionally ranking positional candidates in a large locus on mouse Chr 6 (45.9 Mb to 127.8 Mb) associated with histamine hypersensitivity (Histh). Histh is characterized by systemic vascular leakage and edema in response to histamine challenge, which can lead to multiple organ failure and death. Although Histh risk is strongly influenced by genetics, little is known about its underlying molecular or genetic causes, due to genetic and physiological complexity of the trait. To dissect this complexity, we ranked genes in the Histh locus by predicting functional association with multiple Histh-related processes. We integrated these predictions with single nucleotide polymorphism (SNP) association data derived from a survey of 23 inbred mouse strains. The top-ranked genes included Cxcl12, Ret, Cacna1c, and Ctnm3, all of which had strong functional associations and were proximal to SNPs segregating with Histh. These results demonstrate the power of network-based computational methods to nominate highly plausible quantitative trait genes even in highly challenging cases involving large QTLs and extreme trait complexity.
B.2. Introduction

Identifying causal variants within quantitative trait loci (QTLs) is a central problem of genetics, but genetic linkage often prevents narrowing QTLs to less than several megabases (Mb). Thus, QTLs may contain hundreds of candidate genes. Instead of revealing the exact gene (or genes) responsible for trait variation, QTL mapping produces positional candidate genes. Rigorously narrowing a QTL by fine mapping with congenic strains can take years or decades, particularly in organisms such as mice that have long generation times. Moreover, high-resolution congenic mapping often reveals that the overall QTL effect is due to multiple linked genes within the QTL rather than a single gene (Yazbek et al., 2011; Parker et al., 2013). Thus, positional data alone are generally insufficient to nominate candidate genes for subsequent biological follow up. To overcome the limitations of mapping data, researchers look within a QTL for plausible candidate genes. However, these selections are typically done by ad hoc criteria using prior knowledge or a literature search. This strategy is strongly biased toward prior knowledge and is highly error prone due to missing annotations. There is a need for rigorous and systematic strategies to distinguish among positional candidate genes for mechanistic follow up.

We developed a novel approach to rank positional candidates based on functional association with a trait. To avoid annotation or literature bias, we use functional genomic networks (FGNs), which encode predicted functional associations among all genes in the genome. FGNs such as the Functional Networks of Tissues in Mouse (FNTM) (Goya et al., 2015) and HumanBase (Greene et al., 2015), are Bayesian integration networks that
combine gene co expression, protein-protein binding data, ontology annotation and other data to predict functional associations among genes. With these networks we can expand on known gene-trait associations to identify sub-networks of trait-associated genes that include novel genes, including in the QTL of interest.

Recent studies with functional genomic networks, for example FNTM, have demonstrated their power to associate novel genes with specific phenotype terms (Guan et al., 2010) or biological processes (Ju et al., 2013). For example, Guan et al. used a support vector machine (SVM) classifier to identify a gene network associated with bone mineralization and made validated predictions of novel genes that lay outside of all published QTLs for bone mineralization phenotypes (Guan et al., 2010). Subsequent studies using similar network-based techniques have made novel predictions of hypertension- and autism-associated genes (Greene et al., 2015; Krishnan et al., 2016). We have expanded these methods to rank genes in a mapped QTL based on multiple putative functional terms and to integrate these rankings with genetic association p values from strain surveys. Our method produces a final ranked list for all genes in the QTL that incorporates both the functional and positional scores of each candidate gene.

Our strategy first builds trait-associated gene lists from structured biological ontologies (e.g., the Gene Ontology (Acencio et al., 2019) and the Mammalian Phenotype Ontology) (Smith et al., 2012) and public transcriptomic data from the Gene expression Omnibus (GEO) (Edgar et al., 2002; Barrett et al., 2013). We then applied machine learning classifiers to the functional networks of tissues in mice (FNTM) (Goya et al., 2015) to identify network-based signatures of the trait-related gene lists. This strategy
allows us to predict gene-trait associations that are not currently annotated within a structured ontology, overcoming the missing annotation problem.

We applied our approach to a large QTL associated with histamine hypersensitivity (Histh) in mice. Histh in mice is a lethal response to a histamine injection. In insensitive mice, a histamine injection produces an inflammatory response that resolves without further treatment. Mice with the Histh response develop excitation and ear blanching, followed by progressive respiratory distress, vasodilation, anaphylactic shock, and death (Vaz et al., 1977; Wang et al., 2014). Histh can be induced in a subset of mouse strains by sensitization with Complete Freund’s Adjuvant (CFA). Histh also develops spontaneously in SJL/J animals older than six months of age.

We previously mapped Histh to a locus on Chr 6 (45.9Mb to 127.8 Mb; the Histh locus), which was confirmed using a congenic line (B10.S-HisthSJL) (Raza et al., 2019). Because of the large size of this locus, additional information is required to identify causal variants. To narrow down candidates, we integrated novel genetic association data from interval-specific congenic recombinant lines (ISCRLs) and an inbred strain survey with our network-based functional predictions of Histh-related genes. By augmenting positional data with functional predictions, we dramatically reduced the candidate gene list to a tractable set of high-quality candidates that are implicated in Histh-related processes.
B.3. Results

B.3.1. Generation of Interval Specific Recombinant Congenic Lines (ISRCL) across the Histh locus.

In prior work, we mapped the genetic locus regulating susceptibility to age- and/or inflammation (CFA)-dependent sensitivity to histamine on Chr 6 in SJL mice (Raza et al., 2019). The B10.S-HisthSJL congenic mice exhibit Histh and carry a large ~ 83 Mb region of SJL between 45.9 Mb to 127.8 Mb on the resistant B10.S background (Raza et al., 2019). This large QTL includes 628 protein coding genes. To narrow this region, we generated five ISRCLs using B10.S-HisthSJL x B10.S backcross mice and assessed their susceptibility to Histh (Figure B.1). Under an additive model, these data suggest that Histh is composed of four sub-QTL which we have designated Histh1, Histh2, Histh3, and Histh4, each contributing 17%, 19%, 14% and 10%, respectively, to the overall penetrance. The sub-QTLs are assigned # 6362992, 6362994, 6362996 and 6362997 accession numbers in Mouse Genome Informatics (MGI, www.informatics.org). Importantly, for each sub-QTL this makes positional candidate gene identification using interactive high-resolution congenic mapping impractical.

B.3.2. Inbred strain survey of Histh.

To investigate whether the Histh phenotype is unique to SJL, we assessed histamine responses for 23 inbred mouse strains (including SJL and B10.S; Table B.1). These strains were chosen using haplotype structure across the Histh interval to identify additional mouse strains that are likely to share a susceptible Histh allele (data not shown). 129X1/SvJ, ALR/LtJ, BPN/3J, FVB/NJ, NOD/ShiLtJ, NU/J, SJL/BmJ and SWR/J mice
were identified as having similar haplotype structure as SJL at the Histh locus. ALR/LtJ and SJL/BmJ mice required embryo recovery and were therefore not included. Histh phenotyping identified FVB/NJ, SWR/J, and NU/J mice as Histh-susceptible, whereas 129/X1/SvJ, NOD/ShiLtJ, and BPN/3J were resistant. Taken together with our earlier data, these results indicate that Histh susceptibility segregates among a unique subset of SJL/J-related strains (Petkov et al., 2004).

**B.3.3. Targeted genetic association analysis for Histh.**

Our result from previous linkage analysis (Raza et al., 2019) and congenic mapping localized Histh to an ~83 Mb region on Chr 6 between 45.9 Mb to 127.8 Mb. Given that Histh susceptibility is restricted to a unique subset of inbred strains, particularly the closely related SJL/J, FVB/NJ, and SWR/J, we performed a targeted association analysis (Benson et al., 2017) between SNPs in the Histh locus across all 23 inbred strains. We tested the association of 13,598 SNPs across the Histh locus using efficient mixed-model association (EMMA) (Kang et al., 2008). A total of 84 SNPs in 23 genes showed significant associations ($p<3.68\times10^{-6}$) (Figure B.2). The majority of the significant hits were intronic (71%), non-coding (12%), intergenic (4%) or regulatory (5%) variants. Interestingly, there was overlap between three of the four Histh sub-QTLs (Figure B.2) and SNP-association peaks.

**B.3.4. Network-based prediction of Histh-associated genes.**

To predict functional candidates among the positional candidates in the Histh locus, we delineated a list of Histh-associated biological processes and trained machine learning classifiers to identify subnetworks of functional genomic networks associated with
each of these processes. An overview of our workflow is in Figure B.3. We first defined
gene sets that were related to seven terms that are functionally related to the Histh
phenotype.

The terms and their justifications are as follows:

- **Type I hypersensitivity/Anaphylaxis**: The death response following systemic
  histamine challenge exhibits symptoms of type I hypersensitivity/anaphylaxis
  including respiratory distress, vasodilation, and anaphylactic shock (Vaz et al.,
  1977).
- **Cardiac**: There is evidence suggesting that anaphylactic shock in mice is
  associated with decreased cardiac output, rather than solely a function of systemic
  vasodilation (Wang et al., 2014).
- **Histamine**: Histh is elicited by a systemic histamine challenge (Raza et al., 2019).
- **G-protein coupled receptor**: Histamine receptor H1 (Hrh1) signaling is required
  for the Histh phenotype, and all histamine receptors belong to the family of G-
  protein coupled receptors (Hill et al. 1997).
- **Aging**: Spontaneous Histh develops after six months of age in sensitive mouse
  strains (Raza et al., 2019).
- **Inflammation**: Treatment with pro-inflammatory CFA induces Histh in sensitive
  mouse strains.
- **Tuberculosis**: Histh is induced in some mouse strains by CFA, which contains
  inactivated *Mycobacterium tuberculosis* (Raza et al., 2019).
• Vascular permeability: The Histh response includes vascular leakage in skin and skeletal muscles as assessed by Miles’ assay (Raza et al., 2019).

We used GeneWeaver, the Gene Expression Omnibus (GEO), and PubMed to retrieve gene sets associated with each of these terms (see Materials and Methods). The gene sets ranged in size from 651 to 1466 genes. Because Guan et al. (Guan et al., 2010) found that SVMs trained on gene sets with around 300 genes performed best for network-based functional prediction, we clustered large gene sets into modules of approximately 300 genes and analyzed each module separately (see Materials and Methods). Multiple members of these gene sets are encoded in the Histh locus. For example, e.g. Hrh1 was a member of the Anaphylaxis gene set. To reduce bias in classification, we removed all such genes from each gene set before SVM training. We then trained an ensemble of 100 SVMs on each module gene set. We calculated ROC curves for each model to quantify the ability of each set of SVMs to distinguish genes inside the module gene set from all genes outside the module gene set. AUCs ranged from 0.9 to 0.975 indicating that the SVMs were able to classify the genes in each list robustly. In other words, each gene set used to define a putative Histh-related process forms a distinct subnetwork of the full functional genomic network.

We then applied the trained SVM models to the positional candidate genes in the Histh locus. By classifying each positional candidate, we can identify genes that are likely to be functionally associated with each module gene set. For example, for the Anaphylaxis module gene set, the histamine receptor Hrh1 received a positive score indicating that the SVMs predicted it belonging to the Anaphylaxis gene set despite its absence from the
training set. This example provides a positive control and shows that the SVMs identify biologically relevant patterns in the functional genomic network. In addition to the SVM score, we calculated a false positive rate (FPR) for each gene (see Materials and Methods). Low FPRs indicate high confidence in the classification.

**B.3.5. Integration of functional enrichment with genetic association.**

Genes that are predicted to be highly functionally related to the trait may not have functionally variant alleles in the study population and may therefore be unlikely to drive the observed strain differences in Histh. To identify genes that were likely to have functionally relevant polymorphisms, we integrated functional scores with SNP association p values to focus only on those candidates that satisfied both criteria. By plotting the maximum functional score for a gene, \(-\log_{10}(FPR_{SVM})\) versus the \(-\log_{10}(p_{EMMA})\) (normalized to the max values; see Materials and Methods), we can identify genes that were predicted to be both highly functionally related to Histh phenotype and likely to have functional polymorphisms that segregated with Histh susceptibility (Figure B.4). The blue line in Figure B.4 traces along the Pareto front of the gene set in this space. For any gene on this line, finding a gene with a stronger functional association means finding a gene with lower SNP p value, and *vice versa*. The genes near the Pareto front have either segregating polymorphisms or are predicted to be functionally related to Histh, or both. All such genes are potentially good candidates for experimental follow-up.

To rank the candidates with a single score, we defined a final gene score \((S_{cg})\) for each gene, which is the sum of the (normalized) \(-\log_{10}(FPR_{SVM})\) and the \(-\log_{10}(p_{EMMA})\) (Figure B.5). This score prioritizes candidates in the upper right quadrant with
simultaneously high positional and functional scores. The genes in the upper right quadrant—\textit{Cxcl12, Ret} and \textit{Cacna1c}—had near-maximal scores along both axes and were therefore ranked as the best candidates for follow-up.

In addition to identifying the top-ranked gene over the full \textit{Histh} locus, we identified a top-ranked gene for each sub-QTL identified through congenic mapping. Figure B.5A shows the functional associations across all modules of the top 20 genes ordered by final gene score ($S_{cg}$).
B.4. Discussion

In this analysis, we identified a small set of positional candidate genes in a large locus by combining computational predictions of functional association with Histh and SNP associations. The final list of genes is highly plausible and can be followed up relatively easily with modern genetic editing techniques.

High-quality candidates for Histh.

Three genes in the final ranked list deserve attention: Cxcl12, Ret, and Cacna1c. Of all genes in the locus, these three lie on the Pareto front with both low genetic association p values and high functional scores (Figure B.4). The top-ranked gene, Cxcl12 (a.k.a. stromal cell-derived factor 1), is chemotactic for mast cells via the chemokine receptor Cxcr4 (Ghannadan et al., 2002). Mast cells are major drivers of pathological events in anaphylaxis (Lieberman et al., 2016), demonstrating that the final predictions are highly relevant to Histh. The second-ranked gene Ret encodes a pleiotropic tyrosine protein kinase involved in cell differentiation, growth, migration, and survival (Motenko et al., 2015), inflammation (Rusmini et al., 2013) and the development of the cardiovascular system (Hiltunen et al., 2000). Alleles of this gene could conceivably modify multiple processes underlying Histh, including the both the anatomical background susceptible to Histh and the acute response to histamine. Ret was significantly associated with multiple functional gene sets (Figure B.5A). The third-ranked gene, Cacna1c, encodes the voltage-dependent calcium channel Cav1.2, which is expressed in the heart, muscle, and endocrine glands (Mouse Genome Informatics Mouse Genome Informatics Web Site). Mutations in Cacna1c are associated with electrophysiological alterations in the heart (Napolitano et al.,
suggested a possible role for Cacna1c in impaired cardiac function in Histh. Interestingly, SNPs in human CACNA1C were recently associated with chronic spontaneous urticaria (i.e., spontaneous episodes of hives and/or angioedema) and antihistamine drug response (Yan et al., 2018). These results suggest a direct connection between Cacna1c and anaphylactic or hypovolemic shock.

All of the above genes lie in the Histh4 locus, which accounts for only a portion of the total variation in the Histh phenotype. In the Histh3 locus, the highest-ranked candidate gene was Cntn3, which encodes for contactin 3, an activator protein of the small GTPase Arf. Cntn3 is a member of the contactin family of immunoglobulins. Genetic variants of human CNTN3 are associated with an enlargement of the aorta, acute heart rate recovery, and abdominal aortic aneurysm, suggesting a potential connection to impaired cardiac function during histamine challenge (Elmore et al., 2009). Intriguingly, CNTN3 is near a segregating SNP for Systemic Capillary Leak Syndrome (SCLS) from a human GWAS. SCLS is an extremely rare disease characterized by transient but potentially lethal episodes of diffuse vascular leakage of proteins and fluids into peripheral tissues, resulting in massive whole-body edema and hypotensive shock. The pathological mechanisms and genetic basis for SCLS remain elusive (Xie et al., 2013), but SCLS shares many phenotypic properties with Histh in mice. In particular, SCLS attacks are diagnosed based on the clinical triad of hypotension, elevated hematocrit, and hypoalbuminemia, all of which naturally occur in the Histh-sensitive SJL mouse strain (Raza et al., 2019). The potential association between CNTN3 and SCLS, therefore, lends credence to its possible functional role in Histh as well. Indeed, CNTN3 was not only a positional candidate in the SCLS
GWAS but was contained within functional terms that were enriched among the top positional candidate genes (cf. Table 5 of Xie et al. (Xie et al., 2013), indicating that CNTN3 functions in concert with other genetic risk factors for SCLS.

In the Histh1 locus, the top hits were Creb5 and Tril. Creb5 codes for cyclic AMP-Responsive Element-Binding Protein 5. Creb5 has high expression in the heart (Fagerberg et al., 2014) and has been implicated in cardiac function and pathology (Schisler et al., 2015). Tril is Tlr4 interactor with leucine-rich repeats and is a functional component of Tlr4 complex involved with LPS signaling and is highly expressed in the kidney (Carpenter et al., 2009), indicating a potential role for Tril in blood pressure regulation. Tril(-/-) mice also produce lower levels of multiple proinflammatory cytokines and chemokines within the brain after E. coli and LPS challenge (Wochal et al., 2014), suggesting a potential role in immune modulation. There were no significant hits in the Histh2 locus.

Further experimental validation will be required to confirm the association between our any of the above candidates and Histh. However, the above genes each have compelling functional associations that can inform follow up studies.

**Computation and quantitative trait gene prediction.**

Definitive functional validation of a quantitative trait gene (QTG) has traditionally required either congenic mapping to resolve an extremely narrow QTL, or *ad hoc* nomination of a candidate gene for direct experimentation. The advent of modern genetic technologies, such as CRISPR/Cas9 (Hsu et al., 2014), allow relatively fast and inexpensive allelic manipulations, so the burden of QTG prediction is moving toward a
regime in which a small handful of strong candidates can be followed up individually. Importantly, many QTLs, including Histh, contain multiple causal variants, so fine-mapping alone cannot provide definitive validation. Therefore, computational tools that can identify a small number of reasonable candidates can be a significant aid in biological follow-up. We have presented an integrative strategy for ranking genes in a QTL by combining predicted functional associations to the trait with SNP associations. Our method produces a full ranked list of genes in the locus providing researchers with the potential to validate multiple targets. To this end, the Histh QTL represents an extreme use case for QTG prediction—a large, polygenic QTL associated with a physiologically complex trait.

One major limitation to our approach is the decision of which functional terms to include for network-based prediction. The better tailored this set is to the trait of interest, the greater confidence we can have in the final predictions. In principle, the inclusion of a spurious functional term could skew the rankings toward genes that are functionally associated with the spurious term but irrelevant to the trait of interest. One potential way around this issue is to use functional data, such as transcriptomics, directly from the mapping population. However, in some cases, including Histh, the relevant tissue in which to measure gene expression may not be obvious. Alternatively, one could consider distinct rankings for each functional term. In any case, the researcher will have to exercise some measure of judgment in the prioritization process. However, by transferring the judgments from a large list of positional candidate genes to a smaller and more tractable list of trait-related biological processes, we have shown that we can arrive at a strong set of follow up
candidates that would have evaded naive \textit{p}-value filters and are relatively unbiased by findings published in the literature.

The final output of our method, a ranked list of positional candidate genes, is easy to interpret, and provides researchers with a clear set of hypotheses to test in the lab. While this approach cannot definitively identify the causal gene or genes in a locus, it does provide a much-reduced set of plausible candidates to test.
B.5. Material and Methods

B.5.1. Animals.
A total of 23 mouse strains (129X1/SvJ, A/J, AKR/J, B10.SH2s/SgMcJ (B10.S), BALB/cJ, BPL/1J, BPN/3J, C3H/HeJ, C57BL/6J, CBA/J, CZECHII/EiJ, DBA/1J, DBA/2J, FVB/NJ, JF1/MsJ, MOLF/EiJ, MRL/MpJ, MSM/MsJ, NOD/ShiLtJ, NU/J, PWD/PhJ, PWK/PhJ, SJL/J and SWR/J were purchased from the Jackson Laboratory (Bar Harbor, ME). All mice, including B10.S-\textit{Histh}^{SJL} and B10.S-\textit{Histh}^{SJL} ISRC lines, were generated and maintained under specific pathogen-free conditions in the vivarium of the Given Medical Building at the University of Vermont according to National Institutes of Health guidelines. All animal studies were approved by the Institutional Animal Care and Use Committee of the University of Vermont.

B.5.2. Histh Phenotyping.
On D0 mice were injected i.p. with complete Freund’s adjuvant (CFA) (Sigma-Aldrich, St. Louis, MO) supplemented with 200 mg of \textit{Mycobacterium tuberculosis} H37Ra (Difco Laboratories, Detroit, MI). On D30 histamine hypersensitivity was determined by i.v. injection of histamine (mg/kg dry weight free base) in phosphate buffered saline (PBS). Deaths were recorded at 30 min post injection and the data are reported as the number of animals dead over the number of animals studied. Significance of observed differences was determined by Chi-square with \( p \)-values <0.05 significant.

B.5.3. DNA extraction and genotyping.
DNA was isolated from mouse tail clippings as previously described (Sudweeks \textit{et al.}, 1993). Briefly, individual tail clippings were incubated with 300\( \mu \)L cell lysis buffer
(125µg/mL proteinase K, 100 mM NaCl, 10mM Tris-HCl (pH 8.3), 10 mM EDTA, 100mM KCl, 0.50% SDS) overnight at 55°C. The next day, 150µL of 6M NaCl were added followed by centrifugation for 10 min at 4°C. The supernatant layer was transferred to a fresh tube containing 300µL of isopropanol. After centrifuging for two minutes, the supernatant was discarded, and pellet washed with 70% ethanol. After a final two min centrifugation, the supernatant was discarded, and DNA was air dried and resuspended in 50µL TE.

Genotyping: Genotyping was performed using either microsatellite markers in a standard PCR reaction or sequence specific SNP primers in a phototyping reaction. Polymorphic microsatellites were selected to have a minimum polymorphism of 8bp for optimal identification by agarose gel electrophoresis. Briefly, primers were synthesized by IDT-DNA (Coralville, IA) and diluted to a concentration of 10µM. PCR amplification was performed using Promega GoTaq. The cycling conditions included a two-minute initial denaturation step at 94°C followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds followed by a final extension step at 72°C for five minutes. Amplicons were subjected to 2% agarose gel electrophoresis and visualized by ethidium bromide and UV light.

Phototyping: Genotyping was performed using sequence specific primers that differ only at the 3’ nucleotide corresponding to each allele of the identified SNP (Bunce et al., 1995). Each primer set was designed using Primer3 to have a Tm of 58-60°C, synthesized by IDT-DNA (Coralville, IA), and used at a concentration of 100mM (primer sequences are available online.
PCR reactions were subjected to multistage (high, medium and low stringency) cycling conditions as described and if found to be necessary, the cycle conditions at each stage were adjusted to accommodate the optimal annealing temperature. Amplicons were electrophoresed with 10µl Orange G loading buffer on a 1.5% agarose gel stained with ethidium bromide and visualized by UV light. The presence of a SNP specific allele was scored by observing an amplicon of the expected size in either reaction. Cycling conditions are available in Supplemental Figure B.1.

**B.5.4. Generation of Histh congenic lines and GigaMUGA.**

B10.S-HisthSJL ISRC lines were generated by identifying recombinant haplotypes across the Histh interval among (B10.S-HisthSJL × B10.S) × B10.S BC1 mice and then fixed as homozygous lines (Figure B.1). To identify potential contaminating background loci segregating among the strains and to further refine the recombination break points of each line, the lines were further genotyped using GigaMUGA arrays (143,259 markers) by the commercial service of Neogen/Geneseek (Lincoln, NE).

**B.5.5. Targeted genetic association testing.**

We retrieved genotype data (both coding and non-coding) of the 23 mouse strains from the databases at the Sanger Institute (https://www.sanger.ac.uk/science/data/mouse-genomes-project) and The Jackson Laboratory (https://phenome.jax.org/). The lack of representation of genotype data from B10.S, BPN/3J, BPL/1J, CZECHII/EiJ, JF1/Ms, MOLF/EiJ, MRL/MpJ, NU/J, PWD/J and SJL/J in these databases were compensated by the following:
**The Chromosome Region Capture Sequencing**

*Fragment DNA.* 3 µg of genomic DNA was sheared into fragments of approximately 200 bp with the Covaris E220 system (Covaris, USA), and purification was performed with 1.4-fold volume of AMPure XP Beads (Beckman, USA).

*DNA Library Construction.* After the purification of the sheared DNA, the library was constructed with SureSelect Library Prep Kit (Agilent, USA). End-repair was performed with the volume of 10× End Repair Buffer, dNTP Mix, T4 DNA Polymerase, Klenow DNA Polymerase and T4 Polynucleotide Kinase, reacted at 20 °C for 30 min, and 1.8× the volume of AMPure XP Beads was added for purification; adding A at the 3’ end was performed with the volume of 10×Klenow Polymerase Buffer, dATP and Exo(-) Klenow, reacted at 37 °C for 30 min. T4 DNA Ligase Buffer, SureSelect Adaptor Oligo Mix and T4 DNA Ligase were added and reacted at 20 °C for 15 min. Adaptor-ligated library was amplified through added in SureSelect Primer, SureSelect ILM Indexing Pre Capture PCR Reverse Primer, 5×Herculase II Rxn Buffer, 100 mM dNTP Mix and Herculase II Fusion DNA Polymerase, and the reaction procedure is: 98 °C pre-denaturation 2 min, 98 °C denaturation 30 Sec, 65 °C annealing 30 Sec, 72 °C extension 30 Sec, amplified for 4 rounds. Purification was performed with 1.8X Agencourt AMPure XP beads after each enzymatic reaction. The adaptor-ligated library around range of 225 ~ 275bp was finally obtained.

*Hybridization capture.* Prepared library was executed hybridization capture experiment with the SureSelect Target Enrichment Kit (Agilent, USA). The prepared library reacted with SureSelect Block Mix in 95 °C 5min, followed by maintaining in 65
°C and then Hybridization Buffer, capture library mix was added in and reacted at 65 °C 24hrs, finally Dynabeads M-280 streptavidin (Life, USA) was used for the enrichment of the Captured DNA library (Gnirke et al., 2009; Mamanova et al., 2010).

Index amplification. 5×Herculase II Rxn Buffer, 100 mM dNTP Mix, SureSelect ILM Indexing Post Capture Forward PCR Primer and Herculase II Fusion DNA Polymerase were added in the enriched captured DNA library for index amplification. The reaction procedure is 98°C Pre-denaturation 2 Min, 98 °C denaturation 30 Sec, 57 °C annealing 30 Sec, 72 °C extension 30 Sec, amplification 12 rounds, followed by the purification using 1.8 times the volume of AMPure XP Beads. A Sequencing library of 250-350 bp range was finally obtained (Gnirke et al., 2009).

Sequencing. A 10 ng library was used for cluster generation in cBot with the TruSeq PE Cluster Kit (illumina, USA) followed by bidirectionally sequenced in Illumina Hiseq 2500 to obtain the data of 2x150 bp.

Whole-genome sequencing

DNA Library Construction. For whole-genome sequencing, DNA libraries was constructed according to Illumina recommended protocols. Briefly, 3 μg of genomic DNA was sheared into fragments of approximately 300-400 base pairs with the Covaris E220 system, followed by end-repair, A-tailing, and ligation of the Illumina multiplexing PE adaptors. Purification was performed with 1.8X Agencourt AMPure XP beads after enzymatic reactions. An agarose gel electrophoresis with a concentration of 2% to separate DNA products was performed, and DNA fragments with a length between 300 and 400 bp 231
were recycled and purified according to the user guide of Qiagen Gel Extraction Kit. A PCR enrichment experiment was performed to ensure that DNA products to be successfully sequenced was enough.

*Library inspection.* After construction of the library, preliminary quantification was performed using Qubit 2.0, and the library was diluted to 1 ng/ul, and then the insert size of the library was detected using Agilent 2100. If the insert size was as expected, Q-PCR was performed to accurately quantify the effective concentration of the library (library effective concentration >10 nM) to ensure library quality.

*Sequencing.* Finally, these DNA fragments were subjected to the Illumina Hiseq 2000 platform for pair-end sequencing (PE150). The raw image data files obtained by high-throughput sequencing (Illumina) were converted into Sequenced Reads by CASAVA, and the results were stored in FASTQ format. The read length was 150 bp.

*Data processing and analysis.* To ensure the quality of subsequent information analysis, the original sequence was filtered with the software SolexaQA to get high quality Clean Reads (Cox *et al.,* 2010). Efficient high-quality sequencing data was mapped to the reference genome mm10 by BWA software (Li *et al.,* 2009a), samtools (Li *et al.,* 2009b) was used for sorting, picard tools was used for duplication, and GATK was used for Indel Realignment and Base Recalibration (McKenna *et al.,* 2010). Finally, HaplotypeCaller of GATK is used for mutation detection. The VCF format file was filtered with VCFtools (Danecek *et al.,* 2011). The SNP filtered results of each sample are annotated by ANNOVAR software (Wang *et al.,* 2010), which mainly includes three aspects: annotation
based on gene, genomic region and function. GATK software was also used to detect InDel, and pindel was used to detect SV (Ye et al., 2009), which is divided into four types: deletions (>5bp), Insertions (>5bp), inversions and tandem duplication. The VCF format files were converted to Plink files with VCFtools.

All these data sources were collated to generate genotype information for a total of 13,598 SNPs across the Histh locus. To calculate associations between genetic polymorphisms and Histh, we used efficient mixed-model association (EMMA) (Kang et al., 2008). This method treats genetic relatedness as a random variable in a linear mixed model to account for population structure, thereby reducing false associations between SNPs and the measured trait. We used the likelihood ratio test function (emma.ML.LRT) to generate $p$ values. Significance was defined with a Bonferroni correction ($p = 0.05/13,598$). Genomic coordinates included for each SNP using the latest mouse genome build GRCm38.p5/mm10.

**B.5.6. Trait-related gene sets.**

The positional candidate genes were ranked based on their predicted association with seven functional terms related to the Histh phenotype: “aging”, “mycobacterium tuberculosis”, “cardiac”, “G protein coupled receptor”, “histamine”, “inflammation”, “type I hypersensitivity”, and “vascular permeability.” We used Gene Weaver (Baker et al., 2012) to identify genes associated with each term. We entered each term into the Gene Weaver homepage (https://geneweaver.org). We restricted the search to human, rat, and mouse genes, and to curated lists only. Mouse homologs for each gene were retrieved using batch query tool in MGI (http://www.informatics.jax.org/batch_data.shtml). In addition,
we used Gene Expression Omnibus (GEO) and PubMed to retrieve expression data sets for each phenotype term. Final gene lists consisted of the unique set of genes associated with each process term.

**B.5.7. FNTM network.**

We trained support vector machines (SVMs) to classify genes in each gene list using features derived from the Functional Network of Tissues in Mouse (FNTM) (Goya et al., 2015). In this network, genes are nodes, and the edge weights between them are continuous values between 0 and 1 predicting the degree to which each pair of genes is functionally related. Larger values indicate higher predicted functional relatedness. Functional relatedness in this network was predicted through Bayesian integration of data sets from multiple sources, including gene expression, protein-protein interaction data, and annotation to GO terms (Goya et al., 2015). We downloaded the top edges of the mouse network on January 15, 2018 from https://fntm.princeton.edu.

**B.5.8. Clustering gene sets.**

Guan et al. (Guan et al., 2010) noted that support vector machines trained on 200 to 300 genes yielded the best classification accuracy. Two of our gene lists had fewer than 100 genes. For all lists over 400 genes, we reduced the size of our training sets by clustering each term gene set into modules using the fast greedy (Newman, 2004) algorithm in the R/igraph package (Csardi, 2006). We applied the fast greedy algorithm iteratively until all modules comprised fewer than 400 genes. Using a maximum modules size of 300 overly fragmented the networks yielding many modules with fewer than 100 genes.
**B.5.9. Machine learning.**

To classify novel genes as belonging to a functional module, we trained SVMs using the connection weights in the FNTM network as features, as described in (Guan et al., 2010). Briefly, an annotated set of genes (Figure B.6A, blue nodes) is used as a set of known positives for the corresponding functional module. Other genes in this module are expected to be strongly functionally connected to these known positives, i.e. have high probability of functionally interacting with known positives. Each gene, therefore, is represented as a feature vector of connection weights to the known positives, which can be visualized as a sub-matrix of the adjacency matrix of the network (Figure B.6B). Correspondingly, the rows of this matrix are labeled as either known positive or not (Figure B.6B, blue dots vs. gray dots). We used the e1071 package in R (Dimitriadou et al., 2008) to train SVMs to distinguish the known positive genes from an equal-sized set of genes selected at random from outside the known positive list using the network-based feature vectors (Figure B.6C). The trained model can then annotate novel genes as belonging to the functional module by classifying all gene in the genome (Figure B.6C, green bordered nodes). We trained 100 SVMs on each module selecting a new set of random genes for each run. We used a linear kernel and 10-fold cross-validation for each SVM. We trained each SVM over a series of cost parameters. We started with the sequence $1 \times 10^{-5}$ to $1 \times 10^{2}$ by factors of 10, and iteratively narrowed the range of cost parameters until we found a series of eight cost parameters that maximized the accuracy of the SVM (see Workflow).

We calculated the area under the ROC curves (AUC) over all runs in the following way: For a sequence ranging from the minimum SVM score to the maximum SVM score,
we quantified all true positives (TP), true negative (TN), false positives (FP) and false negatives (FN). The TP genes in this case were those genes from the known positives that were correctly classified as being in the module (above the SVM score cutoff). TN genes in this case were those genes outside the module that were correctly classified as being outside the module (below the SVM score cutoff). We calculated the AUC across the average curve for all 100 SVMs for each module.

**B.5.10. Positional Candidate Scoring.**

We used the trained SVMs to score each positional candidate gene in the *Histh* locus. The score for each gene gave an estimate of how functionally related each gene was to each module based on its connection weights to the known module genes in the FNTM mouse network. Genes with large positive scores were predicted by the SVMs to interact functionally with the genes in the module, while genes with negative scores were predicted to not functionally interact with the module genes. To be able to compare SVM scores across different trained models, we calculated a false positive rate ($FPR$) for each gene and each SVM as follows: For each gene’s SVM score we calculated the number of true positives ($TP$), true negatives ($TN$), false positives ($FP$) and false negatives ($FN$) classified by the SVM. The $FPR$ for a given SVM score was calculated as $FP/(FP + TN)$.

The final functional score for each was the $\max(-\log_{10}(FPR))$ across all modules. This meant that genes with a high functional score for a single module, but low functional scores for other modules received higher overall scores than genes with moderately high scores across all modules.
B.5.11. Combined Gene Score.

High-quality candidate genes in the locus should not only be functionally related to the trait of interest, but should also segregate with the trait of interest. We thus defined a combined gene score ($S_{cg}$) that combined these two aspects of the analysis:

$$S_{cg} = \frac{-\log_{10}(p_{EMMA})}{\max_{\text{pos.cand.}} -\log_{10}(p_{EMMA})} + \frac{-\log_{10}(FPR_{SVM})}{\max_{\text{pos.cand.}} -\log_{10}(FPR_{SVM})},$$

where the denominators of the two terms on the right hand side are the maximum values of $-\log_{10}(p_{EMMA})$ and $-\log_{10}(FPR_{SVM})$ over all positional candidates in Histh, respectively, which normalizes the functional and positional scores to be comparable to each other. EMMA $p$ values for SNPs were assigned to the nearest gene within 1 megabase using the R package biomaRt (Durinck et al., 2005; Durinck et al., 2009). Genes for which more than one SNP was assigned were given the maximum $-\log_{10}(p_{EMMA})$ across all SNPs associated with that gene. The rows of this matrix are sorted by the maximum gene score across all gene lists.

B.5.12. Data Availability.

A reproducible workflow in R markdown is available on GitHub (https://github.com/MahoneyLab/HisthFunctionalRankings). This workflow contains all code required to reproduce the figures and results presented in this manuscript. The data used as input for the workflow, as well as intermediate and final-results are available on Figshare (https://figshare.com).
B.6. Acknowledgements

We would like to thank Laura Cort for supervising students during genotyping of congenic mice. ALT and JMM are supported by a grant (R21 LM012615) from the National Library of Medicine of the United States National Institutes of Health (NIH). AR, DNK, EPB, and CT were supported by grants from the NIH and the National Multiple Sclerosis Society (NMSS). DNK was supported by NIH grants from the National Institute of Neurological Disease and Stroke (R01 NS097596), National Institute of Allergy and Infectious Disease (R21 AI145306), and the NMSS (RR-1602-07780).
B.7. References


Rusmini, M., Griseri, P., Lantieri, F., Matera, I., Hudspeth, K. L., Roberto, A., Mikulak, J., Avanzini, S., Rossi, V., Mattioli, G., Jasonni, V., Ravazzolo, R., Pavan, W. J., Pini-


B.8. Figures

Figure B.1. Interval specific recombinant congenic (ISRC) mapping places Histh candidates in four genetic loci.

ISRC lines were injected (D0) with complete Freund’s adjuvant (CFA) and subsequently challenged (D30) with i.v. injection of histamine to determine histamine hypersensitivity. Deaths were recorded at 30 min post injection and the data are reported as the number of animals dead over the number of animals studied. Significance of observed differences was determined by a $\chi^2$ test with p-values <0.05 considered significant.
**Figure B.2. Targeted genetic association analysis for Histh.**

Negative log-transformed $p$ values of SNP associations with Histh. Genomic coordinates (mm10 Mbp) of each SNP are shown along the x-axis. Each circle denotes a single SNP. Gene names are included for SNPs that crossed $p$-value threshold of $3.68 \times 10^{-6}$ shown with a red dotted line. The location of Histh sub-QTLs are shown at the top of the figure.
Figure B.3. Workflow Overview.

The workflow is broken into blocks by color, each with a bolded title. Each block shows how data (blue rectangles) were operated on (gray rectangles) to achieve results (green rectangles). Arrows show the general flow of work and dependence (and independence) of individual analyses.
Figure B.4. Two axes of gene scoring.

Gene names are plotted by their $-\log(p_{EMMA})$ on the x-axis and the $-\log(FPR_{SVM})$ on the y-axis. Both scores were scaled by their maximum value for better comparison. Genes farther to the right were associated with SNPs that segregated with Histh. Genes higher up on the y-axis are associated with stronger functional association with gene modules. The blue line marks the Pareto front. Genes on this line maximize the two scores and are the best candidates based on the combination of both scores.
Figure B.5. Final gene scores.

Gene functional values were combined with SNP associations to assign each gene a final gene score ($S_{cg}$). Higher gene scores indicate better candidates. (A) Heat map showing the final score of each of the top 20-ranked genes for each gene module. To aid visualization of the strongest candidates,
asterisks in each cell indicate where candidate genes were associated with a module with an $FPR_{SVM}$ $\leq 0.2$. (B) The top panel shows individual SNPs plotted at their genomic location (x-axis) and their $\log_{10}(p_{EMMA})$ (y-axis). All SNPs with nominally significant p value ($p \leq 0.05$) are plotted. The horizontal line indicates the Bonferroni corrected significance cutoff ($p \leq 0.05/13598$). The bottom panel shows genes plotted at their genomic location (x-axis) and their final gene score ($S_{cg}$) (y-axis) to demonstrate how the final ranked genes align with the SNP association data.
Figure B.6. Network-based machine learning for functionally annotating genes.

(A) Known-positive genes annotated to a functional term (blue nodes) are typically densely interconnected in a functional network. (B) The adjacency matrix of a network is a tabular representation of the connectivity structure of the network in which each row/column corresponds to a node of the network, and connected pairs of nodes have non-zero values in the corresponding cell of the matrix. Note that in general the connections are weighted, but for display we are only showing present/absent links (white/black cells). The connections from every gene in the genome to the known positives form a sub-matrix of the adjacency matrix called the feature matrix (vertical red lines), whose rows are the feature vectors for each gene. (C) Using the network-based feature vectors for each gene, we train SVMs to distinguish known positives (blue dots) from random genes in the genome (gray dots) to identify the full sub-network corresponding to the true positive genes (green bordered dots and dotted red lines in panels A,B).
### B.9. Tables

**Table B.1. A survey of Histh phenotypes across 23 inbred mouse strains.**

<table>
<thead>
<tr>
<th>Strain</th>
<th>HA</th>
<th>Strain</th>
<th>HA</th>
<th>Strain</th>
<th>HA</th>
<th>Strain</th>
<th>HA</th>
</tr>
</thead>
<tbody>
<tr>
<td>129X1/SvJ</td>
<td>0/8</td>
<td>C3H/HeJ</td>
<td>0/8</td>
<td>DBA/2J</td>
<td>0/8</td>
<td>PWK/PhJ</td>
<td>0/6</td>
</tr>
<tr>
<td>A/J</td>
<td>0/8</td>
<td>C57BL/10J</td>
<td>0/8</td>
<td>JF1/Ms</td>
<td>0/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKR/J</td>
<td>0/8</td>
<td>C57BL/6J</td>
<td>0/7</td>
<td>MOLF/EiJ</td>
<td>0/8</td>
<td>FVB/J</td>
<td>6/8</td>
</tr>
<tr>
<td>BALB/cJ</td>
<td>0/8</td>
<td>CBA/J</td>
<td>0/8</td>
<td>MRL/MpJ</td>
<td>0/8</td>
<td>NU/J</td>
<td>5/8</td>
</tr>
<tr>
<td>BPL/1J</td>
<td>0/8</td>
<td>CZECHII/EiJ</td>
<td>0/8</td>
<td>NOD/ShiLtJ</td>
<td>0/8</td>
<td>SJL/J</td>
<td>12/12</td>
</tr>
<tr>
<td>BPN/3J</td>
<td>0/8</td>
<td>DBA/1J</td>
<td>0/8</td>
<td>PWD/PhJ</td>
<td>0/12</td>
<td>SWR/J</td>
<td>6/8</td>
</tr>
</tbody>
</table>

Cohorts of CFA injected 8 to10 week old mice were challenge 30 days later with 75 mg/kg HA by i.v. injection, and deaths recorded at 30min. Results are expressed as the #of animals dead/#of animals studied.
B.10. Supplemental Data.

**Supplemental Table B.1. Cycling conditions for PCR.**

<table>
<thead>
<tr>
<th>Cycling conditions for SNP</th>
<th>96°C</th>
<th>60s</th>
<th>1 cycle</th>
<th>96°C</th>
<th>60s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70°C</td>
<td>45s</td>
<td>annealing</td>
<td>72°C</td>
<td>25s</td>
</tr>
<tr>
<td></td>
<td>96°C</td>
<td>25s</td>
<td></td>
<td>65°C</td>
<td>50s</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>30s</td>
<td>25s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>96°C</td>
<td>30s</td>
<td></td>
<td>55°C</td>
<td>60s</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>90s</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

cool by ramping to 20°C for 30s
APPENDIX C: GENETIC ANALYSIS OF BPHSE: A NOVEL GENE COMPLEMENTING RESISTANCE TO BORDETELLA PERTUSSIS-INDUCED HISTAMINE SENSITIVITY.

Abbas Raza¹, Sean A. Diehl¹, Laure Case⁷, Dimitry Krementsov¹,², Dawei Li³, Jason Kost³, Anna Tyler⁷, Matthew Mahoney⁶, Elizabeth P. Blankenhorn⁸, and Cory Teuscher¹,⁵

¹Departments of Medicine, ²Department of Biomedical and Health Sciences, ³Department of Microbiology and Molecular Genetics, ⁵Department of Pathology and ⁶Department of Neurol Sci, University of Vermont, Burlington, VT, 05401, USA; ⁷The Jackson Laboratory, Bar Harbor, ME, 04609, USA; ⁸Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, 19129, USA

Address correspondence to: Cory Teuscher, 89 Beaumont Avenue Given C Bldg, Rm 329A, University of Vermont Larner College of Medicine, Burlington, VT 05405;
c.teuscher@uvm.edu.

Keywords:
Abbreviations: Bphs/Bphs, Bordetella pertussis induced histamine sensitivity; Bphse/Bphse, enhancer of Bordetella pertussis induced histamine sensitivity; HA, histamine; Hrh1/H₁R, histamine receptor H₁; PTX, Pertussis toxin, T1H, Type 1 hypersensitivity; Vasc.perm, Vascular permeability; LD, linkage disequilibrium; GPCR, G
protein coupled receptor; ER, Endoplasmic reticulum; Endoplasmic membrane protein complex (EMC); Endoplasmic reticulum-associated degradation (ERAD).
C.1. Abstract

Susceptibility to *Bordetella pertussis* and pertussis toxin (PTX) induced histamine sensitization (Bphs/Bphs) is controlled by histamine H₁ receptor (*Hrh1/H₁R*) alleles, with the susceptible (Bphs<sup>S</sup>/H₁R<sup>S</sup>) and resistant (Bphs<sup>R</sup>/H₁R<sup>R</sup>) alleles differing by three amino acids (P263L, V313M and P331S) within the third intracellular loop associated with signal transduction, protein folding, and trafficking. Functionally, the two alleles equally activate G<sub>αq/11</sub>, the G protein family members that couple H₁R signaling to second messenger signaling backways, indicating that susceptibility and resistance is not due to differential activation of G<sub>αq/11</sub>. In contrast, the two alleles exhibit differential cell surface expression and altered intracellular trafficking, with the H₁R<sup>R</sup> allele preferentially retained within the endoplasmic reticulum (ER), and all three residues comprising the H₁R<sup>R</sup> haplotype required for altered expression. Given the importance of H₁R signaling in health and disease we phenotyped and sequenced the third intracellular loop of H₁R among a large panel of inbred laboratory and wild-derived mouse strains to identify potential allelic co-adaptations capable of complementing Bphs<sup>R</sup> in mice with a H₁R<sup>R</sup> allele. Taking this approach, we identified eight inbred mouse strains with a H₁R<sup>R</sup> allele that are Bphs<sup>S</sup> (Bphs<sup>S</sup>/H₁R<sup>R</sup>). Genetic analyses mapped the locus complementing Bphs<sup>R</sup> to mouse Chr 6, in linkage disequilibrium with *Hrh1*; designated Bphs-enhancer (*Bphse*). Interval-specific SNP based association testing and functional enrichment resulted in the identification of nine candidate genes for *Bphse* within a 113-116Mb interval, including *Atp2b2*, *Atg7*, *Pparg*, *Syn2*, *Ift122*, *Raf1*, *Mkrn2*, *Timp4* and *Gt(Rosa)26Sor*. These findings reveal an additional mechanism by which genetic factors control Bphs<sup>S</sup>.
C.2. Introduction

Histamine (2-[4-imidazole]-ethylamine) is an endogenous biogenic monoamine that is synthesized, stored intracellularly within granules, and following cellular activation released by mast cells, basophils, platelets, neurons, and enterchromaffin-like cells in the stomach (Panula et al., 2015). After release, free histamine (HA) mediates its effects by binding to four different 7-transmembrane G-protein-coupled receptors (GPCRs): histamine receptor H\textsubscript{1} (Hrh1/H1R), Hrh2/H2R, Hrh3/H3R, and Hrh4/H4R, expressed on target cells in various tissues (Parsons et al., 2006).

Histamine acting through these receptors influences a diverse array of physiological processes, including brain function, neurotransmission, secretion of pituitary hormones, cell proliferation and differentiation, hematopoiesis, embryonic development, wound healing and regeneration, and the regulation of gastrointestinal, cardiovascular, and secretory functions (Parsons et al., 2006). In addition, HA plays a major role in inflammation and the regulation of innate and adaptive immune responses in both normal and pathologic states (Branco et al., 2018). In fact, HA has been linked to more than twenty-five different physiological functions and is one of the most extensively studied chemical compounds with physiologic activity (Falus et al., 2004).

Historically, HA is most well-known for its role in HA-shock and anaphylaxis (Peavy et al., 2008a). Histamine was first synthesized in 1907 (Windaus et al., 1907), and in 1910 isolated in sufficient quantities for experimentation by Barger from the parasitic rye mold, ergot (Claviceps purpurea) (Barger G, 1910). They showed that HA caused smooth muscle contraction, and subsequently, in a series of crucial experiments, Dale
found that HA induced a shock-like syndrome when injected into mammals (Dale et al., 1910). In addition, HA caused bronchiolar constriction, constricted cardiac and pulmonary arteries and stimulated cardiac contraction. Further research firmly established that HA was a natural constituent of the body and a mediator of anaphylactic shock (Peavy et al., 2008b).

Studies of species differences in susceptibility to HA-shock revealed significant differences with guinea pigs and rabbits being highly susceptible whereas mice and rats, were in general, remarkably resistant (Munoz et al., 1968). However, Parfentjev showed in 1948 that prior exposure of mice to Bordetella pertussis vaccine increased the susceptibility of mice to HA up to 100-fold (Parfentjev et al., 1948). Susceptible strains died within 30 minutes as a result of hypotensive (drop in blood pressure) and hypovolemic (drop in blood volume) shock (Iff et al., 1966). This phenotype was designated Bphs for B. pertussis-induced histamine sensitivity (Sudweeks et al., 1993). Subsequent studies showed that B. pertussis-treated mice exhibited increased vascular permeability and alteration in blood-tissue barrier functions in association with susceptibility to HA-shock and other vasoactive agents including serotonin (Bpss) and bradykinin (Bpbs) (Gao et al., 2003b). The sensitization phenotype is unique to B. pertussis, as other bacterial or viral infections/exposures do not elicit increased sensitivity to HA-shock (Kind, 1953). The sensitizing activity elicited by exposure to B. pertussis was subsequently shown to be a function of pertussis toxin (PTX)-catalyzed ADP-ribosylation of the α subunit of heterotrimeric G_{i/o} proteins (G_{α_{i/o}}) (Katada et al., 1982; Kurose et al., 1983; Diehl et al., 2014).
Genetic studies using inbred strains of mice established that susceptibility to Bphs, Bpss, and Bpbs are under unique genetic control (Diehl et al., 2014) and that Bphs was controlled by a single dominant locus (Wardlaw, 1970). Using a forward genetic approach, we identified Bphs as Hrh1/H1R. H1R-susceptible (BphsS/H1RS) and -resistant (BphsR/H1RR) alleles differ by three amino acids (P263L, V313M and P331S) (Ma et al., 2002a) within the disordered third intracellular loop of GPCRs that is implicated in signal transduction, protein folding, and trafficking (Venkatakrishnan et al., 2014; Latorraca et al., 2017). Functionally, H1RS and H1RR alleles equally activate Gaq/11, the G protein family members that couple H1R signaling to second messenger signaling pathways (Parsons et al., 2006; Monczor et al., 2016), indicating that the genetic control of susceptibility and resistance to Bphs is not inherently due to differential activation of either Gaq or Gal11. However, the two alleles exhibit differential cell surface expression and altered intracellular trafficking, with the H1RR allele selectively retained within the endoplasmic reticulum (ER). Importantly, all three residues (263L-313M-331S) comprising the H1RR haplotype are required for altered expression (Noubade et al., 2008).

Given the importance of H1R signaling in health and disease we phenotyped and sequenced the H1R alleles of a panel of inbred laboratory and wild-derived strains to identify potential allelic co-adaptations capable of complementing BphsR in mice with a H1RR allele. We identified eight inbred strains that phenotyped as BphsS, despite carrying an H1RR allele. Genetic analysis, reported herein, identified a dominant locus linked to Hrh1 capable of complementing BphsR. We have designated this locus Bphse for Bphs-
enhancer. Interval-specific SNP-based association testing and functional enrichment are used to identify candidate genes for *Bphse*.
C.3. Results

C.3.1. H₁R is highly conserved in mice.

Given the role of H₁R signaling in diverse normal and pathologic states, particularly HA-shock anaphylaxis, we undertook a genetic approach to screen for evolutionarily selected mechanisms that may be capable of complementing Bphsᵣ. Toward this end, we sequenced the third intracellular loop of Hrh1 across a panel of 91 inbred laboratory and wild-derived strains of mice (Table C.1). Surprisingly, we did not identify any additional genetic variants other than the P263L, V313M and P331S haplotypes associated with Bphsˢ/H₁Rˢ and Bphsᵣ/H₁Rᵣ (Ma et al., 2002a). Of the 91 strains, 22 carry the Bphsᵣ/H₁Rᵣ allele whereas 69 carry the Bphsˢ/H₁Rˢ allele. The evolutionary distribution of the two alleles was assessed using a published mouse phylogenetic family tree (Petkov et al., 2004) (Supplementary Figure C.1). The Bphsᵣ/H₁Rᵣ allele was primarily restricted to wild-derived Group 7 strains and a selected sub-branch of Group 1 Bagg albino derivatives whereas the Bphsˢ/H₁Rˢ allele was distributed across Groups 2-6. Interestingly, Group 7 is highly heterogeneous and includes representatives of M. m. domesticus (PERA, PERC, ZALENDE and TIRANO), M. m. musculus (PWK, PWD, CZECHI, CZECHII and WSB), M. m. castaneus (CAST), M. m. molossinus (JF1, MSM, MOLF, MOLD, MOLC), M. m. hortulanus (PANCEVO/EiJ), M. m. spretus (SPRET), M. m. praetextus (IS), or hybrids of M. m. musculus and M. m. domesticus (SKIVE), M. m. musculus and M. m. poschiavinus (RBF) and of M.m.castaneous and M.m.domesticus (CALB) (Boursot et al., 1993; Beck et al., 2000; Bult et al., 2019)
C.3.2. Identification of co-adaptation complementing Bphs\textsuperscript{R}.

Compared to classical inbred strains, wild-derived mice exhibit sequence variation at approximately every 100-200 base pairs (Poltorak \textit{et al.}, 2018) and are, in general, more resistant to a variety of pathogens, most notably viral infections (Guenet \textit{et al.}, 2003; Harper, 2008; Dejager \textit{et al.}, 2009; Bearoff \textit{et al.}, 2016). This genetic variability represents a rich source of evolutionary selected diversity and has the potential to lead to the identification of genes controlling novel regulatory features arising from host-pathogen co-evolutionary adaptation. To screen for functional co-adaptations capable of complementing Bphs\textsuperscript{R}, we phenotyped a panel of Group 1 and 7 mice with the H\textsubscript{1}R\textsuperscript{R} allele for susceptibility to Bphs.. Among the seventeen H\textsubscript{1}R\textsuperscript{R} strains that we phenotyped, eight were Bphs\textsuperscript{S} (Table C.2). Importantly, the Bphs\textsuperscript{S}/H\textsubscript{1}R\textsuperscript{R} strains are confined primarily to a Group 7 wild-derived strains (Supplementary Figure C.1) whereas most H\textsubscript{1}R\textsuperscript{R} Group 1 strains were Bphs\textsuperscript{R}. We therefore selected a subset of Bphs\textsuperscript{S}/H\textsubscript{1}R\textsuperscript{R} (MOLF, PWK) and Bphs\textsuperscript{R}/H\textsubscript{1}R\textsuperscript{R} (AKR, CBA, C3H, MRL) strains for further studies.

To confirm the existence of a co-adaptation capable of restoring Bphs susceptibility in mice with an H\textsubscript{1}R\textsuperscript{R} allele and to assess its heritability, we studied F\textsubscript{1} hybrids between the select strains of interest and H\textsubscript{1}RKO mice (Table C.3). (B6 × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}} and (C3H.Bphs\textsuperscript{S} × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}} served as Bphs\textsuperscript{S}/H\textsubscript{1}R\textsuperscript{S} controls while (C3H × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}}, (CBA × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}}, (MRL × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}} and (AKR × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}} served as Bphs\textsuperscript{R}/H\textsubscript{1}R\textsuperscript{R} controls, respectively. Both H\textsubscript{1}R\textsuperscript{S} by H\textsubscript{1}RKO F\textsubscript{1} hybrids were Bphs\textsuperscript{S} whereas none of the H\textsubscript{1}R\textsuperscript{R} by H\textsubscript{1}RKO F\textsubscript{1} hybrids were susceptible, in agreement with our prior finding (Ma \textit{et al.}, 2002a). In contrast, both (MOLF × H\textsubscript{1}RKO)\textsubscript{F\textsubscript{1}} and (PWK ×
H1RKO)F1 mice were BphsS. These data support the existence of one or more dominant loci in MOLF and PWK capable of complementing BphsR.

C.3.3. A functional linkage disequilibrium domain on Chr 6 regulates susceptibility to Bphs.

Given the evidence from inbred strains of mice indicating that a quarter or more of the mammalian genome consists of chromosomal regions containing clusters of functionally related genes, i.e., functional linkage disequilibrium domains (Petkov et al., 2005; Graber et al., 2006), we hypothesized that the dominant locus complementing BphsR/H1R may reside within a Hrh1/H1R functional LD domain. To test this, we generated and phenotyped 114 (MOLF × H1RKO) × H1RKO backcross (BC) mice (Table C.3). As expected, none of the 54 homozygous H1RKO mice phenotyped as BphsS. In contrast, of the 60 H1R/−MOLF mice, 54 were BphsS and 6 were BphsR, indicating that the locus capable of complementing BphsR is linked to Hrh1. We have designated this locus Bphse for Bphs-enhancer.

Further support for linkage of Bphse to Hrh1 comes from genetic linkage analysis using 170 (AKR × PWK) × AKR BC mice and informative Chr 6 marker loci (Table C.4). Marker loci from rs36385580 through D6mit135 exhibited significant linkage to Bphse with maximal linkage across the interval bounded by D6Mit102 and rs31698248; thereby, placing Bphse within a 93-120Mb interval encompassing Hrh1. We next confirmed the existence and location of Bphse by congenic mapping. Marker-assisted selection was used to introgress the BphseMOLF and BphsePWK intervals onto the BphsR C3H and AKR backgrounds, respectively. Starting at N5 through N10, heterozygous and homozygous BC
mice were phenotyped for Bphs (Figure C.1). Compared to C3H (C3H and C3H.Bphse^{C3H/C3H}) and AKR (AKR and AKR.Bphse^{AKR/ AKR}) mice both C3H.Bphse^{C3H/MOLF} and AKR.Bphse^{AKR/PWK} mice were Bphs\. Overall, Bphse^{C3H/MOLF} and Bphse^{AKR/PWK} mice were significantly more susceptible to Bphs than Bphse^{C3H/C3H} and Bphse^{AKR/ AKR} mice (χ² = 60.63, df=1, p<0.0001). Taken together, these results confirm the location of Bphse within an interval encompassing Hrh1, providing further support for the existence of a functional LD domain controlling overall responsiveness to HA.

### C.3.4. Identification of candidate genes for Bphse.

Using genotype data collected from Mouse Phenome Database, Mouse Genomes Project, and Nimblegen Sequence Capture (See Material & Methods) for the seventeen Bphs^R/H_1R^R and Bphs^S/H_1R^R strains (Table C.2), we searched for genetic associations between single nucleotide polymorphisms (SNPs) and Bphs^S\. Not surprisingly, this approach yielded poor resolution as the sample size was limited (data not shown). To circumvent the sample size limitation in genetic association testing, we utilized all available Bphs phenotype data (Supplementary Table C.1) and searched for SNPs that could explain Bphs^S irrespective of Hrh1/H_1R genotype. Using efficient mixed-model association (EMMA) (Kang et al., 2008), we identified 69 SNPs in 28 genes that were associated with Bphs^S (data not shown). Most significant SNPs were intronic (64%) followed by non-coding (13%), regulatory (3%) and intergenic (1%) variants).

As a complementary approach to identify loci associated with Bphs^S, we utilized machine-learning computation, using functional genomic networks (Guan et al., 2010) to identify network-based signatures of biological association. To this end, we used prior
knowledge to generate a list of Bphs-associated biological processes and retrieved gene sets functionally associated with each term. The terms and their justifications are as follows:

- **Type I hypersensitivity/Anaphylaxis**: The death response following systemic HA challenge exhibits symptoms of type I hypersensitivity/anaphylaxis including respiratory distress, vasodilation, and anaphylactic shock (Munoz et al., 1965).

- **Cardiac**: There is evidence suggesting that anaphylactic shock in mice is caused by decreased cardiac output, rather than systemic vasodilation (Wang et al., 2014).

- **Histamine**: Bphs is induced by a systemic HA challenge (Ma et al., 2002b).

- **G-protein coupled receptor**: H₁R signaling is required for the Bphs phenotype (Ma et al., 2002b), and all HA receptors belong to the family of G-protein coupled receptors (Seifert et al., 2013).

- **Pertussis toxin**: Bphs is induced in mouse strains by PTX (Munoz et al., 1968).

- **Vascular permeability**: The Bphs response includes vascular leakage in skin and muscles (Munoz et al., 1965) (Gao et al., 2003a).

- **Endoplasmic reticulum (ER)/endoplasmic membrane protein complex (EMC), and endoplasmic reticulum-associated degradation (ERAD)**: The two H₁R alleles exhibit differential protein trafficking and cell surface expression whereby *Hrh1R* is retained in endoplasmic reticulum (Noubade et al., 2008).
The EMC and ERAD are intimately involved in regulating GPCR translocation to the plasma membrane (Lackman et al., 2014; Chitwood et al., 2018).

Each of the six gene sets define a putative Bphs-related process that forms a distinct subnetwork of the full functional genomic network. Using this approach, we identified several hundred genes within the Bphse locus that are functionally associated with each biological process (data not shown).

Genes that are predicted to be highly functionally related to the trait may not have functionally variant alleles in the study population and may therefore be unlikely to drive the observed strain differences in Bphs. We, therefore, integrated the genetic association (-log_{10} p_{EMMA}) with functional enrichment (-log_{10} FPR) to focus on genes that were significant in both approaches (Figure C.4A). The final ranking was calculated by defining a final gene score (S_{cg}) for each gene, which is the sum of the (normalized) -log_{10}(FPR) and the -log_{10}(p_{EMMA}) (Figure C.2B). This approach revealed several candidates including Hrh1, which is a positive control shown to regulate Bphs among laboratory inbred strains of mice (Ma et al., 2002b). The nine candidates for Bphse are ranked using S_{cg}: Atp2b2, Atg7, Pparg, Syn2, Ifit122, Raf1, Mkn2, Timp4 and Gt(ROSA)26Sor. Each of these candidates are associated with sub-phenotypes of Bphs supporting the utility of this approach (Table C.5). Interestingly, all predicted candidates localize around 113-116Mb overlapping the Bphs locus (Figure C.2B), in agreement with the presence of a functional LD domain containing Hrh1 and Bphse that regulate overall responsiveness to HA.
In summary, we have identified a small set of positional candidate genes in a large locus by combining linkage analysis, congenic mapping, genetic and computational predictions of functional association with Bphs susceptibility in mice. The final list of genes is highly plausible and can be followed up relatively easily with modern genetic editing techniques.
C.4. Discussion

*Bordetella pertussis* and its purified toxin, PTX, mediates several phenotypes *in vivo* including hypersensitivity to HA, serotonin, bradykinin, active and passive anaphylaxis, endotoxins, X-irradiation, cold stress, hypoglycemia, hypoproteinemia, changes in vascular permeability and enhancement of organ specific autoimmune diseases like EAE, EAU (Munoz *et al.*, 1968; Ma *et al.*, 2002a). Our previous work had found that the genetic control of HA hypersensitivity among laboratory inbred strains of mice following PTX challenge is dependent on the haplotype of histamine 1 receptor with H₁R<sup>S</sup> allele exhibiting susceptibility whereas H₁R<sup>R</sup> imparting resistance to Bphs (Ma *et al.*, 2002a). Phylogenetically, the H₁R<sup>R</sup> allele is restricted to wild-derived Group 7 strains and a selected sub-branch of Group 1 Bagg albino derivatives whereas the H₁R<sup>S</sup> allele is distributed across all groups (Supplementary Figure C.1). Herein, we present data from several wild derived inbred strains of mice that harbor H₁R<sup>R</sup> but surprisingly a subset of them exhibit histamine hypersensitivity. This is suggestive of an evolutionary adaptation in the genomes of these Group 7 (e.g., MOLF) mice that can complement Bphs/ H₁R<sup>R</sup> resistance. We have designated this locus *Bphse* for Bphs-enhancer.

Our results with the genetic cross [(MOLF × C3H) × C3H backcross mice] confirm the existence of this co-adaptation capable of complimenting susceptibility to Bphs in mice with an H₁R<sup>R</sup> allele. We also found that this complementing dominant locus (*Bphse*) does require H₁R as none of the BC1 mice that genotype as H₁R<sup>−/−</sup> exhibit Bphs (Table C.3). Among BC1 mice that genotype as H₁R<sup>MOLF<sup>+</sup></sup>, 10% phenotyped as Bphs<sup>R</sup> suggesting that the co-adaptation that complements the H₁R<sup>R</sup> allele in wild-derived mice is linked to *Hrh1*.
at > 10.7 cM. Linkage scan using microsatellite markers validates significant linkages to Chr 6, with most significance around \( Hrh1 \) (Table C.4). In addition, we show the physical location of this locus by making congenic mice (C3H.\( Bphse^{MOLF+/-} \) and AKR.\( Bphse^{PWK+/-} \)) that captures the \( Bphse \) locus on Chr 6 (59.3-128.8Mbp, Figure C.1) and replicate the phenotype. To our knowledge, this is the first study assessing Bphs susceptibility among wild derived inbred strains of mice.

Several factors can influence histamine sensitivity after PTX inoculation including age, sex, strain of mice, route of sensitization in addition to the haplotype of \( Hrh1 \) (Munoz et al., 1968; Ma et al., 2002a). In our phenotyping experiments, we used 8-12-week-old mice of each sex and did not find any sex differences. This agrees with earlier studies that found no such effects in Bphs (Fink et al., 1954). We also tested the route of administration of PTX and histamine challenge using intraperitoneal and intravenous route and found no difference (data not shown). We have not tested the effect of age on Bphs susceptibility amongst the various strains, however work from Munoz and others have reported significant effect of age (Munoz et al., 1953). It is quite possible that some of the strains that show resistance to Bphs may exhibit susceptibility as they age.

\( H_1R \) is a GPCR. Normally, GPCRs function at the cell surface, but mutated GPCRs can become sequestered in the ER. We have earlier shown that the two \( H_1R \) alleles exhibit differential cell surface expression and altered intracellular trafficking, with the \( H_1R^R \) allele being retained within the ER. Synthesis of GPCRs in the ER is, by itself, insufficient to result in effective cell surface delivery and function. It is possible that the retention of \( H_1R^R \) is a result of improper folding and conformation, and is therefore unable to exit the ER.
We have earlier tested that the two H1R alleles can activate downstream Gαq and Ga11 proteins equally well in vitro suggesting the defect lies not in the signaling but in the location of the two alleles. Both conventional and nonconventional chaperone and escort proteins play key roles in promoting proper folding, preventing the transport of incorrectly folded proteins to the Golgi, and regulation of receptor trafficking to the cell surface (Achour et al., 2008). If the defect cannot be corrected, it can lead to disease or in the case of H1R, resistance to Bphs. There are several examples associated with GPCR mutations and incorrect folding e.g. mutations in the vasopressin V2 receptor are linked to nephrogenic diabetes insipidus (Robben et al., 2005) in the photon receptor rhodopsin associated with retinitis pigmentosa (Illing et al., 2002) and in the gonadotropin-releasing hormone receptor (GRHR) resulting in hypogonadotropic hypogonadism (de Roux et al., 1997). It is highly possible that Bphse encodes a chaperone or an adaptor protein for H1R that regulates trafficking to the surface thereby rescuing the signaling by H1R. It will be interesting to test the surface expression of H1R among strains that show Bphs susceptibility (Table C.2). Results using bone marrow chimeras suggest that susceptibility to Bphs lies in the non-hematopoietic compartment (Lu et al., 2010) so several cell types (endothelial, epithelial, stromal cells) are potential candidates for this cell surface expression analysis.

Histh is an autosomal recessive genetic locus that regulates age and/inflammation dependent responses to histamine ((Tyler et al., 2019)). Interestingly, Histh is in strong linkage with Bphse and Hrh1 on Chr 6. This is a classic example of the physical linkage of functionally related genes, i.e., functional LD domains as seen in a quarter or more of the
mammalian genome (Petkov et al., 2005; Graber et al., 2006). LD-domains can be thought of as gene clusters whose products act in a similar pathway and tend to segregate together in linkage studies. Thus, it is possible that Bphse may encode genes related to H₁R folding, trafficking, or signaling. Unfortunately, the region implicated for Bphse is very large and contain hundreds of genes. Until recently, interval specific recombinant congenic mapping was the gold standard to delimit large quantitative trait loci (QTLs) associated with a phenotype (Ma et al., 2002a; Schallschmidt et al., 2018). Of the thousands of QTLs for various phenotypes and diseases, only a small fraction of genes have been identified through sub-congenic mapping, phenotyping and sequencing. The identification of candidate genes from large genomic regions has been revolutionized with the advent of advanced sequencing technologies and genome wide association studies (GWAS) (Kelly et al., 2017). For example, the Sanger Institute has sequenced 36 mouse genomes and The Jackson Laboratory, in conjunction with the University of North Carolina, has genotyped several hundred laboratory inbred strains using the Mouse Diversity Array, which altogether provide an almost complete picture of genetic variation among the various strains. Our approach, however, is different from other mouse GWAS studies that have been done to identify candidate loci for several diseases (Klein et al., 2016; Kelly et al., 2017). Instead of running a full genome scan among Bphs strains, we tested association of susceptibility exclusively across the Bphse locus (Chr 6:59-128Mbp). This allowed us to use the information gathered from the genetic cross and congenic mapping and delimit the region to be screened for association. Our first screen using genetoype and phenotype data across seventeen H₁R⁺ mouse strains did not yield any significant hits, mostly due to
limitation in number of strains used. To overcome this problem, we excluded H₁R genotype as the co-variante. Given that several dozen laboratory inbred mouse strains (Group 2, 3 and 4) have been described in literature that are phenotyped for Bphs (Linthicum et al., 1982; Diehl et al., 2014), and also to circumvent the sample size limitation in genetic association testing, we searched for genetic polymorphism among 50 mouse strains that could explain overall responsiveness to histamine. Hrh1, which is our positive control and associated with Bphs susceptibility among classical laboratory inbred strains (Ma et al., 2002b) was identified as a significant hit supporting the validity of this approach. In addition to Hrh1, several loci were predicted to be associated with Bphs susceptibility (data not shown).

Recently, a quantitative trait gene prediction tool has been described that utilizes functional genomics information (gene co-expression, protein-protein binding data, ontology annotation and other functional data) to rank candidate genes within large QTLs associated with respective phenotype (Guan et al., 2010). This methodology has validated several of the GWAS/QTL mapping findings and discovered novel associations with phenotype of interest. We utilized this approach for a complex phenotype (Bphs) that is believed to involve cardiac, vascular and anaphylactic responses (Munoz et al., 1965; Ma et al., 2002b; Wang et al., 2014). Because the selection of phenotype-associated gene sets is critical for final gene predictions, several terms were used to incorporate sub-phenotypes equivalent to Bphs in the expectation that use of multiple phenotype terms would help identify candidate loci for Bphse. Integration of functional predictions with genetic association (Scg, Figure C.2) allowed us to focus on only those candidates that reached significance in both approaches. Importantly, Hrh1, one of the genes implicated in Bphs
susceptibility is among the the six candidates that were identified. Interestingly, the location of all ten predicted candidates around 113-116Mb supports the presence of a functional LD domain containing *Hrh1.Bphse* and *Histh* that regulate overall responsiveness to histamine. Several of the top candidates have potential relevance to phenotypes associated with Bphs, including anaphylaxis and mast cell degranulation, G protein coupled signaling, potential role as chaperones, and cardiovascular effects (Table C.5). Among them, peroxisome proliferator-activated receptor-gamma (*Pparg*), is the only gene that together with *Hrh1* is functionally enriched in all sub-phenotypes used to define Bphs. *Pparg* encodes a nuclear receptor protein belonging to the peroxisome proliferator-activated receptor (Ppar) family. Activation of PPARg suppresses mast cell maturation and is involved in allergic disease (Tachibana et al., 2008; Ushio et al., 2011). Because mast cells are major drivers of pathological events in anaphylaxis (Lieberman et al., 2016), identification of *Pparg* may be highly relevant to Bphs. In addition, increased PPARg expression is associated with cardiac dysfunction (Sikder et al., 2018), one of the sub phenotypes associated with anaphylactic shock following histamine release (Wang et al., 2014). It will be interesting to quantify the mRNA expression of some of these candidates between susceptible and resistant strains and whether they interact with the H1R haplotype. Taken together these data support that allelic co-adaptations within a functional *Hrh1* LD domain encompasses a *Bphse* gene in some wild-derived strains that is capable of complementing Bphs.
C.5. Materials and Methods

C.5.1. Animals.

AKR/J (AKR), BPL/1J, C3H/HeJ (C3H), C3H/HeN, CAST/EiJ, C57BL/6J (B6), CBA/J (CBA), CBA/N, CZECHII/EiJ, I/LnJ, JF1/MsJ, MOLD/EiJ, MOLF/EiJ (MOLF), MRL/MpJ (MRL), MSM/Ms, PWD/PhJ, PWK/PhJ (PWK), RF/J, SF/CamEiJ, and SKIVE/EiJ were purchased from the Jackson Laboratory (Bar Harbor, Maine). B6.129P-Hrh1\textsuperscript{tm1Wat} (H1RKO) (Banu et al., 1999), C3H.SJL-Hrh1\textsuperscript{Bphs-s} (C3H.Bphs\textsuperscript{s}) (Ma et al., 2002a), (B6 × H1RKO)\textsubscript{F1}, (C3H × H1RKO)\textsubscript{F1}, (CBA × H1RKO)\textsubscript{F1}, (AKR × H1RKO)\textsubscript{F1}, (MRL × H1RKO)\textsubscript{F1}, (AKR × PWK)\textsubscript{F1}, (C3H × MOLF)\textsubscript{F1}, (MOLF × H1RKO) × H1RKO, (AKR × PWK) × AKR, (C3H × MOLF) × C3H, C3H.Bphs\textsuperscript{MOLF+/−}, C3H.Bphse\textsuperscript{C3H}, AKR.Bphse\textsuperscript{PWK+/−} and AKR.Bphse\textsuperscript{AKR} were generated and maintained under specific pathogen free conditions in the vivarium of the Given Medical Building at the University of Vermont according to National Institutes of Health guidelines. All animal studies were approved by the Institutional Animal Care and Use Committee of the University of Vermont.

C.5.2. Bphs Phenotyping.

Bphs phenotyping was carried out as previously described (Ma et al., 2002a). Briefly, mice were injected with purified PTX (List Biological Laboratories, Inc.) in 0.025 M Tris buffer containing 0.5 M NaCl and 0.017% Triton X-100, pH 7.6. Control animals received carrier. Three days later, mice were challenged by injection with histamine (milligrams per kilogram of body weight [dry weight], free base) suspended in phosphate-
buffered saline (PBS). Deaths were recorded at 30 min post-challenge. The results are expressed as the number of animals dead over the number of animals studied.

**C.5.3. DNA sequencing of third intracellular loop of Hrh1.**

DNA for 91 inbred laboratory and wild-derived strains of mice was purchased from the Mouse DNA resource at Jackson laboratories (www.jax.org) and used in an *Hrh1* specific PCR reaction using the following primer sets: forward-740F, 5’-TGCCAAGAAACCTGGGAAAG-3’, and reverse-1250R, 5’-CAACTGCTTGCTGCTCCTC-3’. Thermocycling was carried out for a 15 µl reaction mix with 2 mM MgCl₂, 200 µM dNTPs, 0.2 µM primers, 1 unit of Taq polymerase and ~50 ng of genomic DNA together with an initial 2-min 97°C denaturation followed by 35 cycles of 97°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec. The final extension was for 5 min at 72°C. *Hrh1* amplicons from each mouse strain were gel purified (Qiagen Cat# 28115) and DNA sequencing reactions were performed with the BigDye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) using 740F or 1250F reverse primers. The reaction products were resolved on an ABI Prism 3100 DNA sequencer at the DNA analysis facility at University of Vermont. DNA sequencing data were assembled and analyzed using MultiAlign (Corpet, 1988). Each potential nucleotide sequence polymorphism was confirmed by comparing it with the actual chromatographic profiles using Chromas v2.6.5 (https://technelysium.com.au/wp/)

**C.5.4. DNA isolation and Genotyping.**

DNA was isolated from mouse tail clippings as previously described (Sudweeks et al., 1993). Briefly, individual tail clippings were incubated with cell lysis buffer (125
mg/ml proteinase K, 100 mM NaCl, 10 mM Tris-HCl (pH 8.3), 10 mM EDTA, 100 mM KCl, 0.50% SDS, 300 ml) overnight at 55°C. The next day, 6M NaCl (150 ml) was added followed by centrifugation for 10 min at 4°C. The supernatant layer was transferred to a fresh tube containing 300 µl isopropanol. After centrifuging for 2 min, the supernatant was discarded, and the pellet was washed with 70% ethanol. After a final 2 min centrifugation, the supernatant was discarded, and DNA was air dried and resuspended in TE. Genotyping was performed using microsatellite, sequence specific and Hrh1 primers. Microsatellite primers: Polymorphic microsatellites were selected to have a minimum polymorphism of 8bp for optimal identification by agarose gel electrophoresis. Briefly, primers were synthesized by IDT-DNA (Coralville, IA) and diluted to a concentration of 10 µM. PCR amplification was performed using Promega GoTaq according standard conditions and amplicons were subjected to 2% agarose gel electrophoresis and visualized by ethidium bromide and UV light. Sequence-specific primers: Genotyping was performed using sequence specific primers that differ only at the 3’ nucleotide corresponding to each allele of the identified SNP (Bunce et al., 1995). Each primer set was designed using Primer3 to have a Tm of 58-60°C and synthesized by IDT-DNA (Coralville, IA) and used at a concentration of 100 µM. PCR reactions were subjected to cycling conditions as described and if found to be necessary, the annealing temperature at each stage was adjusted to accommodate the optimal Tm. Amplicons were electrophoresed with 10 µl Orange G loading buffer on a 1.5% agarose gel stained with ethidium bromide and visualized by UV light. The presence of a SNP specific allele was scored by observing an amplicon of the expected size in either reaction.
C.5.5. H1RKO mice genotyping.

Wild-type and Hrh1/- alleles were genotyped as previously described (Ma et al., 2002a). Approximately 60 ng of DNA was amplified (GeneAmp PCR system 9700, Applied Biosystems, Foster City, CA). The DNA was amplified by incubation at 94°C for 3 min followed by 35 cycles of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 30 sec. At the end of the 35 cycles, the DNA was incubated at 72°C for 10 min and 4°C for 10 min. The amplified DNA was analyzed by gel electrophoresis in a 1.5% agarose gel. The DNA was visualized by staining with ethidium bromide.

C.5.6. Linkage analysis and generation of Bphse congenic.

Segregation of genotype frequency differences with susceptibility and resistance to Bphs in (MOLF × H1RKO) × H1RKO and (AKR × PWK) × AKR mice were tested by χ² in 2 × 2 contingency tables. C3H.BphseMOLF+/-, C3H.BphseC3H, AKR.BphsePWK+/- and AKR.BphseAKR congenic mice were derived by marker assisted selection. (AKR × PWK) × AKR and (C3H × MOLF) × C3H mice that were heterozygous across the Bphse interval at N2 and at each successive BC generation were selected for continued breeding. Bphse congenic mice were maintained as heterozygotes.

C.5.7. Targeted genetic association testing.

Genotype data (SNPs in both coding and non-coding) of 50 mouse strains (Supplementary Table C.1) that were phenotyped for Bphs either by us or described in the literature (Munoz et al., 1953; Wardlaw, 1970; Linthicum et al., 1982; Gao et al., 2003b; Diehl et al., 2014), was retrieved from public databases at the Sanger Institute (https://www.sanger.ac.uk/science/data/mouse-genomes-project) and The Jackson
Laboratory (https://phenome.jax.org/). The lack of representation of genotype data from BPL/1J, C3H/HeN, CZECHII/EiJ, JF1/Ms, MOLD/EiJ, MOLF/EiJ, MRL/MpJ, MsM/Ms, NU/J, PWD/J, RJ/J, SF/CamiJ, SKIVE/EiJ and SJL/J in these databases were compensated by the following:

The Chromosome Region Capture Sequencing

*Fragment DNA.* 3 μg of genomic DNA was sheared into fragments of approximately 200 bp with the Covaris E220 system (Covaris, USA), and purification was performed with 1.4-fold volume of AMPure XP Beads (Beckman, USA).

*DNA Library Construction.* After the purification of the sheared DNA, the library was constructed with SureSelect Library Prep Kit (Agilent, USA). End-repair was performed with the volume of 10× End Repair Buffer, dNTP Mix, T4 DNA Polymerase, Klenow DNA Polymerase and T4 Polynucleotide Kinase, reacted at 20 °C for 30 min, and 1.8× the volume of AMPure XP Beads was added for purification; adding A at the 3’ end was performed with the volume of 10× Klenow Polymerase Buffer, dATP and Exo(-) Klenow, reacted at 37 °C for 30 min. T4 DNA Ligase Buffer, SureSelect Adaptor Oligo Mix and T4 DNA Ligase were added and reacted at 20 °C for 15 min. Adaptor-ligated library was amplified through added in SureSelect Primer, SureSelect ILM Indexing Pre Capture PCR Reverse Primer, 5× Herculase II Rxn Buffer, 100 mM dNTP Mix and Herculase II Fusion DNA Polymerase, and the reaction procedure is: 98 °C pre-denaturation 2 min, 98 °C denaturation 30 Sec, 65 °C annealing 30 Sec, 72 °C extension 30 Sec, amplified for 4 rounds. Purification was performed with 1.8X Agencourt AMPure
XP beads after each enzymatic reaction. The adaptor-ligated library around range of 225 ~ 275bp was finally obtained.

*Hybridization capture.* Prepared library was executed hybridization capture experiment with the SureSelect Target Enrichment Kit (Agilent, USA). The prepared library reacted with SureSelect Block Mix in 95 °C 5min, followed by maintaining in 65 °C and then Hybridization Buffer, capture library mix was added in and reacted at 65 °C 24hrs, finally Dynabeads M-280 streptavidin (Life, USA) was used for the enrichment of the Captured DNA library (Gnirke *et al.*, 2009; Mamanova *et al.*, 2010).

*Index amplification.* 5×Herculase II Rxn Buffer, 100 mM dNTP Mix, SureSelect ILM Indexing Post Capture Forward PCR Primer and Herculase II Fusion DNA Polymerase were added in the enriched captured DNA library for index amplification. The reaction procedure is 98°C Pre-denaturation 2 Min, 98 °C denaturation 30 Sec, 57 °C annealing 30 Sec, 72 °C extension 30 Sec, amplification 12 rounds, followed by the purification using 1.8 times the volume of AMPure XP Beads. A Sequencing library of 250-350 bp range was finally obtained (Gnirke *et al.*, 2009).

*Sequencing.* A 10 ng library was used for cluster generation in cBot with the TruSeq PE Cluster Kit (illumina, USA) followed by bidirectionally sequenced in Illumina Hiseq 2500 to obtain the data of 2x150 bp.

Whole-genome sequencing
**DNA Library Construction.** For whole-genome sequencing, DNA libraries was constructed according to Illumina recommended protocols. Briefly, 3 µg of genomic DNA was sheared into fragments of approximately 300-400 base pairs with the Covaris E220 system, followed by end-repair, A-tailing, and ligation of the Illumina multiplexing PE adaptors. Purification was performed with 1.8X Agencourt AMPure XP beads after enzymatic reactions. An agarose gel electrophoresis with a concentration of 2% to separate DNA products was performed, and DNA fragments with a length between 300 and 400 bp were recycled and purified according to the user guide of Qiagen Gel Extraction Kit. A PCR enrichment experiment was performed to ensure that DNA products to be successfully sequenced was enough.

**Library inspection.** After construction of the library, preliminary quantification was performed using Qubit 2.0, and the library was diluted to 1 ng/ul, and then the insert size of the library was detected using Agilent 2100. If the insert size was as expected, Q-PCR was performed to accurately quantify the effective concentration of the library (library effective concentration >10 nM) to ensure library quality.

**Sequencing.** Finally, these DNA fragments were subjected to the Illumina Hiseq 2000 platform for pair-end sequencing (PE150). The raw image data files obtained by high-throughput sequencing (Illumina) were converted into Sequenced Reads by CASAVA, and the results were stored in FASTQ format. The read length was 150 bp.

**Data processing and analysis.** To ensure the quality of subsequent information analysis, the original sequence was filtered with the software SolexaQA to get high quality
Clean Reads (Cox et al., 2010). Efficient high-quality sequencing data was mapped to the reference genome mm10 by BWA software (Li et al., 2009a), samtools (Li et al., 2009b) was used for sorting, picard tools was used for duplication, and GATK was used for Indel Realignment and Base Recalibration (McKenna et al., 2010). Finally, HaplotypeCaller of GATK is used for mutation detection. The VCF format file was filtered with VCFtools (Danecek et al., 2011). The SNP filtered results of each sample are annotated by ANNOVAR software (Wang et al., 2010), which mainly includes three aspects: annotation based on gene, genomic region and function. GATK software was also used to detect InDel, and pindel was used to detect SV (Ye et al., 2009), which is divided into four types: deletions (>5bp), Insertions (>5bp), inversions and tandem duplication. The VCF format files were converted to Plink files with VCFtools.

All these data sources were collated to generate genotype information for 13,257 SNPs across the Bphse locus on Chr 6. To calculate associations between genetic polymorphisms and Bphs, we used efficient mixed-model association (EMMA) (Kang et al., 2008). This method treats genetic relatedness as a random variable in a linear mixed model to account for population structure, thereby reducing false associations between SNPs and the measured trait. We used the likelihood ratio test function (emma.ML.LRT) to generate p-values. Significance was assessed with Bonferroni multiple correction testing. The -log transformed p-values were plotted using GraphPad Prism7 and genomic coordinates included for each SNP using the latest mouse genome build GRCm38.p5/mm10.
C.5.8. Trait-related gene sets.

The positional candidate genes were ranked based on their predicted association with six functional terms related to the Bphs phenotype: “cardiac”, “G-protein coupled receptor”, “histamine”, “pertussis toxin”, “type I hypersensitivity”, and “vascular permeability” and “ER/EMC/ERAD” Gene Weaver (Baker et al., 2012) was used to identify genes annotated with each term. Each term was entered into Gene Weaver homepage (https://geneweaver.org)and search restricted to human, rat, and mouse genes, and to curated lists only. Mouse homologs for each gene were retrieved using batch query tool in MGI (http://www.informatics.jax.org/batch_data.shtml). In addition, using Gene Expression Omnibus (GEO) and PubMed additional gene expression data sets were retrieved for each phenotype term. Final gene lists consisted of the unique set of genes associated with each process term.

C.5.9. Functional enrichment and ranking of Bphs associated genes.

We associated genes with Bphse-related functions as described in Tyler et al. (Tyler et al., 2019). Briefly, we used the connectivity weights in the Functional Network of Tissues in Mouse (FNTM) (Goya et al., 2015) as features for training support vector machines. Each feature consisted of the connection weights from a given gene to genes in the functional module. To improve classification and reduce over-generalization we clustered each functional gene set into modules each less than 400 genes (Guan et al., 2010). For each of these modules, we trained 100 SVMs to classify the module genes from a balanced set of randomly chosen genes from outside the module. We used 10-fold cross validation and a linear kernel. We also trained each SVM over a series of cost parameters
identified by iteratively narrowing the cost parameter window to identify a series of eight cost parameters that maximized classification accuracy. We then used the train modules to score each positional candidate gene in the Bphse locus. To compare scores across multiple trained models, we converted SVM scores to false positive rates.

C.5.10. Combined gene score.

To create the final ranked list of positional candidate genes, we combined the SNP association scores with the functional predictions derived from the SVMs. We scaled each of these scores by its maximum value across all positional candidates and summed them together to derive a combined gene score ($S_{cg}$) that incorporated both functional predictions and genetic influence:

$$S_{cg} = \frac{-\log_{10}(p_{EMMA})}{\max_{\text{pos.cand.}}(-\log_{10}(p_{EMMA}))} + \frac{-\log_{10}(FPR_{SVM})}{\max_{\text{pos.cand.}}(-\log_{10}(FPR_{SVM}))},$$

where the denominators of the two terms on the right-hand side are the maximum values of $-\log_{10} (p_{EMMA})$ and $-\log_{10} (FPR_{SVM})$ over all positional candidates in Bphse, respectively, which normalizes the functional and positional scores to be comparable to each other. SNPs were assigned to the nearest gene within 1Mb. If more than one SNP was assigned to a gene, we used the maximum negative $\log_{10}$ $p$-value among all SNPs assigned to the gene.
C.6. Acknowledgements

ALT and JMM are supported by a grant (R21 LM012615) from the National Library of Medicine of the United States National Institutes of Health (NIH). AR, DNK, EPB, and CT were supported by grants from the NIH and the National Multiple Sclerosis Society (NMSS). DNK was supported by NIH grants from the National Institute of Neurological Disease and Stroke (R01 NS097596), National Institute of Allergy and Infectious Disease (R21 AI145306), and the NMSS (RR-1602-07780).
C.7. References


an Autoimmune Disease Locus, as Histamine Receptor H1. *Science*, 297(5581), 620-623.


C.8. Figures

<table>
<thead>
<tr>
<th>Strain</th>
<th>D6Mit74 (48,726,556)</th>
<th>D6Mit35 (59,333,905)</th>
<th>D6Mit3722 (128,834,894)</th>
<th>Bphs</th>
<th>% Aff</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3H/HeJ</td>
<td>C C C C C</td>
<td></td>
<td></td>
<td>4/30</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>MOLF/MpJ</td>
<td>M M M M M</td>
<td></td>
<td></td>
<td>20/22</td>
<td>90</td>
<td>&lt;0.0001a</td>
</tr>
<tr>
<td>C3H.Bphse\textsuperscript{C3H/C3H}</td>
<td>C C C C C</td>
<td></td>
<td></td>
<td>7/50</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>C3H.Bphse\textsuperscript{C3H/MOLF}</td>
<td>C He He He C</td>
<td></td>
<td></td>
<td>19/32</td>
<td>60</td>
<td>&lt;0.0002a</td>
</tr>
<tr>
<td>AKR/J</td>
<td>A A A A A</td>
<td></td>
<td></td>
<td>5/30</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>PWK/J</td>
<td>P P P P P</td>
<td></td>
<td></td>
<td>15/18</td>
<td>83</td>
<td>&lt;0.0001b</td>
</tr>
<tr>
<td>AKR.Bphse\textsuperscript{AKR/AKR}</td>
<td>A A A A A</td>
<td></td>
<td></td>
<td>8/47</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>AKR.Bphse\textsuperscript{AKR/PWK}</td>
<td>A He He He A</td>
<td></td>
<td></td>
<td>35/40</td>
<td>88</td>
<td>&lt;0.0001b</td>
</tr>
<tr>
<td>C3H+AKR</td>
<td></td>
<td>9/60</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>MOLF+PWK</td>
<td></td>
<td>35/40</td>
<td></td>
<td></td>
<td>88</td>
<td>&lt;0.0001c</td>
</tr>
</tbody>
</table>

\textit{Bphse\textsuperscript{C3H/C3H}+Bphse\textsuperscript{AKR/AKR}}

|                   |                      | 15/97                |                         | 15   |       | <0.0001c|

\textit{Bphse\textsuperscript{C3H/MOLF+Bphse\textsuperscript{AKR/PWK}}}

|                   |                      | 54/72                |                         | 75   |       | <0.0001c|

\textsuperscript{a}Relative to MOLF/MpJ; \textsuperscript{b}Relative to PWK/J; \textsuperscript{c}Relative to C3H/J and AKR/J; C=C3H, M=MOLF, A=AKR, P=PWK and He=heterozygous.

Figure C.1. Congenic mapping of \textit{Bphse} confirms linkage to \textit{Bphs/Hrh1}.
Figure C.2. Integration of genetic and functional mapping approaches to predict candidates for Bphse.
(A). The plot shows negative log-transformed false positive rate of functional enrichment on the y-axis. The x-axis denotes the corresponding negative log-transformed genetic association scores. Both scores were scaled by their maximum value for better comparison. Significance thresholds are marked as red dotted line. (B). Candidates that reached significance in both genetic and functional approaches are displayed using a final gene score ($S_{cg}$) on y-axis and the genomic location on x-axis. Higher gene scores indicate better candidates. The boundaries of Bphs and
Bphse quantitative trait loci are shown to indicate the linkage of Bphse predicted candidates with Hrh1.
C.9. Tables

Table C.1. Distribution of $H_1R^S$ and $H_1R^R$ alleles in inbred laboratory and wild-derived mouse strains.

<table>
<thead>
<tr>
<th>Susceptible Haplotype</th>
<th></th>
<th>Resistant Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Pro$^{263}$, Val$^{312}$, Pro$^{330}$)</td>
<td>(Leu$^{263}$, Met$^{312}$, Ser$^{330}$)</td>
<td></td>
</tr>
<tr>
<td>129X1/SvJ</td>
<td>C57BR/cdJ</td>
<td>P/J</td>
</tr>
<tr>
<td>129S1/SvImJ</td>
<td>C57L/J</td>
<td>PANCEVO/EiJ</td>
</tr>
<tr>
<td>129T2/SvEmSJ</td>
<td>C58/J</td>
<td>PERA/EiJ</td>
</tr>
<tr>
<td>A/HeJ</td>
<td>CALB/RkJ</td>
<td>PERC/EiJ</td>
</tr>
<tr>
<td>A/J</td>
<td>CE/E</td>
<td>PL/J</td>
</tr>
<tr>
<td>A/WySnJ</td>
<td>DBA/1J</td>
<td>RBF/DnJ</td>
</tr>
<tr>
<td>ALR/LtJ</td>
<td>DBA/2J</td>
<td>RIIS/J</td>
</tr>
<tr>
<td>ALS/LtJ</td>
<td>DDY/JcI3SidSeyFrkJ</td>
<td>SB/LeJ</td>
</tr>
<tr>
<td>B10.S/DvTee</td>
<td>EL/SuzSeFrkJ</td>
<td>SEA/GnJ</td>
</tr>
<tr>
<td>B10.S/McdgJ</td>
<td>FVB/NCr</td>
<td>SEC/1ReJ</td>
</tr>
<tr>
<td>BALB/cByJ</td>
<td>IS/CamRkJ</td>
<td>SENCARA/PtJ</td>
</tr>
<tr>
<td>BALB/cJ</td>
<td>KK/HIJ</td>
<td>SENCARB/PtJ</td>
</tr>
<tr>
<td>BDP/J</td>
<td>LEWES/EiJ</td>
<td>SENCARC/PtJ</td>
</tr>
<tr>
<td>BPH/2J</td>
<td>LG/J</td>
<td>SJL/J</td>
</tr>
<tr>
<td>BPL/1J</td>
<td>LP/J</td>
<td>SJL/BmJ</td>
</tr>
<tr>
<td>BPN/3J</td>
<td>MA/MyJ</td>
<td>SM/J</td>
</tr>
<tr>
<td>BTBRT+</td>
<td>MOR/RkJ</td>
<td>SPRET/EiJ</td>
</tr>
<tr>
<td>BSXB/MpJ</td>
<td>NOD/LtJ</td>
<td>ST/BJ</td>
</tr>
<tr>
<td>C57BL/10J</td>
<td>NON/LtJ</td>
<td>SWR/J</td>
</tr>
<tr>
<td>C57BL/10SnJ</td>
<td>NOR/LtJ</td>
<td>SWXL-4/TyJ</td>
</tr>
<tr>
<td>C57BL/6ByJ</td>
<td>NZB/BINJ</td>
<td>TIRANO/EiJ</td>
</tr>
<tr>
<td>C57BL/6J</td>
<td>NZO/HILtJ</td>
<td>YBR/EiJ</td>
</tr>
<tr>
<td>C57BLKS/J</td>
<td>NZW/LacJ</td>
<td>ZALENDE/EiJ</td>
</tr>
</tbody>
</table>

292
Table C.2. Bphs susceptibility of mice with the H₁R<sup>R</sup> allele.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Histamine (mg/kg)</th>
<th>Total</th>
<th>% Aff</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>C3H/HeJ</td>
<td>0/3</td>
<td>0/2</td>
<td>0/2</td>
<td>0/7</td>
</tr>
<tr>
<td>C3H/HeN</td>
<td>0/2</td>
<td>0/2</td>
<td>0/2</td>
<td>0/6</td>
</tr>
<tr>
<td>C3H</td>
<td>0/5</td>
<td>0/4</td>
<td>0/4</td>
<td>0/13</td>
</tr>
<tr>
<td>AKR/J</td>
<td>1/3</td>
<td>0/2</td>
<td>0/2</td>
<td>1/7</td>
</tr>
<tr>
<td>BPL/1J</td>
<td>1/2</td>
<td>2/2</td>
<td>2/2</td>
<td>5/6</td>
</tr>
<tr>
<td>CAST/EiJ</td>
<td>1/3</td>
<td>0/3</td>
<td>0/3</td>
<td>1/9</td>
</tr>
<tr>
<td>CBA/J</td>
<td>0/3</td>
<td>0/2</td>
<td>0/2</td>
<td>0/7</td>
</tr>
<tr>
<td>CBA/N</td>
<td>0/3</td>
<td>0/2</td>
<td>0/2</td>
<td>0/7</td>
</tr>
<tr>
<td>CBA</td>
<td>0/6</td>
<td>0/4</td>
<td>0/4</td>
<td>0/14</td>
</tr>
<tr>
<td>CZECHII/EiJ</td>
<td>4/4</td>
<td>2/4</td>
<td>2/2</td>
<td>8/10</td>
</tr>
<tr>
<td>I/LnJ</td>
<td>2/7</td>
<td>0/3</td>
<td>-</td>
<td>2/10</td>
</tr>
<tr>
<td>JF1/MsJ</td>
<td>2/3</td>
<td>2/3</td>
<td>-</td>
<td>4/6</td>
</tr>
<tr>
<td>MOLD/EiJ</td>
<td>2/2</td>
<td>1/2</td>
<td>2/2</td>
<td>5/6</td>
</tr>
<tr>
<td>MOLF/EiJ</td>
<td>2/2</td>
<td>5/5</td>
<td>5/5</td>
<td>12/12</td>
</tr>
<tr>
<td>MSM/Ms</td>
<td>0/3</td>
<td>0/3</td>
<td>-</td>
<td>0/6</td>
</tr>
<tr>
<td>MRL/MpJ</td>
<td>0/3</td>
<td>0/2</td>
<td>0/2</td>
<td>0/7</td>
</tr>
<tr>
<td>PWD/PhJ</td>
<td>5/7</td>
<td>-</td>
<td>-</td>
<td>5/7</td>
</tr>
<tr>
<td>PWK/PhJ</td>
<td>2/2</td>
<td>2/2</td>
<td>2/2</td>
<td>6/6</td>
</tr>
<tr>
<td>RF/J</td>
<td>2/2</td>
<td>2/2</td>
<td>2/2</td>
<td>6/6</td>
</tr>
<tr>
<td>SF/CamEiJ</td>
<td>0/4</td>
<td>0/2</td>
<td>-</td>
<td>0/6</td>
</tr>
<tr>
<td>SKIVE/EiJ</td>
<td>2/7</td>
<td>1/6</td>
<td>0/2</td>
<td>3/15</td>
</tr>
</tbody>
</table>

<sup>a</sup>Relative to C3H mice
Table C.3. Bphs susceptibility in (H1RKO × H1RKO) F1 hybrids.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Histamine (mg/kg)</th>
<th>Total</th>
<th>% Aff</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1RKO</td>
<td>0/2 0/2 0/2 0/2 0/2 0/10 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C57BL/6J</td>
<td>3/3 3/3 3/3 2/2 1/2 12/15 90  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B6 × H1RKO) F1</td>
<td>4/4 4/4 4/4 3/4 3/3 18/19 95  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H.Bphs&lt;sup&gt;sjl&lt;/sup&gt;</td>
<td>3/3 2/2 2/2 2/2 2/2 11/11 100  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bphs&lt;sup&gt;sjl&lt;/sup&gt; × H1RKO) F1</td>
<td>2/2 2/2 2/2 2/2 2/2 10/10 100  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H/HeJ</td>
<td>1/3 0/2 0/2 0/2 0/2 1/11 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C3H × H1RKO) F1</td>
<td>0/2 0/2 0/2 0/2 0/2 0/10 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBA/J</td>
<td>0/3 0/2 0/2 0/2 0/2 0/11 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CBA x H1RKO) F1</td>
<td>0/2 0/3 0/2 0/2 0/2 0/11 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKR/J</td>
<td>1/3 0/2 0/2 0/2 0/2 1/11 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AKR x H1RKO) F1</td>
<td>0/2 0/2 0/2 0/2 0/2 0/10 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRL/MpJ</td>
<td>0/3 0/2 0/2 0/2 0/2 0/11 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MRL x H1RKO) F1</td>
<td>2/3 0/2 0/2 0/2 0/2 2/11 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWK/PhJ</td>
<td>3/3 3/3 2/2 1/2 0/2 9/12 75  0.0005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PWK x H1RKO) F1</td>
<td>3/3 2/2 2/2 1/2 1/2 11/13 85  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOLF/MpJ</td>
<td>2/2 2/2 2/2 2/2 0/2 8/10 80  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MOLF x H1RKO) F1</td>
<td>2/2 2/2 2/2 2/2 0/2 8/10 80  &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MOLF x H1RKO) x H1RKO</td>
<td>114 Aff Unaff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1R&lt;sup&gt;-/-&lt;/sup&gt;</td>
<td>54 0 54   0/54 0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1R&lt;sup&gt;MOLF-/-&lt;/sup&gt;</td>
<td>60 54 6   54/60 90% &lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Relative to H1RKO mice
Table C.4. Linkage of chromosome 6 marker loci to Bphse.

<table>
<thead>
<tr>
<th>Marker</th>
<th>bp</th>
<th>$\chi^2$</th>
<th>p-value</th>
<th>A Ho</th>
<th>A He</th>
<th>Un Ho</th>
<th>Un He</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs36385580</td>
<td>59,353,905</td>
<td>28.8</td>
<td>7.95E-08</td>
<td>28</td>
<td>52</td>
<td>65</td>
<td>20</td>
<td>165</td>
</tr>
<tr>
<td>rs38650989</td>
<td>72,592,521</td>
<td>30.6</td>
<td>3.21E-08</td>
<td>28</td>
<td>52</td>
<td>66</td>
<td>19</td>
<td>165</td>
</tr>
<tr>
<td>D6Mit186</td>
<td>73,387,511</td>
<td>29.5</td>
<td>5.49E-08</td>
<td>30</td>
<td>53</td>
<td>66</td>
<td>19</td>
<td>168</td>
</tr>
<tr>
<td>D6Mit102</td>
<td>93,463,949</td>
<td>38.2</td>
<td>6.36E-10</td>
<td>25</td>
<td>58</td>
<td>66</td>
<td>19</td>
<td>168</td>
</tr>
<tr>
<td>D6Mit65</td>
<td>101,387,523</td>
<td>42.3</td>
<td>7.92E-11</td>
<td>25</td>
<td>58</td>
<td>68</td>
<td>17</td>
<td>168</td>
</tr>
<tr>
<td>D6Mit149</td>
<td>106,005,405</td>
<td>38.5</td>
<td>5.44E-10</td>
<td>27</td>
<td>56</td>
<td>68</td>
<td>17</td>
<td>168</td>
</tr>
<tr>
<td>Hrh1</td>
<td>114,397,936</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs31698248</td>
<td>120,207,163</td>
<td>41.6</td>
<td>1.09E-10</td>
<td>26</td>
<td>56</td>
<td>69</td>
<td>16</td>
<td>167</td>
</tr>
<tr>
<td>D6Mit254</td>
<td>125,356,646</td>
<td>35.6</td>
<td>2.42E-09</td>
<td>26</td>
<td>56</td>
<td>66</td>
<td>19</td>
<td>167</td>
</tr>
<tr>
<td>rs30853093</td>
<td>125,365,703</td>
<td>34.5</td>
<td>4.32E-09</td>
<td>26</td>
<td>57</td>
<td>65</td>
<td>20</td>
<td>168</td>
</tr>
<tr>
<td>rs30662734</td>
<td>125,370,997</td>
<td>34.5</td>
<td>4.32E-09</td>
<td>26</td>
<td>57</td>
<td>65</td>
<td>20</td>
<td>168</td>
</tr>
<tr>
<td>rs36868180</td>
<td>127,629,804</td>
<td>32.7</td>
<td>1.06E-08</td>
<td>27</td>
<td>56</td>
<td>65</td>
<td>20</td>
<td>168</td>
</tr>
<tr>
<td>D6Mit135</td>
<td>128,834,894</td>
<td>29.2</td>
<td>6.53E-08</td>
<td>27</td>
<td>56</td>
<td>63</td>
<td>22</td>
<td>168</td>
</tr>
</tbody>
</table>

Segregation of genotype frequency differences with susceptibility (Affected=A) and resistance (unaffected=Un) to Bphs in (AKR × PWK) × AKR mice were tested by $\chi^2$ in 2 × 2 contingency tables. He=AKR/PWK allele, Ho=AKR allele
Table C.5. Predicted candidates for *Bphse* and their annotation in Bphs functional networks.

<table>
<thead>
<tr>
<th>Gene</th>
<th>bp</th>
<th>Cardiac</th>
<th>GPCR</th>
<th>HA</th>
<th>PTX</th>
<th>T1H</th>
<th>V.P</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 <em>Gt(ROSA)26Sor</em></td>
<td>113074173</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2 <em>Atp2b2</em></td>
<td>113893222</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3 <em>Hrh1</em></td>
<td>114440616</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4 <em>Atg7</em></td>
<td>114751855</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5 <em>Syn2</em></td>
<td>115208454</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6 <em>Timp4</em></td>
<td>115247048</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7 <em>Pparg</em></td>
<td>115425675</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8 <em>Mkrn2</em></td>
<td>115612263</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9 <em>Raf1</em></td>
<td>115647351</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10 <em>Ift122</em></td>
<td>115890084</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

GPCR = G-protein coupled receptors, HA = Histamine, PTX = Pertussis toxin, T1H = Type 1 hypersensitivity/anaphylaxis, V.P = vascular permeability and ER = Endoplasmic reticulum/endoplasmic membrane protein complex, and endoplasmic reticulum-associated degradation.
### C.10. Supplementary data

#### Supplementary Table C.1. List of 50 inbred mouse strains used in genetic association testing.

<table>
<thead>
<tr>
<th></th>
<th>Bphs Susceptible (41)</th>
<th>Bphs Resistant (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>129S1/SvImJ</td>
<td>C57BL/6J</td>
<td>LP/J</td>
</tr>
<tr>
<td>129T2/SvEmS</td>
<td>C57BLKS/J</td>
<td>MA/MyJ</td>
</tr>
<tr>
<td>129X1/SvJ</td>
<td>C57BR/cdJ</td>
<td>MOLD/RkJ</td>
</tr>
<tr>
<td>A/J</td>
<td>C57L/J</td>
<td>MOLF/EiJ</td>
</tr>
<tr>
<td>A/WySnJ</td>
<td>C58/J</td>
<td>NOD/ShiLtJ</td>
</tr>
<tr>
<td>BALB/cJ</td>
<td>CZECHII/EiJ</td>
<td>NON/ShiLtJ</td>
</tr>
<tr>
<td>BALB/cByJ</td>
<td>DBA/1J</td>
<td>NOR/LtJ</td>
</tr>
<tr>
<td>BPL/1J</td>
<td>DBA/2J</td>
<td>NU/J</td>
</tr>
<tr>
<td>C57BL/10J</td>
<td>FVB/NJ</td>
<td>NZB/BlNJ</td>
</tr>
<tr>
<td>C57BL/10ScNJ</td>
<td>JF1/MsJ</td>
<td>NZW/LacJ</td>
</tr>
<tr>
<td>C57BL/6ByJ</td>
<td>LG/J</td>
<td>P/J</td>
</tr>
</tbody>
</table>
Supplementary Figure C.1. Mouse family tree showing the phylogenetic relationships among 102 inbred and wild-derived inbred strains.

The seven mouse groups as described as by Petkov et al. 2004 with the strains examined in this study highlighted. Group 1, Bagg albino derivatives; Group 2, Swiss mice; Group 3, Japanese and New Zealand inbred strains; Group 4, C57/58 strains; Group 5, Castle mice; Group 6, C.C. Little DBA and related strains; Group 7, wild-derived strains.