



**HAL**  
open science

# Génétique de la Sclérose En Plaques: Héritabilité manquante et Charge génétique

Vincent Damotte

► **To cite this version:**

Vincent Damotte. Génétique de la Sclérose En Plaques: Héritabilité manquante et Charge génétique. Génétique humaine. Université Pierre et Marie Curie - Paris VI, 2013. Français. NNT: . tel-00929806

**HAL Id: tel-00929806**

**<https://theses.hal.science/tel-00929806>**

Submitted on 13 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

PRÉSENTÉE A

**L'UNIVERSITÉ PIERRE ET MARIE CURIE**

ÉCOLE DOCTORALE : Cerveau, Cognition, Comportement

Par **Vincent DAMOTTE**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : Génétique Humaine

## **GÉNÉTIQUE DE LA SCLÉROSE EN PLAQUES : Héritabilité manquante et Charge génétique**

Directeur de recherche : Bertrand FONTAINE

Soutenue le : 18 Septembre 2013

Devant la commission d'examen formée de :

M. Alain TREMBLEAU	PU	Université Pierre et Marie Curie	Président
M. Thomas BOURGERON	PU	Université Denis Diderot	Rapporteur
M. David BRASSAT	PU-PH	Université Paul Sabatier	Rapporteur
M. François CORNELIS	PU-PH	Université d'Auvergne	Examineur
Mme Isabelle COURNU-REBEIX	IR	INSERM	Encadrante
M. Bertrand FONTAINE	PU-PH	Université Pierre et Marie Curie	Directeur de thèse







## *Remerciements*

En premier lieu, je tiens à remercier :

Monsieur le **Professeur Bertrand Fontaine**, pour m'avoir accueilli au sein de votre laboratoire et pour m'avoir donné l'opportunité d'effectuer ma thèse en étroite collaboration avec d'autres équipes à travers le monde. Je suis bien conscient de la chance que j'ai eu et je vous en suis reconnaissant.

Madame le **Docteur Isabelle Cournu-Rebeix**, pour l'environnement de travail que tu m'as offert, pour ta sympathie, et pour m'avoir accordé toute ta confiance pour mener à bien ces projets.

Monsieur le **Professeur Alain Trembleau**, pour m'avoir fait l'honneur de présider mon jury de thèse.

Messieurs les **Professeurs Thomas Bourgeron** et **David Brassat**, pour avoir accepté d'être rapporteurs de cette thèse, de consacrer du temps pour juger ce travail et d'apporter votre regard critique.

Monsieur le **Professeur François Cornélis**, pour m'avoir fait l'honneur de participer à ce jury, merci pour l'intérêt que vous portez à ce travail.

Je tiens ensuite à remercier toutes les personnes de l'équipe, encore présentes ou déjà parties, avec qui j'ai passé d'agréables moments.

Je remercie tout particulièrement celles avec qui j'ai travaillé étroitement : **Bertrand** et **Isabelle**, bien sûr, mais aussi **Léna**, **Mohamed**, **Violetta** et **Claire**.

Je vais essayer de n'oublier personne : Merci à **Cécile**, **Sophie**, **Claire-So**, **Marie-Jo**, **Stéphanie**, **Nadège**, **Marie**, **Mathilde**, **Elodie**, **Carine**, **Elodie**, **Nadia**, **Servane**, **Erwan**, **Alexandre** et **Aurélié**.

Merci à tous pour ces moments passés avec vous !

Je remercie également **Pierre-Antoine** pour l'opportunité que tu m'as offert de venir passer 3 mois à UCSF, ainsi que toute l'équipe *UCSF MS Genetics* pour leur accueil et leur sympathie.

Je voudrai aussi, en plus d'**Isabelle** et **Bertrand**, adresser un merci tout particulier à **Cécile**, **Mohamed** et **Léna** : merci d'avoir pris du temps pour relire ce manuscrit !



Je veux une nouvelle fois tous vous remercier très chaleureusement, merci pour ces 4 ans que j'ai passés au laboratoire en votre compagnie !

De tout cœur, je vous souhaite à tous que la vie vous apporte tout ce que vous désirez, tant sur le plan professionnel que personnel !

Pour finir, je souhaiterai bien évidemment remercier mes **parents**, mes **sœurs** et mes **amis**: merci pour votre soutien durant ces 3 années de thèse ... et *avant* !





## *Résumé*

La Sclérose En Plaques (SEP) est une maladie multifactorielle résultant de l'interaction de facteurs environnementaux, aujourd'hui inconnus, et de facteurs génétiques.

Ces dernières années, la recherche génétique a connu de grandes avancées. En 2011, une importante étude internationale permis d'identifier 57 régions non-HLA de prédisposition à la SEP, expliquant ainsi 17% de sa composante génétique, les 83% restants étant appelés « hérédabilité manquante ».

Au cours de mon doctorat, j'ai mené ou participé à différents projets dans le but d'identifier cette hérédabilité manquante et dont les hypothèses reposaient sur l'augmentation de la taille des échantillons pour augmenter la puissance statistique des études, l'identification des variants causaux des régions génomiques mises en évidence lors des études précédentes, l'étude des régions non investiguées jusqu'à présent et des variants interagissant au sein d'un même réseau biologique et l'étude de charge génétique individuelle.

Un des projets auquel j'ai participé a permis d'augmenter le nombre de variants connus associés à la SEP à 110, hors HLA. Cependant, 77% de la composante génétique de la maladie reste encore à identifier.

D'autres hypothèses sont à explorer comme par exemple l'identification de variants de prédisposition spécifiques à une sous-population, l'identification de variants de risque localisés dans des régions déjà connues mais qui ne seraient présents que chez les individus non-porteurs des allèles de risque déjà identifiés dans ces régions.

Le défi majeur des années à venir sera de corrélérer les variants de risque à leurs effets fonctionnels et des profils génétiques spécifiques à un phénomène SEP exhaustif.



## *Abstract*

Multiple Sclerosis (MS) is a multifactorial disease, resulting from an interaction between genetics and unknown environmental factors. During the last years, the identification of several predisposing genetic polymorphisms constituted a real breakthrough in the understanding of MS genetic architecture.

In 2011, a large international study identified 57 non-MHC genomic regions associated with MS susceptibility. However, the genetic part of the disease explained by these factors was evaluated to only be 17% of total MS heritability.

During my PhD, I led or participated to several projects aiming to identify the so-called « missing heritability », based on different strategies: to increase the sample sizes in order to amplify the statistical power of the studies, to identify the causative variants within previously highlighted genomic regions, to study un-investigated genomic regions and genetic interactions between variants of a same biological pathway, to study the individual genetic burden.

One of the projects extended the number of MS non-MHC-variants to 110, increasing the known part of heritability from 17% to 23%. In order to go further in MS genetics, other hypotheses should be explored, as (a) the identification of variants specific to one sub-population (French, English ...), (b) the identification of new risk variants localized in previously highlighted regions in populations of non-carriers of known risk alleles of these regions, (c) the study of the Copy-Number Variations.

Finally, the major challenge of next years will be to correlate genetic risk variants to functional effects and to comprehensive MS Phenome.



## *Abréviations*

ADN	Acide Désoxyribonucléique
ANKRD55	Ankyrin Repeat Domain-containing protein 55
ANZgene	Australia and New Zealand MS Genetics Consortium
AUC	Area Under the ROC Curve
BHE	Barrière Hémato-Encéphalique
CAM	Cell Adhesion Molecule
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
CLDN	Claudin
CMH	Complexe Majeur d'Histocompatibilité
CNV	Copy-Number Variation
CIS	Clinically Isolated Syndrome
CPA	(1) Cellule Présentatrice d'Antigènes (Immunologie) (2) Candidate Pathway Analysis (Analyses <i>Pathways</i> )
CXCR5	C-X-C (motif) Chemokine Receptor type 5
EAE	Encéphalomyélite Auto-immune Expérimentale
EBV	Epstein-Barr Virus
EDSS	Extended Disability Status Scale
GeneMSA	Genetic Multiple Sclerosis Associations
GPC5	Glypican 5
GWAS	Genome-Wide Association Studies
GWPA	Genome-Wide Pathway Analysis
HEA	Hétérogénéité Allélique Extrême
HLA	Human Leukocyte Antigen
ICAM	Intercellular Adhesion Molecule
IFN- $\beta$	Interferon beta
IFN- $\gamma$	Interferon gamma
IL	Interleukin
IL2RA	Interleukin 2 Receptor Alpha chain
IL7RA	Interleukin 7 Receptor Alpha chain
IM	Immuno-Modulateur
IMSGC	International Multiple Sclerosis Genetics Consortium
IP	Indice de Progression
IRM	Imagerie par Résonance Magnétique
IS	Immuno-Suppresseur
ITGA4	Integrine Alpha 4
ITGAL	Integrine Alpha L
JAM	Junctional Adhesion Molecule
LCR	Liquide Céphalo-Rachidien
LFA-1	Lymphocyte Function-Associated Antigen 1
LB	Lymphocyte B



LT	Lymphocyte T
MAC-1	Macrophage-1 antigen
MAF	Minor Allele Frequency (= Fréquence d'Allèle Mineur)
MANBA	Mannosidase, Beta A, Lysosomal
MBP	Myelin Basic Protein
MGAT5	Mannosyl (alpha-1,6-)-Glycoprotein beta-1,6-N-Acetyl-glucosaminyltransferase
MSGB	Multiple Sclerosis Genetic Burden
MSSS	Multiple Sclerosis Severity Score
NFKB1	Nuclear Factor of Kappa light polypeptide gene enhancer in B-cells 1
OCLN	Occludin
OR	Odds Ratio
PP	Progressive Primaire
PSGL-1	P-Selectin Glycoprotein Ligand-1
ROC	Receiver Operating Characteristic
RPS6KB1	Ribosomal Protein S6 Kinase Beta-1
RR	Rémittente
S1P	Sphingosine-1-Phosphate
SELL	L-selectin
SELP	P-selectin
SELE	E-selectin
SEP	Sclérose En Plaques
SNC	Système Nerveux Central
SNP	Single Nucleotide Polymorphism
SOX8	SRY (sex determining region Y)-box 8
SP	Secondairement Progressive
TCR	T Cell Receptor
T <sub>FH</sub>	Lymphocyte T Follicular Helper
T <sub>H</sub>	Lymphocyte T Helper (lymphocyte T auxiliaire)
Thy-1	Thymocyte antigen 1
TNF- $\alpha$	Tumor Necrosis Factor alpha
TNFRSF6B	Tumor Necrosis Factor Receptor SuperFamily, member 6B, decoy
TYK2	Tyrosine Kinase 2
UCSF	University of California San-Francisco
VCAM1	Vascular Cell Adhesion Molecule 1
VDRE	Vitamin D Response Element
VLA-4	Very Late Antigen-4
wGRS	weighted Genetic Risk Score
ZBTB46	Zinc Finger and BTB domain containing protein 46





## *Liste des Publications*

Ce travail de thèse a donné lieu aux publications suivantes :  
En bleu, publications présentées dans cette thèse

- **Premier ou co-premier auteur :**

Noriko Isobe\*, [Vincent Damotte\\*](#), Enza Lo Re\*, Maria Ban, Derek Pappas, Lena Guillot-Noel, Isabelle Rebeix, Alastair Compston, Thomas Mack, Wendy Cozen, Bertrand Fontaine, Stephen L. Hauser, Jorge R. Oksenberg, Stephen Sawcer, Pierre-Antoine Gourraud. Genetic burden in multiple sclerosis families. *Genes Immun.* 2013 Aug 1. doi: 10.1038/gene.2013.37.

[Vincent Damotte](#), Pierre-Antoine Gourraud, Léna Guillot-Noel, Mohamed El Behi, Isabelle Cournu-Rebeix\*, Bertrand Fontaine\*. A pilot study of the role of OCLN, CLDN3 and CLDN5 genes in multiple sclerosis susceptibility and severity. (en préparation)

[Vincent Damotte](#), Léna Guillot-Noël, Nikolaos Patsopoulos, Lohith Madireddy, Mohamed El Behi, Philip De Jager, Sergio Baranzini, Isabelle Cournu-Rebeix\*, Bertrand Fontaine\*. A gene pathway analysis highlights the role of cellular adhesion molecules in Multiple Sclerosis susceptibility. (en préparation)

\* Égale contribution

- **Co-auteur :**

Babron MC, Perdry H, Handel AE, Ramagopalan SV, **Damotte V**, Fontaine B, Müller-Mihsook B, Ebers GC, Clerget-Darpoux F. Determination of the real effect of genes identified in GWAS: the example of IL2RA in multiple sclerosis. *Eur J Hum Genet.* 2012 Mar;20(3):321-5.

Lill CM, Liu T, Schjeide BM, Roehr JT, Akkad DA, **Damotte V**, Alcina A, Ortiz MA, Arroyo R, Lopez de Lapuente A, Blaschke P, Winkelmann A, Gerdes LA, Luessi F, Fernandez O, Izquierdo G, Antigüedad A, Hoffjan S, Cournu-Rebeix I, Gromöller S, Faber H, Liebsch M, Meissner E, Chanvillard C, Touze E, Pico F, Corcia P; ANZgene Consortium, Dörner T, Steinhagen-Thiessen E, Baeckman L, Heekeren HR, Li SC, Lindenberger U, Chan A, Hartung HP, Aktas O, Lohse P, Kümpfel T, Kubisch C, Epplen JT, Zetl UK, Fontaine B, Vandebroek K, Matesanz F, Urcelay E, Bertram L, Zipp F. Closing the case of APOE in multiple sclerosis: no association with disease risk in over 29 000 subjects. *J Med Genet.* 2012 Sep;49(9):558-62.



Lill CM, Schjeide BM, Graetz C, Liu T, [Damotte V](#), Akkad DA, Blaschke P, Gerdes LA, Kroner A, Luessi F, Cournu-Rebeix I, Hoffjan S, Winkelmann A, Touze E, Pico F, Corcia P, Otaegui D, Antigüedad A, Alcina A, Comabella M, Montalban X, Olascoaga J, Matesanz F, Dörner T, Li SC, Steinhagen-Thiessen E, Lindenberger U, Chan A, Rieckmann P, Hartung HP, Aktas O, Lohse P, Buttman M, Kämpfel T, Kubisch C, Zettl UK, Epplen JT, Fontaine B, Zipp F, Vandenbroeck K, Bertram L. Genome-wide significant association of ANKRD55 rs6859219 and multiple sclerosis risk. *J Med Genet.* 2013 Mar;50(3):140-3.

Faraco J, Lin L, Kornum BR, Kenny EE, Trynka G, Einen M, Rico TJ, Lichtner P, Dauvilliers Y, Arnulf I, Lecendreux M, Javidi S, Geisler P, Mayer G, Pizza F, Poli F, Plazzi G, Overeem S, Lammers GJ, Kemlink D, Sonka K, Nevsimalova S, Rouleau G, Desautels A, Montplaisir J, Frauscher B, Ehrmann L, Högl B, Jennum P, Bourgin P, Peraita-Adrados R, Iranzo A, Bassetti C, Chen WM, Concannon P, Thompson SD, [Damotte V](#), Fontaine B, Breban M, Gieger C, Klopp N, Deloukas P, Wijmenga C, Hallmayer J, Onengut-Gumuscu S, Rich SS, Winkelmann J, Mignot E. ImmunoChip study implicates antigen presentation to T cells in narcolepsy. *PLoS Genet.* 2013 Feb;9(2):e1003270

International Multiple Sclerosis Genetics Consortium, Lill CM, Schjeide BM, Graetz C, Ban M, Alcina A, Ortiz MA, Pérez J, [Damotte V](#), Booth D, Lopez de Lapuente A, Broer L, Schilling M, Akkad DA, Aktas O, Alloza I, Antigüedad A, Arroyo R, Blaschke P, Buttman M, Chan A, Compston A, Cournu-Rebeix I, Dörner T, Epplen JT, Fernández O, Gerdes LA, Guillot-Noël L, Hartung HP, Hoffjan S, Izquierdo G, Kemppinen A, Kroner A, Kubisch C, Kämpfel T, Li SC, Lindenberger U, Lohse P, Lubetzki C, Luessi F, Malhotra S, Mescheriakova J, Montalban X, Papeix C, Paredes LF, Rieckmann P, Steinhagen-Thiessen E, Winkelmann A, Zettl UK, Hintzen R, Vandenbroeck K, Stewart G, Fontaine B, Comabella M, Urcelay E, Matesanz F, Sawcer S, Bertram L, Zipp F. MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. *Brain.* 2013 Jun ;136(Pt 6) :1778 :82

International Multiple Sclerosis Genetics Consortium. Baranzini SE, Khankhanian P, Patsopoulos NA, Li M, Stankovich J, Cotsapas C, Søndergaard HB, Ban M, Barizzone N, Bergamaschi L, Booth D, Buck D, Cavalla P, Celius EG, Comabella M, Comi G, Compston A, Cournu-Rebeix I, D'alfonso S, [Damotte V](#), Din L, Dubois B, Elovaara I, Esposito F, Fontaine B, Franke A, Goris A, Gourraud PA, Graetz C, Guerini FR, Guillot-Noel L, Hafler D, Hakonarson H, Hall P, Hamsten A, Harbo HF, Hemmer B, Hillert J, Kemppinen A, Kockum I, Koivisto K, Larsson M, Lathrop M, Leone M, Lill CM, Macciardi F, Martin R, Martinelli V, Martinelli-Boneschi F, McCauley JL, Myhr KM, Naldi P, Olsson T, Oturai A, Pericak-Vance MA, Perla F, Reunanen M, Saarela J, Saker-Delye S, Salvetti M, Sellebjerg F, Sørensen PS, Spurkland A, Stewart G, Taylor B, Tienari P, Winkelmann J, Zipp F, Ivinson AJ, Haines JL, Sawcer S, DeJager P, Hauser SL, Oksenberg JR. Network-based pathway analysis in multiple sclerosis with GWAS data from 15,000 cases and 30,000 controls. *Am J Hum Genet.* 2013 May 22. pii: S0002-9297(13)00180-8. doi: 10.1016/j.ajhg.2013.04.019.



International Multiple Sclerosis Genetics Consortium, Beecham A, Beecham A, Patsopoulos N, Xifara D, Davis M, Kemppinen A, Cotsapas C, Shah T, Spencer C, Booth D, Goris A, Oturai A, Saarela J, Fontaine B, Hemmer B, Martin C, Zipp F, D'Alfonso S, Martinelli-Boneschi F, Taylor B, Harbo H, Kockum I, Hillert J, Olsson T, Ban M, Oksenberg J, Agliardi C, Alfredsson L, Alizadeh M, Anderson C, Andrews R, Søndergaard H.B., Band G, Walton A, Baranzini S, Barizzone N, Barrett J, Bellenguez C, Bergamaschi L, Bernardinelli L, Berthele A, Biberacher V, Binder T, Blackburn H, Bomfim I, Brambilla P, Broadley S, Brochet B, Brundin L, Buck D, Butzkueven H, Caillier S, Camu W, Carpentier W, Cavalla P, Celius E, Coman I, Comi G, Corrado L, Cosemans L, Cournu-Rebeix I, Cree B, Cusi D, **Damotte V**, Defer G, Delgado S, Deloukas P, Di Sapio A, Dilthey A, Donnelly P, Dubois B, Duddy M, Edkins S, Elovaara I, Esposito F, Evangelou N, Fiddes B, Field J, Franke A, Freeman C, Galimberti D, Gieger C, Gourraud P-A, Graetz C, Graham A, Grummel V, Guaschino C, Hadjixenofontos A, Hakonarson H, Halfpenny C, Hall G, Hall P, Hamsten A, Harley J, Harrower T, Hawkins C, Hellenthal G, Hillier C, Hobart J, Hoshi M, Hunt S, Jagodic M, Jelčić I, Jochim A, Kendall B, Kermode A, Kilpatrick T, Koivisto K, Konidari I, Korn T, Kronsbein H, Langford C, Larsson M, Lathrop M, Lebrun-Frenay C, Lechner-Scott J, Lee M, Leone M, Leppä V, Liberatore G, Lie B, Lill C, Lindén M, Link J, Luessi F, Lycke J, Macciardi F, Mannisto S, Manrique C, Martin R, Martinelli V, Mason D, Mazibrada G, McCabe C, Mero I-L, Moutsianas L, Myhr K-M, Nagels G, Nicholas R, Nilsson P, Piehl F, Pirinen M, Price S, Reunanen M, Robberecht W, Robertson N, Rodegher M, Rog D, Salvetti M, Schnetz-Boutaud N, Sellebjerg F, Selter R, Shaunak S, Shields S, Siffrin V, Slee M, Sorensen P.S., Sorosina M, Sospedra M, Spurkland A, Strange A, Sundqvist E, Thijs V, Thorpe J, Ticca A, Tienari P, Visser E, Vucic S, Westerlind H, Wiley J, Wilkins A, Wilson J, Winkelmann J, Frohlich I, Zajicek J, Zindler E, Ivinson A, Stewart G, Hafler D, Hauser S, Compston A, Haines J, Pericak-Vance M, De Jager P, McVean G, Sawcer S, McCauley J. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet.* 2013 Sep 29. doi: 10.1038/ng.2770. [Epub ahead of print]



# Table des matières

<b>Préambule</b> .....	<b>27</b>
<b>1. La Sclérose En Plaques : une maladie multifactorielle</b> .....	<b>31</b>
<b>1.1. Composante environnementale de la Sclérose En Plaques</b> .....	<b>33</b>
1.1.1. Arguments .....	33
1.1.2. Hypothèses .....	33
1.1.2.1. Ensoleillement et Vitamine D.....	33
1.1.2.2. Exposition virale.....	35
1.1.2.3. Autres facteurs .....	35
<b>1.2. Composante génétique de la Sclérose En Plaques</b> .....	<b>37</b>
1.2.1. Arguments .....	37
1.2.2. Hypothèses .....	39
1.2.3. Stratégies d'étude: Gènes candidats versus Pangénomique.....	41
1.2.3.1. L'approche gène candidat, ou approche dite avec à priori .....	41
1.2.3.2. L'approche criblage anonyme du génome, pangénomique, ou sans à priori.....	41
<b>1.3. Les connaissances actuelles en génétique de la Sclérose En Plaques</b> .....	<b>45</b>
1.3.1. Le locus du Complexe Majeur d'Histocompatibilité .....	45
1.3.2. Les études de liaison .....	47
1.3.3. Les études d'association.....	47
<b>1.4. Corrélation génétique - mécanismes biologiques</b> .....	<b>55</b>
1.4.1. Activation immunitaire .....	55
1.4.2. Passage de la Barrière Hémato-Encéphalique .....	57
1.4.3. La chronicité .....	59
<b>1.5. Manifestations cliniques des mécanismes biologiques</b> .....	<b>63</b>
1.5.1. Démyélinisation .....	63
1.5.2. L'inflammation chronique .....	65
1.5.3. Installation du handicap.....	67
1.5.3.1. Score de handicap .....	67
1.5.3.2. Score de sévérité.....	69
1.5.4. Les traitements de la Sclérose En Plaques.....	69
1.5.4.1. Traitements de fond de première intention .....	69
1.5.4.2. Traitements de fond de seconde intention .....	71
<b>2. Sclérose En Plaques et recherche génétique actuelle</b> .....	<b>75</b>
<b>2.1. Hypothèse 1 : Manque de puissance statistique liée à la taille des échantillons étudiés</b> .....	<b>83</b>
<b>2.2. Hypothèse 2 : Les variants identifiés ne sont pas causaux</b> .....	<b>107</b>
<b>2.3. Hypothèse 3 : Les régions génomiques non investiguées pourraient contenir des marqueurs de prédisposition</b> .....	<b>129</b>
<b>2.4. Hypothèse 4 : Variants rares et peu fréquents</b> .....	<b>155</b>
2.4.1. Arguments .....	155
2.4.2. Approches expérimentales pour l'étude des variants rares .....	157
2.4.3. Hypothèses sur le rôle des variants rares dans la prédisposition génétique.....	157
2.4.4. Limites.....	159
<b>2.5. Hypothèse 5 : L'interaction de variants d'un même réseau biologique confère un risque de développer la maladie</b> .....	<b>161</b>
2.5.1. Définition d'une analyse pathway .....	161
2.5.2. Les stratégies d'études pathway.....	163
2.5.3. Méthodes statistiques des analyses pathways .....	165
2.5.3.1. Analyse de sur-représentation (Overrepresentation analysis) .....	165
2.5.3.2. Gene-set enrichment analysis.....	165
2.5.3.3. Set-based methods .....	167





2.5.4. Considérations post-analyses .....	167
<b>2.6. Hypothèse 6 : La maladie s'explique à l'échelle individuelle .....</b>	<b>199</b>
2.6.1. Les études individuelles de prédiction .....	199
2.6.2. Les études familiales d'agrégation .....	201
2.6.3. Limites aux tests d'agrégation .....	203
<b>3. Discussion et Perspectives .....</b>	<b>221</b>
<b>3.1. Autres pistes pour identifier l'héritabilité manquante .....</b>	<b>223</b>
3.1.1. Hypothèse 1 : Stratification de la population Caucasienne .....	223
3.1.2. Hypothèse 2 : Variants flanquants cryptiques .....	225
3.1.3. Hypothèse 3 : Le rôle des Copy-Number Variations .....	229
<b>3.2. Limites des études génétiques .....</b>	<b>229</b>
<b>3.3. Rôle fonctionnel des variants de risques identifiés.....</b>	<b>231</b>
3.2.1. L'étude individuelle des polymorphismes.....	231
3.2.2. L'étude de réseaux de gènes .....	233
<b>3.3. Corrélations Génotype-Phénotype .....</b>	<b>235</b>
3.3.1. Génétique de la sévérité de la Sclérose En Plaques .....	235
3.3.2. Réponses aux traitements .....	237
3.3.3. Phénomène exhaustif de la SEP : le SEPhénomène.....	239
<b>4. Conclusion .....</b>	<b>243</b>
<b>Références bibliographiques .....</b>	<b>245</b>



## **Préambule**

La Sclérose En Plaques (SEP) est une maladie chronique, inflammatoire et démyélinisante du Système Nerveux Central (SNC), c'est à dire affectant le cerveau, la moelle épinière et les nerfs optiques.

Bien que l'origine exacte de la maladie demeure à l'heure actuelle encore inconnue, la principale hypothèse étayée par de nombreuses études évoque une réponse inflammatoire incontrôlée dirigée contre les protéines composant la myéline. La myéline forme autour des axones une gaine protectrice essentielle au passage rapide de l'influx nerveux. Lors de l'attaque inflammatoire observée dans les lésions de SEP, cette gaine est détruite. La myéline peut être resynthétisée grâce aux oligodendrocytes, cellules myélinisantes du SNC, on parle alors de remyélinisation spontanée. Dans ce cas, il y a réapparition d'un courant saltatoire sans altération des axones. Cependant, ce processus spontané de remyélinisation est absent chez certains patients et diminue au cours de la maladie, les axones démyélinisés dégèrent alors et un handicap irréversible s'installe et progresse.

La SEP est une maladie fréquente en population, avec 80 000 personnes atteintes en France et 3000 à 4000 nouveaux cas diagnostiqués chaque année. Dans le monde, le nombre de personnes atteintes est estimé entre 2 et 2,5 millions. Il s'agit de la première cause de handicap moteur non traumatique de l'adulte jeune (Compston and Coles 2002, Compston 2005).

L'âge de début de la maladie se situe en moyenne aux alentours de 30 ans, avec des extrêmes allant du jeune adolescent à l'âge adulte (Compston 2005, Milo and Kahana 2010). Bien que la maladie puisse conduire à des handicaps sévères, la SEP n'est pas une maladie mortelle, l'espérance de vie des patients étant comparable à celle de la population générale.

La SEP touchant les adultes jeunes socialement et économiquement actifs, les impacts socio-économiques de la maladie sont importants. Une étude récente (2005) a estimé le coût de prise en charge des patients SEP par la société à environ 12,5 milliards d'euros en Europe (Sobocki, Pugliatti et al. 2007). Cette estimation prend en compte les coûts directs médicaux (traitements, hospitalisations...) et non-médicaux (aménagement du logement, de la voiture...), ainsi que les coûts indirects dus à la perte de productivité des patients (diminution



précoce des capacités de travail, mise en invalidité ou départ anticipé à la retraite) et de son entourage.

La forte prévalence de la SEP, son impact socio-économique et l'absence de traitement curatif font de cette maladie un enjeu majeur de santé publique pour les années à venir.

La notion de maladie complexe et de prédisposition génétique est aujourd'hui largement admise en ce qui concerne la sclérose en plaques. Cependant, malgré une recherche active portée par de nombreuses équipes à travers le monde, l'identification des facteurs génétiques de prédisposition à la SEP n'a été que très récemment productive.

L'historique de la recherche en génétique de la SEP comporte 3 phases successives :

- dans les années 70, les études sérologiques de Jersild et al. (Jersild, Svejgaard et al. 1972) ont permis l'identification du premier variant de risque associé à la maladie et conférant le risque le plus fort, l'allèle HLA-DRB1\*1501

- puis jusqu'en 2007, la recherche en génétique, fondée sur des analyses de liaison dans des familles SEP, a été plus frustrante que fructueuse, car aucune des avancées n'a pu être confirmée par les études ultérieures

- depuis 2007, grâce à la publication d'études à très grandes échelles et l'utilisation de nouveaux polymorphismes (Single Nucleotide Polymorphisms ou SNPs), l'identification de nombreux facteurs génétiques de prédisposition à la SEP s'est concrétisée.

Le travail que j'ai effectué au cours de mon doctorat s'est donc inscrit dans la continuité des avancées de la compréhension génétique de la maladie.

Ce manuscrit présente :

- les principaux arguments et résultats étayant le rôle de la génétique dans la SEP
- les principales connaissances actuelles sur la prédisposition génétique à la maladie
- la corrélation entre les avancées génétiques et la compréhension des mécanismes biologiques mis en jeu dans la SEP

- mes hypothèses de travail sur « l'héritabilité manquante » et les résultats obtenus au cours de mon doctorat pour chacune de ces hypothèses.

**Figure 1. Répartition géographique de la Sclérose En Plaques.**  
Prévalence pour 100 000 habitants.  
De (Milo and Kahana 2010)

## **1. La Sclérose En Plaques : une maladie multifactorielle**

La sclérose en plaques est une maladie dite multifactorielle ou complexe. Elle se développe chez des sujets génétiquement prédisposés après exposition à des facteurs environnementaux.

La répartition géographique de la prévalence de la SEP n'est pas homogène (Compston 2005, Milo and Kahana 2010). Les études montrent que la SEP est plus fréquente dans les régions tempérées (supérieure à 100-200 cas pour 100 000 personnes) que dans les régions tropicales (inférieure à 5 cas pour 100 000 personnes) (**Figure 1**).

Cependant, la notion de gradient de prévalence de SEP Nord-Sud, longtemps généralisée, est aujourd'hui remise en cause. Il semble que seule la population Australienne, considérée comme homogène car descendant de la population Anglaise, présente cette répartition Nord-Sud (Nylander and Hafler 2012).

Les études épidémiologiques ont aussi montré une variation du risque individuel en fonction du lieu de naissance, de l'ethnie et de l'existence ou non d'un autre cas dans la famille (Compston and Coles 2008).

La puissance et la signification de ces études épidémiologiques sont fortement diminuées par le manque de précision et d'homogénéité des données relevées. En effet, il a été mentionné que de nombreux biais peuvent amener à conclure à de fausses différences de prévalence entre deux pays. Par exemple des critères de diagnostics différents, un accès à des médecins spécialistes plus difficile dans certains pays ou certaines régions d'un pays, et des bases de données nationales incomplètes sont autant de facteurs pouvant entraîner des chiffres de prévalence différents (Benedikz, Magnusson et al. 1994).

Tout ceci implique qu'aujourd'hui encore, il est difficile d'affirmer qui de l'environnement ou de la génétique est le responsable majeur des différences de prévalence de la maladie observées entre différentes ethnies, différents pays voire différentes régions au sein d'un même pays.



**Figure 2. Corrélation entre exposition aux ultraviolets (UV) et prévalence de la maladie en France.**

- a. Exposition aux UV par région
  - b. Prévalence de la SEP par région (nombre de patients pour 100 000 habitants)
- De (Handel, Giovannoni et al. 2010)

## 1.1. Composante environnementale de la Sclérose En Plaques

### 1.1.1. Arguments

Des études de migration de population ont montré que des personnes migrant d'une zone à fort risque vers une zone à faible risque présentent moins de risque de développer la maladie. Cependant ceci n'est pas vérifié dans le cas contraire. Des migrants provenant d'un pays à faible risque de SEP et s'installant dans un pays à fort risque de SEP tendent à garder un faible risque de développer la maladie comparable à celui de leur pays d'origine (Gale and Martyn 1995). D'autres études incluant l'âge de migration des individus se sont attachées à déterminer s'il existait un âge seuil à partir duquel le risque de développer la SEP était établi. Bien que des études semblent établir un âge seuil à environ 15 ans, d'autres n'ont pas montré de corrélation entre l'âge de migration et le risque de développer une SEP (Compston 2005, Mila and Kahana, 2010).

### 1.1.2. Hypothèses

Deux facteurs ont fait l'objet de nombreuses études et par conséquent bénéficient de plus d'arguments quant à leur implication dans le déclenchement de la SEP : la vitamine D (à mettre en relation avec l'ensoleillement) et les virus.

#### *1.1.2.1. Ensoleillement et Vitamine D*

Comme mentionné précédemment, il existe une répartition mondiale hétérogène de la prévalence de la SEP (Milo and Kahana 2010). La corrélation la plus évidente avec la latitude est l'ensoleillement. Plusieurs études ont montré l'existence d'une corrélation inverse entre le risque de SEP et l'exposition solaire avant le début de la maladie (van der Mei, Ponsonby et al. 2001, Marrie 2004, Kampman, Wilsgaard et al. 2007, Sloka, Pryse-Phillips et al. 2008). Par la suite, d'autres études ont montré l'existence d'une corrélation inverse hautement significative entre la prévalence de la SEP et le taux annuel d'UltraViolet (UV) reçu par différentes régions (Handel, Giovannoni et al. 2010, Sloka, Silva et al. 2011) (**Figure 2**).

Ce phénomène de corrélation inverse peut s'expliquer biologiquement par la production de vitamine D, l'exposition lumineuse étant nécessaire à la production de cette vitamine (Holick 2004). Certaines études ont mis en évidence l'effet protecteur que



confèrerait la vitamine D sur le développement de la SEP (Ascherio and Munger 2007). Il est admis aujourd'hui que la vitamine D agit sur la réponse immunitaire, en inhibant la production de cytokines pro-inflammatoires (Milo and Kahana 2010) et en facilitant le développement des lymphocytes T régulateurs (Ascherio and Munger 2007). De plus, certaines études menées sur le modèle murin de SEP, l'Encéphalomyélite Auto-immune Expérimentale (EAE) indiquent que la vitamine D pourrait être bénéfique dans la prévention et le traitement de la SEP (Milo and Kahana 2010).

Il est aussi intéressant de noter la présence d'un excès de naissance de patients au cours de l'été, qui pourrait être mis en relation avec le faible taux de vitamine D des femmes enceintes pendant les mois d'hiver (Fernandes de Abreu, Babron et al. 2009).

#### *1.1.2.2. Exposition virale*

Deux hypothèses virales ont été proposées dans la SEP (Ascherio and Munger 2007):

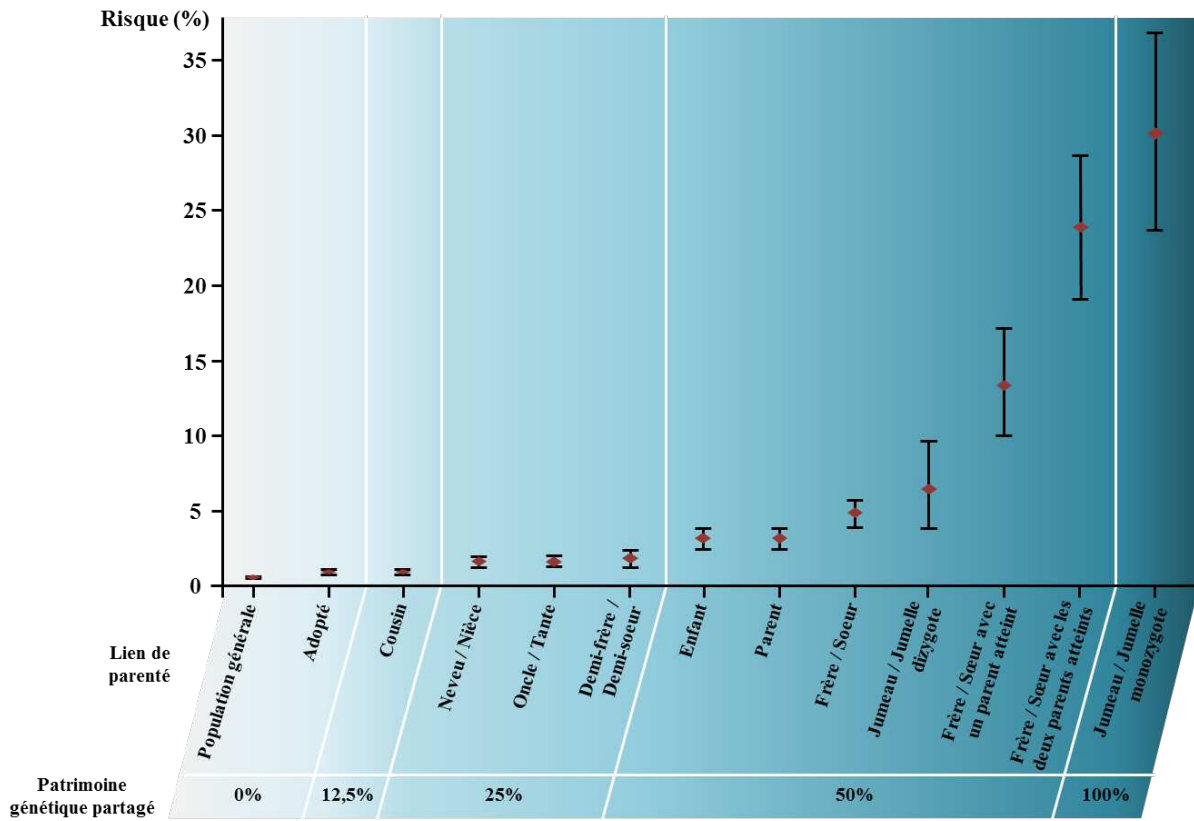
- ***l'hypothèse hygiénique*** postule qu'une succession d'infections par différents pathogènes durant l'enfance protégerait contre la maladie alors qu'un primo-contact avec ces mêmes virus à l'âge adulte déclencherait une SEP. Cette hypothèse est à l'heure actuelle la plus fédératrice.

- ***l'hypothèse dite de prévalence*** postule qu'un pathogène plus commun dans les régions présentant une forte prévalence de la maladie serait à l'origine de la maladie. Ce pathogène serait présent de manière globale et causerait une infection persistante asymptomatique, jusqu'à l'apparition de symptômes dans de rares cas plusieurs années après l'infection primaire.

De nombreux virus ont été soupçonnés de causer la maladie comme le virus de la rougeole, certains virus herpès comme l'EBV (Epstein-Barr Virus), faisant aujourd'hui l'objet de nombreuses études (Ascherio and Munger 2007, Milo and Kahana 2010). Cependant, aucun n'a fait la preuve de son implication réelle.

#### *1.1.2.3. Autres facteurs*

D'autres facteurs environnementaux sont suspectés de jouer un rôle dans le déclenchement de la maladie comme le tabac, les graisses (alimentation) ou certains composés chimiques (Ascherio and Munger 2007, Milo and Kahana 2010), mais aucune étude n'a pu aujourd'hui les impliquer de façon formelle. Certaines études semblent mettre en évidence un rôle du microbiote intestinal. Il a en particulier été démontré qu'une modification de la flore intestinale altèrerait les caractéristiques cliniques de l'EAE (Lavasani, Dzhambazov et al. 2010).



**Figure 3. Risque familial de la Sclérose En Plaques.**

Risque de développer la SEP suivant le degré de parenté avec un individu atteint.

Adapté de (Compston and Coles 2008).

La SEP est plus fréquente chez les femmes, avec 3 à 4 femmes atteintes pour 1 homme atteint (Whitacre 2001, Compston 2005). Certaines études semblent montrer que cette incidence augmentée chez les femmes correspondrait à des particularités physiologiques, en particulier hormonales (Whitacre 2001). Par ailleurs, il a été montré que ce ratio a augmenté au cours des dernières décennies, au détriment des femmes (Orton, Herrera et al. 2006). Ceci pourrait être dû à l'augmentation récente du tabagisme féminin ainsi qu'à la mise sur le marché des pilules contraceptives ou des facteurs de mécanismes épigénétiques aujourd'hui inconnus.

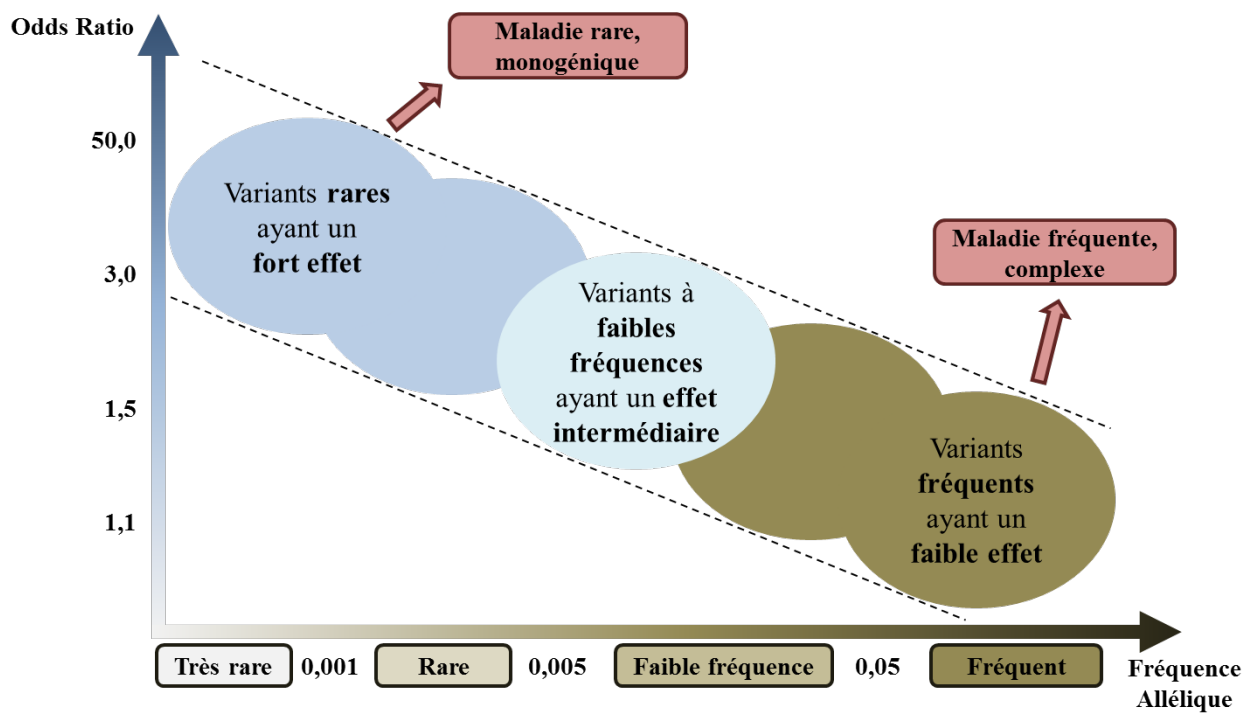
A ce jour, aucun facteur de prédisposition environnementale n'a été identifié de manière certaine. En revanche, au cours de ces dernières années, des avancées majeures dans le domaine de la génétique ont eu lieu.

## **1.2. Composante génétique de la Sclérose En Plaques**

### 1.2.1. Arguments

Sur le plan génétique, des études épidémiologiques ont montré une différence de prédisposition à la SEP suivant les ethnies. Les caucasiens sont hautement susceptibles à la maladie alors que la SEP est moins fréquente dans les populations asiatiques et africaines. Des études familiales ont montré que le risque de développer une SEP pour le(a) jumeau/jumelle d'un patient est augmenté de 25 à 30% par rapport à la population générale. Dans le cas de jumeaux dizygotes, ce risque n'augmente que d'environ 5%. D'autre part, environ 20% des patients ont au moins un apparenté atteint. Le risque de développer la maladie pour une personne apparentée au 1<sup>er</sup> degré à un patient passe de 0,1% à 5% pour les frères/sœurs, 2% pour les parents et 2% pour les enfants. Pour les individus apparentés aux 2<sup>ème</sup> et 3<sup>ème</sup> degrés, ce risque est moins élevé, proche de 1% mais reste quand même supérieur à celui de la population générale (entre 0,1% et 0,2%) (**Figure 3**).

Tous ces arguments assoient donc l'implication de facteurs génétiques dans la survenue de la SEP.



**Figure 4. Fréquence et effet des variants mis en jeu dans les maladies monogéniques et complexes.**

Adapté de (Manolio, Collins et al. 2009).

### 1.2.2. Hypothèses

Deux hypothèses sont avancées pour expliquer l'étiologie des maladies fréquentes telles que la sclérose en plaques (Sawcer 2008) (**Figure 4**) :

- l'hypothèse *maladie fréquente/variants fréquents* (*common disease/common variants hypothesis*) (Reich and Lander 2001, Pritchard and Cox 2002): la prédisposition génétique aux maladies fréquentes serait déterminée par peu de variants génétiques, fréquents en population (fréquence supérieure à 5%), qui chacun ne conférerait qu'un faible risque de développer la maladie, avec un **Odds Ratio (OR)**<sup>1</sup> compris entre 1,1 et 1,5.
- l'hypothèse *multiple variants rares* (*multiple rare variants hypothesis*) ou hypothèse d'*hétérogénéité* (Smith and Luskis 2002): la prédisposition génétique aux maladies fréquentes serait due à une combinaison de variants très peu fréquents en population (fréquence comprise entre 0,1% et 5%) mais ayant chacun un fort effet, avec un OR compris entre 1,5 et 20.

Même si au départ ces hypothèses semblaient être totalement exclusives, un modèle mixte semble aujourd'hui être l'hypothèse la plus probable. La prédisposition aux maladies fréquentes combinerait un fond commun déterminé par quelques variants fréquents et une hétérogénéité déterminée par plusieurs variants rares (Wang, Barratt et al. 2005).

Chacun de ces variants ne serait ni nécessaire ni suffisant pour développer la maladie, ce serait l'interaction entre tous ces variants qui entraînerait la prédisposition génétique à la maladie.

Des études de simulation ont permis d'estimer le nombre de variants nécessaires pour développer une maladie fréquente selon ces 2 hypothèses. Sous l'hypothèse maladie fréquente/variants fréquents, 20 à 100 variants de risques fréquents seraient nécessaires alors que sous l'hypothèse multiple variants rares, plusieurs centaines à plusieurs milliers de variants rares seraient nécessaires pour développer la maladie (Lindsey 2005, Yang, Khoury et al. 2005).

Une étude récemment publiée de Hunt et al. sème le doute quant au probable rôle des variants rares dans les maladies complexes. Les analyses d'associations exécutées sur les données de séquençage et génotypage des exons de 25 gènes associés à au moins deux maladies à composante auto-immune (parmi la SEP, le diabète de type I, la

---

<sup>1</sup> Odds Ratio : Valeur de l'augmentation du risque de développer la maladie conféré par le facteur





maladie de Crohn, la maladie cœliaque, le psoriasis, l'hypothyroïdie), comparant une population de 25 000 patients et de 17 000 contrôles, semblent indiquer que les variants rares ne joueraient pas de rôle majeur dans l'héritabilité des maladies complexes auto-immunes, qui serait ainsi due à l'interaction de plusieurs variants fréquents ayant de faibles effets (Hunt, Mistry et al. 2013).

### 1.2.3. Stratégies d'étude: Gènes candidats versus Pangénomique

Aujourd'hui deux types d'approches sont utilisés pour identifier des gènes de prédisposition aux maladies complexes :

#### *1.2.3.1. L'approche gène candidat, ou approche dite avec à priori*

Un ou quelques gènes sont sélectionnés selon leur fonction biologique d'intérêt dans la survenue de la maladie. Un petit nombre de marqueurs génétiques appartenant aux régions dans lesquelles se trouvent ce(s) gène(s) candidat(s) sont analysés et testés pour leur association avec la maladie.

Cette méthode a l'avantage d'être relativement peu coûteuse et nécessite peu de corrections statistiques car peu de marqueurs sont génotypés et analysés. Cependant cette approche ne considère que des gènes ayant une fonction biologique connue et potentiellement d'intérêt pour la maladie.

#### *1.2.3.2. L'approche criblage anonyme du génome, pangénomique, ou sans à priori*

Aujourd'hui la majorité des études pangénomiques menées pour identifier des facteurs génétiques jouant un rôle dans les maladies complexes sont des études d'association ou GWAS (Genome-Wide Association Studies). Des centaines de milliers jusqu'à plusieurs millions de SNPs (Single Nucleotide Polymorphisms, marqueurs bialléliques), répartis sur l'ensemble du génome, sont génotypés et analysés pour leur association avec un trait phénotypique particulier.

Dans le cadre des études visant à identifier des facteurs de prédisposition à une maladie, cette analyse est effectuée en comparant la fréquence des allèles de ces polymorphismes entre des populations de patients et de contrôles, constituées chacune de milliers d'individus. Les SNPs pour lesquels un allèle est significativement plus fréquent chez les patients que chez les contrôles (ayant une p-value inférieure à la p-value seuil de  $5 \times 10^{-8}$ , considérée comme significative dans les études pangénomiques) sont considérés comme étant

**Figure 5. Stratification des échantillons Caucasiens analysés lors d'une étude GWAS.**

De (IMSGC, WTCCC2 et al. 2011)

des variants de risque associés à la maladie.

En statistiques, lors des tests d'hypothèses, deux types d'erreurs doivent être prises en considération. **L'erreur de type 1**, ou **risque  $\alpha$** , correspond à la probabilité de rejeter l'hypothèse nulle lorsque celle-ci est vraie. Elle correspond aux **faux-positifs**. Elle est souvent fixée à un seuil de 5%. **L'erreur de type 2**, ou **risque  $\beta$** , correspond à la probabilité de ne pas rejeter l'hypothèse nulle lorsque celle-ci est fautive. Elle correspond aux **faux-négatifs**. La **puissance** d'une étude correspond à la probabilité de rejeter l'hypothèse nulle lorsque celle-ci est fautive. Elle est donc égale à  **$1-\beta$** . Une puissance de 80% correspond donc à un risque  $\beta$  de 20%. Du fait du nombre de tests effectués lors des études pangénomiques (environ 1 million), une **correction** du risque  $\alpha$  doit être apportée, ce qui correspond à la **valeur seuil de  $5 \times 10^{-8}$**  communément utilisée dans les études GWAS (0,05/1.000.000, correction de Bonferroni pour les tests multiples) (Risch and Merikangas 1996, Hoggart, Clark et al. 2008). Ainsi, à partir de la p-value seuil de  $5 \times 10^{-8}$ , le résultat d'association a plus de chance d'être un vrai positif qu'un faux positif.

Ce type d'approche a l'avantage de permettre d'étudier un grand nombre de marqueurs et donc de potentiellement identifier un grand nombre de gènes de prédisposition à la maladie, indépendamment de leur fonction biologique. Cependant, le fait de comparer des fréquences alléliques entre une population de patients et une population de contrôles pose le problème de la stratification de populations d'une part entre les ethnies et d'autre part entre les différentes populations au sein d'une même ethnie. Il a en effet été montré que la fréquence des allèles des SNPs varie de façon très significative selon les ethnies (International HapMap 2005). C'est pourquoi, les études se focalisent à chaque fois sur une ethnie en particulier. De même, au sein de chaque ethnie sont apportées des corrections permettant de prendre en compte les différences qu'il peut exister entre différentes populations d'une même ethnie (**Figure 5**) (Hirschhorn and Daly 2005). Ce type d'analyse est très onéreuse et nécessite d'avoir des cohortes de patients et contrôles importantes. C'est pourquoi aujourd'hui, la majorité des échantillons étudiés dans ces GWAS proviennent de la population Caucasienne, car seuls les pays riches peuvent entreprendre de telles études (Bustamante, Burchard et al. 2011).

**Figure 6. Risque génotypique relatif en fonction de la combinaison des allèles au locus HLA-DRB1.**

X/X: individus n'ayant aucun allèle associé à la SEP auxquels sont attribués arbitrairement un risque basal de 1.

X : n'importe quel allèle du locus HLA-DRB1 non associé à la SEP

15, 17, 11, 08, 14, 10, 01 : allèles du locus HLA-DRB1

De (Ramagopalan and Ebers 2009)

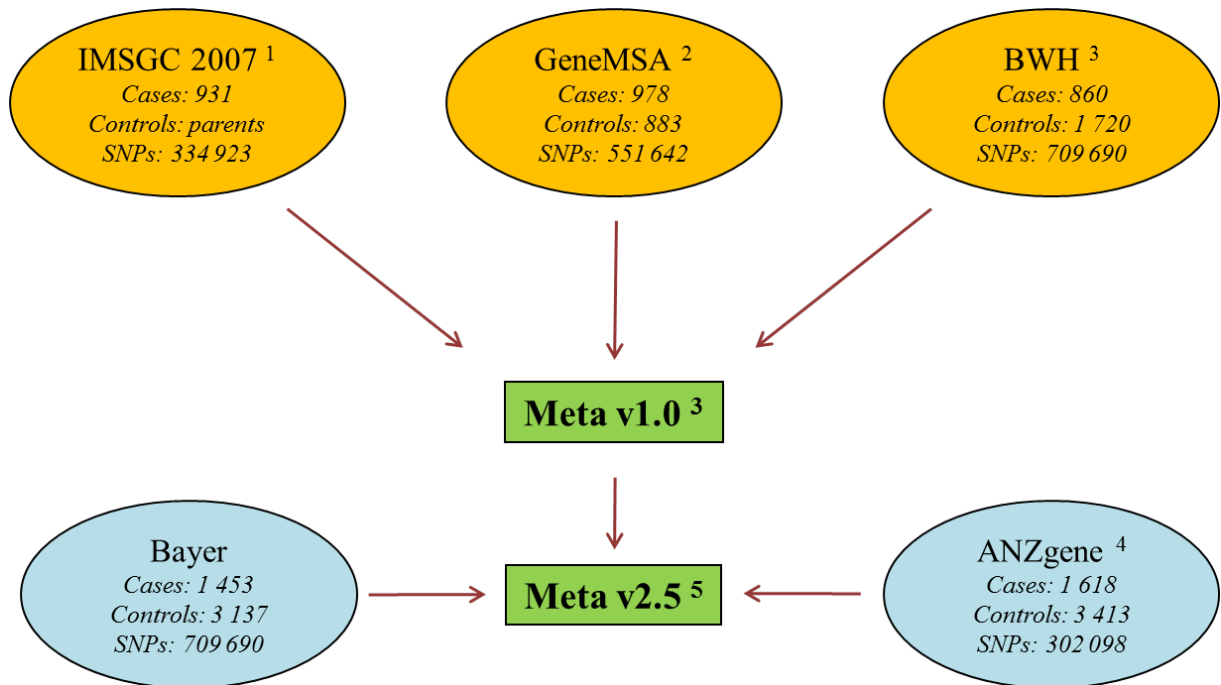
### 1.3. Les connaissances actuelles en génétique de la Sclérose En Plaques

#### *1.3.1. Le locus du Complexe Majeur d'Histocompatibilité*

L'association de la SEP avec les gènes de la région du Complexe Majeur d'Histocompatibilité (CMH) ou Human Leukocyte Antigen (HLA) est connue depuis le début des années 1970. Les gènes de cette région sont divisés en 3 classes : HLA de classe I, II et III et codent pour des glycoprotéines de surface hautement polymorphiques impliquées dans la reconnaissance immunitaire du soi et du non-soi (Gourraud, Harbo et al. 2012). Elles se trouvent à la surface des Cellules Présentatrices d'Antigènes (CPA) et permettent d'activer les Lymphocytes T (LT).

Au début des années 1970, des analyses sérologiques ciblant le CMH ont permis d'identifier une association entre la SEP et les allèles de classe I A3 (Naito, Namerow et al. 1972) et B7 (Jersild, Svejgaard et al. 1972). L'association de HLA-A3 a été montrée comme étant secondaire à celle de HLA-B7, qui elle-même a été montrée comme étant probablement secondaire à l'association des allèles de classe II HLA-DR2 et DQw6 avec la SEP (Jersild, Fog et al. 1973, Bertrams and Kuwert 1976, Compston, Batchelor et al. 1976). Aujourd'hui, l'association de l'haplotype de classe II : HLA-DQB1\*0602, HLA-DQA1\*0102, HLA-DRB1\*1501 (correspondant aux allèles sérologiques HLA-DR2, DQ6) est clairement établi (Barcellos, Oksenberg et al. 2003). De nombreuses études ont confirmé l'association de la maladie avec l'allèle HLA-DRB1\*1501 (Yaouanq, Semana et al. 1997). Les individus porteurs de cet allèle présentent un risque de développer la maladie augmenté de 3 à 4 fois, en faisant ainsi le facteur génétique majeur de prédisposition à la SEP. Cependant, l'existence d'un fort déséquilibre de liaison au sein de la région HLA rend difficile l'identification du ou des réel(s) gènes de prédisposition de l'haplotype HLA-DQB1\*0602, HLA-DQA1\*0102, HLA-DRB1\*1501 (Miretti, Walsh et al. 2005).

Il a été montré que l'allèle HLA-DRB1\*15 était sujet à des phénomènes d'interactions géniques (ou épistasies) dans la SEP. La **Figure 6** montre le risque associé à certaines combinaisons d'allèles du locus HLA-DRB1. Le risque le plus important est attribué aux homozygotes HLA-DRB1\*15, tandis que l'allèle HLA-DRB1\*14 jouerait un rôle protecteur contre la maladie. De plus, l'allèle HLA-DRB1\*08 seul n'augmente que modestement le risque de développer une SEP alors qu'associé à HLA-DRB1\*15, le risque de SEP est doublé (Oksenberg, Baranzini et al. 2008, Ramagopalan and Ebers 2009).



**Figure 7. Genome-Wide Association Studies et Méta-analyses dans la SEP.**

SNP : Single Nucleotide Polymorphisms

- 1 : (IMSGC, Hafler et al. 2007)
- 2 : (Baranzini, Galwey et al. 2009)
- 3 : (De Jager, Jia et al. 2009)
- 4 : (ANZgene 2009)
- 5 : (Patsopoulos, Esposito et al. 2011)

Le mécanisme biologique du gène HLA-DRB1 dans le développement de la SEP est encore mal connue. Des études de structures protéiques de la molécule HLA semblent montrer que les résidus hautement polymorphiques de HLA-DRB1 agiraient sur la forme et la charge du site de fixation de l'antigène, et pourrait ainsi jouer sur l'efficacité de la présentation de ces antigènes aux cellules immunitaires (Oksenberg, Baranzini et al. 2008).

### 1.3.2. Les études de liaison

Les premières analyses génétiques menées dans la sclérose en plaques étaient des criblages anonymes du génome fondés sur des analyses de liaison de marqueurs microsatellites dans des familles multiplex.

Entre 1996 et 2005, 13 études d'analyses de liaison ont été publiées ainsi que 2 méta-analyses. Bien qu'elles aient permis d'identifier des régions d'intérêts en dehors du CMH, l'implication d'aucune d'entre elles n'a pu être confirmée dans les études de réplication (Sawcer 2008).

Les deux raisons majeures pouvant expliquer ces échecs sont le nombre restreint de marqueurs (microsatellites ou SNPs) utilisés dans ces études et le nombre restreint de familles disponibles pour les analyses (730 familles multiplex dans l'étude la plus complète (Sawcer, Ban et al. 2005)).

### 1.3.3. Les études d'association

En 2007, deux études gènes-candidats permettent pour la première fois depuis 1972 d'identifier un gène de prédisposition à la SEP indépendant du locus HLA : le gène IL7RA (Gregory, Schmidt et al. 2007, Lundmark, Duvefelt et al. 2007). Dans le même temps, un criblage anonyme du génome analysé par association et publié par l'IMSGC (International Multiple Sclerosis Genetics Consortium i.e. Consortium International sur la Génétique de la SEP, ne regroupant alors que des équipes américaines et anglaises) permet l'identification de deux gènes de prédisposition à la SEP : le gène IL7RA et le gène IL2RA (IMSGC, Hafler et al. 2007), qui codent respectivement pour une sous-unité du récepteur à l'Interleukine 7 et pour une sous-unité du récepteur à l'Interleukine 2.

Entre 2008 et 2010, 3 GWAS (ANZgene 2009, Baranzini, Wang et al. 2009, Sanna, Pitzalis et al. 2010) ainsi qu'une méta-analyse (appelée meta v1.0 (De Jager, Jia et al. 2009)) combinant plusieurs GWAS entre eux, sont publiés (**Figure 7**) et permettent d'identifier avec certitude 16 régions chromosomiques de prédisposition à la maladie, et 10 autres potentiellement associées à la SEP.





Cependant le plus important des GWAS publié n'incluait qu'au maximum 1600 patients et 3400 contrôles et la méta-analyse la plus importante en incluait respectivement 2600 et 7200 ce qui ne constituait qu'une faible puissance statistique.

C'est en 2011 que l'IMSGC mène le plus important GWAS dans la SEP. Le consortium regroupe aujourd'hui 23 équipes de recherche provenant de 15 pays (Allemagne, Angleterre, Australie, Belgique, Danemark, Espagne, Etats-Unis, Finlande, France, Irlande, Italie, Norvège, Nouvelle-Zélande, Pologne, Suède). Le rationnel de l'étude a été publié par Sawcer et al., en 2008 (Sawcer 2008). Selon ces estimations, la taille nécessaire de l'échantillon d'étude pour la mise en évidence d'un variant de prédisposition ayant une fréquence de 10% et un OR de 1,2, est d'environ 10.000 patients et 10.000 contrôles. Grâce à la mise en commun des ressources biologiques des 23 équipes, l'IMSGC a publié en Août 2011 des résultats portant sur l'analyse de 450 000 SNPs chez 9772 patients et 17376 contrôles (IMSGC, WTCCC2 et al. 2011).

Cette étude a permis d'augmenter à 52 le nombre de régions de prédisposition à la SEP connues hors HLA. Le projet IMSGC a conduit (a) à la confirmation de l'association de 23 des 26 régions chromosomiques suspectées, mises en évidence par les études publiées entre 2007 et 2010, (b) à l'identification de 29 nouvelles régions et (c) à la mise en évidence de 5 potentielles. Chacun des variants identifiés ne confère qu'un faible risque de développer la maladie de 1,1 à 1,3 fois supérieur à celui d'un individu non-porteur.

Il est intéressant de noter que 21 des 57 variants mis en évidence par ce criblage sont des facteurs de prédisposition communs à d'autres maladies dites auto-immunes comme le diabète de type I, la maladie coeliaque, la polyarthrite rhumatoïde, la rectocolite hémorragique, la maladie de Crohn ou encore le psoriasis. Cette constatation renforce l'hypothèse d'une architecture génétique commune à ces maladies complexes. Cependant même s'ils sont communs à plusieurs maladies dys-immunes, les directions des effets associés ne sont pas toujours les mêmes. Ainsi un allèle à risque pour une maladie peut être un allèle protecteur pour une autre maladie (Sirota, Schaub et al. 2009). Quelques études se sont focalisées sur le pourcentage de variants de risques partagés entre les différentes maladies dys-immunes mais elles ne prenaient pas en compte pas les effets associés à ces variants. Sirota et al. ont quant à eux intégrés dans leur étude les directions des effets des variants associés à ces maladies et ont émis l'hypothèse que certains loci prédisposeraient aux



maladies dys-immunes en général et que d'autres, par leur effet protecteur ou à risque pour une ou des maladies, détermineraient la prédisposition spécifique d'un individu à une maladie.

En 2011, une méta-analyse (meta v2.5) combinant tous les GWAS publiés jusqu'à présent, excepté le dernier GWAS de l'IMSGC, a permis d'identifier 3 nouveaux locis associés à la prédisposition à la SEP (Patsopoulos, Esposito et al. 2011) (**Figure 7**).

Le **Tableau** présenté pages 52-53 liste les SNPs mis en évidence comme étant associés à la maladie et connus en 2011.

Chacune de ces régions génomiques comportent de un à plusieurs gènes, exceptées deux d'entre-elles qui semblent ne posséder aucun gène. Au sein des régions où plusieurs gènes sont présents, les analyses d'identification du gène causatif sont en cours.

SNP	Chr	Gène Candidat IMSGC	Autres Gènes Candidats	Odds Ratio	Autre maladie autoimmune	Source
rs4648356	1	<i>MMEL1</i>		1,16	PR, MCo	1
rs118100217	1	<i>EVI5</i>		1,15		1
rs11581062	1	<i>VCAMI</i>	<i>SLC30A7; DPH5</i>	1,12		1
rs1335532	1	<i>CD58</i>		1,22		1
rs1323292	1	<i>RGS1</i>		1,12	MC	1
rs12122721	1	<i>KIF21B</i>	<i>CACNAIS; TMEM9</i>	1,22	CH, MCo, MCr	1
rs6718520	2	<i>THADA</i>		1,17		2
rs12466022	2	<i>No gene</i>		1,11		1
rs7595037	2	<i>PLEK</i>		1,11	MCo	1
rs17174870	2	<i>MERTK</i>	<i>TMEM87B</i>	1,11		1
rs10201872	2	<i>SP140</i>	<i>SP100</i>	1,14		1
rs11129295	3	<i>EOMES</i>		1,11		1
rs669607	3	<i>No gene</i>		1,13		1
rs170934	3	<i>No gene</i>		1,17		2
rs2028597	3	<i>CBLB</i>	<i>ALCAM</i>	1,13		1
rs2293370	3	<i>TMEM39A</i>	<i>CD8; CDGAP</i>	1,16		1
rs9282641	3	<i>CD86</i>	<i>ILDR1</i>	1,21		1
rs2243123	3	<i>IL12A</i>		1,08	MCo	1
rs228614*	4	<i>NFKB1</i>	<i>MANBA; UBE2D3</i>	1,09		1
rs6897932	5	<i>IL7R</i>		1,11	DT1	1
rs4613763	5	<i>PTGER4</i>	<i>TTC33</i>	1,2	MCr	1
rs2546890	5	<i>IL12B</i>		1,11	PS, MCr	1
rs12212193	6	<i>BACH2</i>		1,09	MCo, DT1	1
rs802734	6	<i>THEMIS</i>	<i>PTPRK</i>	1,1	MCo	1
rs11154801	6	<i>MYB</i>		1,13		1
rs17066096	6	<i>IL22RA2</i>		1,14		1
rs13192841	6	<i>OLIG3</i>		1,1	PR	1
rs1738074	6	<i>TAGAP</i>		1,13	MCo	1
rs354033	7	<i>ZNF767</i>		1,11		1
rs1520333	8	<i>IL7</i>		1,1		1
rs4410871	8	<i>MYC</i>		1,11		1
rs2019960	8	<i>PVT1</i>		1,12		1
rs2150702	9	<i>MLANA</i>		1,16		2
rs7090512	10	<i>IL2RA</i>		1,19	PR	1
rs1250550	10	<i>ZMIZ1</i>		1,1	MCo	1
rs7923837	10	<i>HHEX</i>	<i>IDE; KIF11</i>	1,1		1
rs650258	11	<i>CD6</i>		1,12		1
rs630923*	11	<i>CXCR5</i>		1,12		1

SNP	Chr	Gène Candidat IMSGC	Autres Gènes Candidats	Odds Ratio	Autre maladie autoimmune	Source
rs1800693	12	<i>TNFRSF1A</i>		1,12		1
rs10466829	12	<i>CLECL1</i>	CLEC2D; CD69	1,09	DT1	1
rs12368653	12	<i>CYP27B1</i>	AGAP2; CDK4; TSPAN31; MARCH9; METTL1	1,11	PR	1
rs949143	12	<i>MPHOSPH9</i>		1,08		1
rs4902647	14	<i>ZFP36L1</i>		1,11	MCo, DT1	1
rs2300603	14	<i>BATF</i>		1,11		1
rs2119704	14	<i>GALC</i>	GPR65	1,22		1
rs2744148*	16	<i>SOX8</i>	LMF1; SSTR5	1,12		1
rs7200786	16	<i>CLECI6A</i>	DEXI; SOCS1	1,15	DT1	1
rs13333054	16	<i>IRF8</i>		1,12		1
rs9891119	17	<i>STAT3</i>		1,11	MCr	1
rs180515*	17	<i>RPS6KB1</i>		1,09		1
rs7238078	18	<i>MALT1</i>		1,12		1
rs763361	18	<i>CD226</i>		1,03		1
rs1077667	19	<i>TNFSF14</i>		1,16		1
rs8112449	19	<i>TYK2</i>		1,1	DT1	1
rs874628	19	<i>MPV17L2</i>		1,11		1
rs2303759	19	<i>DKK1</i>		1,11		1
rs2425752	20	<i>CD40</i>		1,11	PR	1
rs2248359	20	<i>CYP24A1</i>		1,12		1
rs6062314*	20	<i>TNFRSF6B</i>		1,16		1
rs2283792	22	<i>MAPK1</i>		1,1		1

**Tableau. Liste des SNPs associés à la SEP, connus en 2011.**

\*: SNPs potentiellement associés à la prédisposition à la SEP mais n'ayant pas passé le seuil significatif de  $5 \times 10^{-8}$

Gène Candidats IMSGC : Gènes reportés comme candidats dans la publication de l'IMSGC de 2011 (IMSGC, WTCCC2 et al. 2011).

Autres Gènes Candidats : Autres gènes localisés dans la région du SNP correspondant.

Source : (1) correspond aux SNPs mis en évidence par le GWAS de l'IMSGC de 2011 (IMSGC, WTCCC2 et al. 2011) et (2) correspond aux SNPs mis en évidence par la méta-analyse meta v2.5 (Patsopoulos, Esposito et al. 2011)

SNP : Single Nucleotide Polymorphism

PR : Polyarthrite Rhumatoïde / MCo : Maladie Coeliaque / MCr : Maladie de Crohn / CH : Colique Hémorragique / DT1 : Diabète de Type 1 / PS : Psoriasis

**Figure 8. Représentation graphique du pathway de différenciation des lymphocytes T.**

Les gènes en couleur sont ceux semblant être impliqués du fait de leur proximité avec le SNP montrant une association à la SEP. En rouge, ceux étant confirmés comme étant associés. En orange, ceux n'étant pas confirmés comme associés mais ayant une p-value de découverte  $< 1 \times 10^{-4.5}$ . En jaune, ceux ayant une p-value de découverte  $< 1 \times 10^{-3}$ .

De (IMSGC, WTCCC2 et al. 2011)

## 1.4. Corrélation génétique - mécanismes biologiques

Il est important de noter que près de la moitié des régions de prédisposition identifiées à ce jour contiennent des gènes codant pour des protéines impliquées dans des processus immunitaires incluant la voie biologique des cytokines, des molécules de co-stimulation et des molécules de transduction du signal qui semblent donc être d'excellents candidats au vue de la physiopathologie de la SEP.

Tous ces résultats confortent aujourd'hui l'hypothèse d'une prédisposition immunitaire à l'origine de la sclérose en plaques. L'attaque de la gaine de myéline serait due à l'accumulation de facteurs de risque dans des gènes immunitaires, après exposition à un ou des facteurs environnementaux.

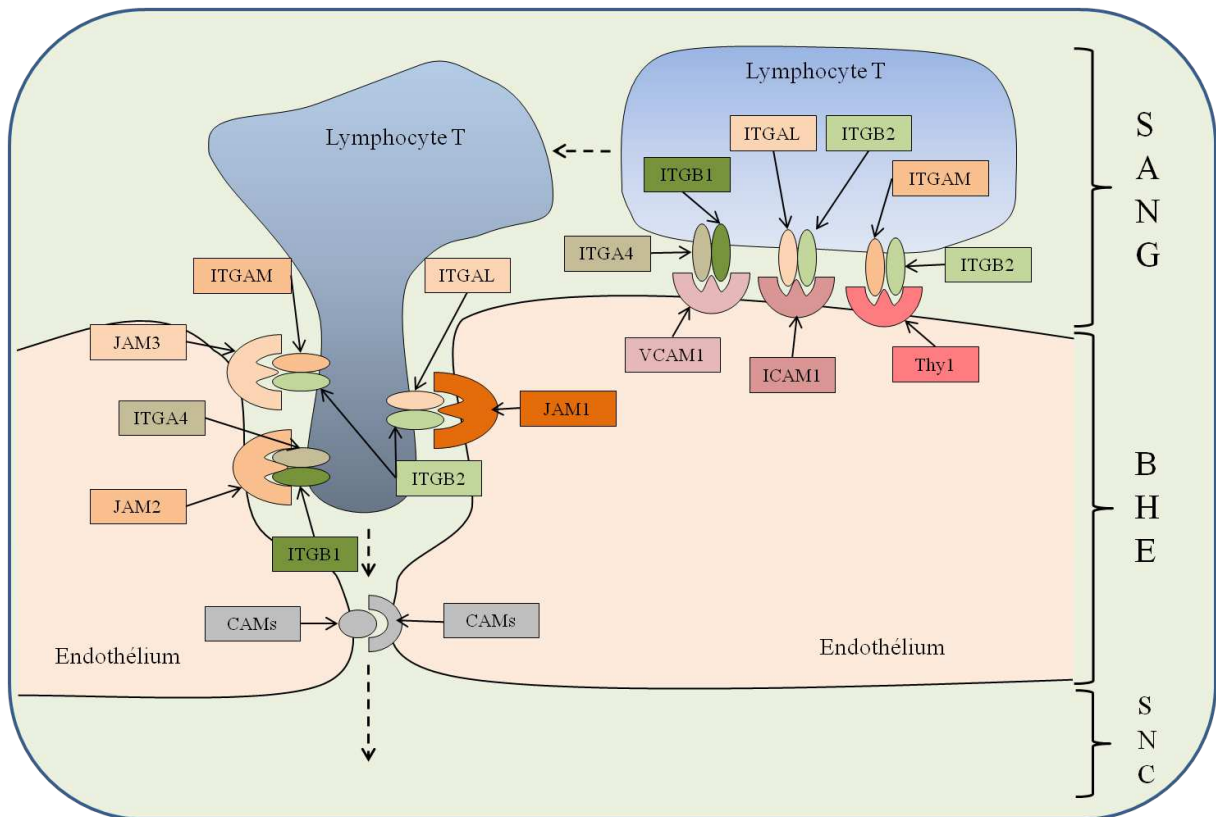
### 1.4.1. Activation immunitaire

Plus de la moitié des gènes identifiés par les études génétiques antérieures (le GWAS de l'IMSGC de 2011 et la meta v2.5) sont dits immunitaires et sont pour la plupart d'entre eux impliqués dans la différenciation des LT. Lors d'une réaction immunitaire, un LT naïf est activé par la reconnaissance via son récepteur antigénique de surface (ou TCR, *T Cell Receptor*) d'un antigène lié à des molécules du CMH de classe I (s'il s'agit d'un lymphocyte de type CD8+) ou de classe II (pour un lymphocyte CD4+) qui sont exprimées à la surface des Cellules Présentatrices d'Antigènes (CPA). Cette activation va être suivie par une phase de prolifération et de différenciation. Dans le cas des LT CD4+ et selon les cytokines sécrétées par la CPA et le contexte inflammatoire, cette différenciation permet aux lymphocytes de passer d'un état naïf mature (ou  $T_{H0}$ ) vers différents stades de cellules T auxiliaires (ou *T helper* ;  $T_H$ ).

Les LT CD4<sup>+</sup> naïfs peuvent ainsi se différencier en 5 sous-types cellulaires (**Figure 8**) :

- les **lymphocytes  $T_{H1}$  et  $T_{H17}$** , favorisant l'inflammation par la production de cytokines dites pro-inflammatoires
- les **lymphocytes  $T_{H2}$  et Treg**, inhibant l'inflammation par la production de cytokines dites anti-inflammatoires
- les **lymphocytes  $T_{FH}$**  (T Follicular Helper) qui sont nécessaires à la formation et la maintenance des centres germinatifs et à la génération de la plupart des **lymphocytes B** mémoires.





**Figure 9. Molécules impliquées dans le franchissement de la Barrière Hémato-Encéphalique (BHE)**

SNC : Système Nerveux Central / CAMs : Cell Adhesion Molecules (exemple : CLDN3, CLDN5, OCLN)

Dans le cas de la SEP, il existe conjointement une différenciation préférentielle des lymphocytes T CD4<sup>+</sup> vers les voies pro-inflammatoires T<sub>H</sub>1 et T<sub>H</sub>17 et un dysfonctionnement des mécanismes régulateurs via les Treg. Les interactions entre lymphocytes T et B aboutissent à l'activation des cellules B, qui différenciées en plasmocytes sécrètent les gammaglobulines intra-thécales observées dans le Liquide Céphalo-Rachidien (LCR) des patients.

#### 1.4.2. Passage de la Barrière Hémato-Encéphalique

Après activation, les lymphocytes T acquièrent la capacité de traverser la Barrière Hémato-Encéphalique (BHE), barrière de cellules endothéliales séparant la circulation sanguine du SNC (Steinman 2001, Fox 2004, Gold and Wolinsky 2011). Ce processus constitue une étape cruciale dans le développement de la SEP. Il en résulte une infiltration massive de LT autoréactifs dans le SNC.

Ces LT autoréactifs sont préactivés en dehors du SNC par des antigènes exogènes, des superantigènes microbiens ou viraux ou des protéines pathogéniques ayant une homologie de structure avec des protéines du soi (on parle alors de mimétisme moléculaire) (Oldstone 1998), dont probablement celles constituant la myéline. Une autre hypothèse est que ces LT pourraient être préactivés directement par des antigènes de la myéline dégradée.

Certains des gènes de prédisposition identifiés, comme le gène *VCAMI*, jouent un rôle prépondérant dans la transmigration des LT dans le SNC. Le franchissement de la BHE par les cellules T se fait par des processus successifs, impliquant l'interaction entre des molécules d'adhésion se trouvant à la surface des LT et d'autres se trouvant à la surface des cellules endothéliales (Engelhardt and Ransohoff 2005, Man, Ubogu et al. 2007, Engelhardt 2008, Weiss, Miller et al. 2009) (**Figure 9**):

- L'étape de roulement à la surface de l'endothélium met en jeu l'interaction de diverses molécules, telles les sélectines (SELL, SELP et SELE), les intégrines VLA-4, MAC-1, ces dernières interagissant respectivement avec les ligands des sélectines (PSGL-1 notamment), VCAM-1 et Thy-1 ;

- L'étape d'adhérence, d'abord faible avec une nouvelle fois la mise en jeu des sélectines, puis ferme avec la mise en jeu des intégrines leucocytaires VLA-4 et LFA-1 qui interagissent avec leur récepteurs VCAM-1 et ICAM-1 et ICAM-2 à la surface des cellules endothéliales de la BHE ;

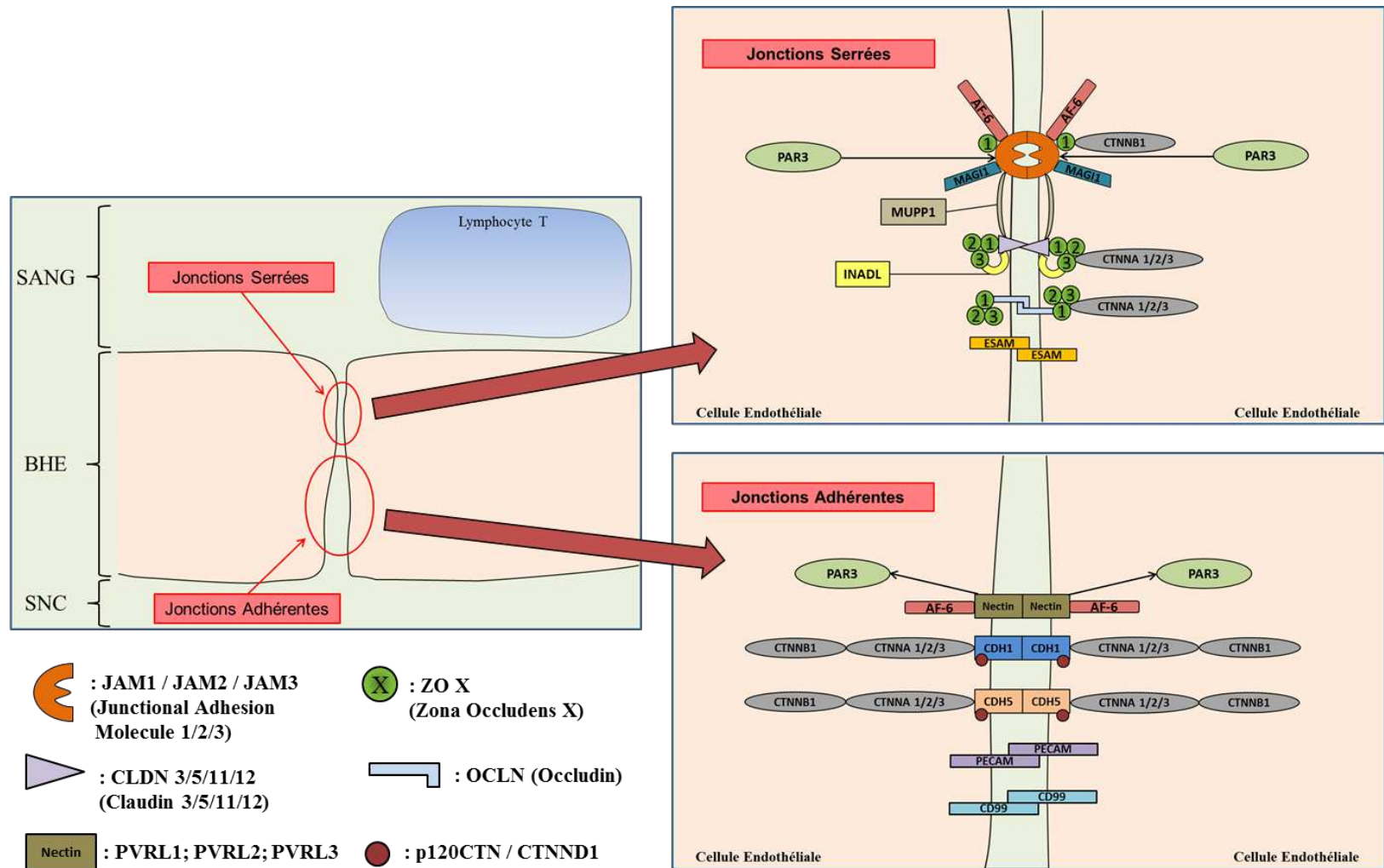


Figure 10. La Barrière Hémato-Encéphalique (BHE) et les molécules la constituant.

- L'étape de transmigration, qui nécessite l'assouplissement des jonctions entre les cellules composant la BHE. Les cellules endothéliales de la BHE sont connectées entre elles par des jonctions adhérentes et des jonctions serrées mettant en jeu diverses molécules d'adhésion, telles les Claudines, l'Occludine, les JAMs (Junctional Adhesion Molecules) (Wolburg and Lippoldt 2002, Ballabh, Braun et al. 2004, Hawkins and Davis 2005, Hartsock and Nelson 2008, Abbott, Patabendige et al. 2010) (**Figure 10**).

L'importance du mécanisme de transmigration des LT au sein du SNC est soulignée par l'existence d'un traitement, le Natalizumab. Il s'agit d'un anticorps monoclonal dirigé contre la sous-unité ITGA4 du ligand VLA-4. Cette thérapie, inhibant l'infiltration des cellules immunitaires dans le SNC via la BHE, constitue aujourd'hui l'un des traitements les plus efficaces dans la SEP.

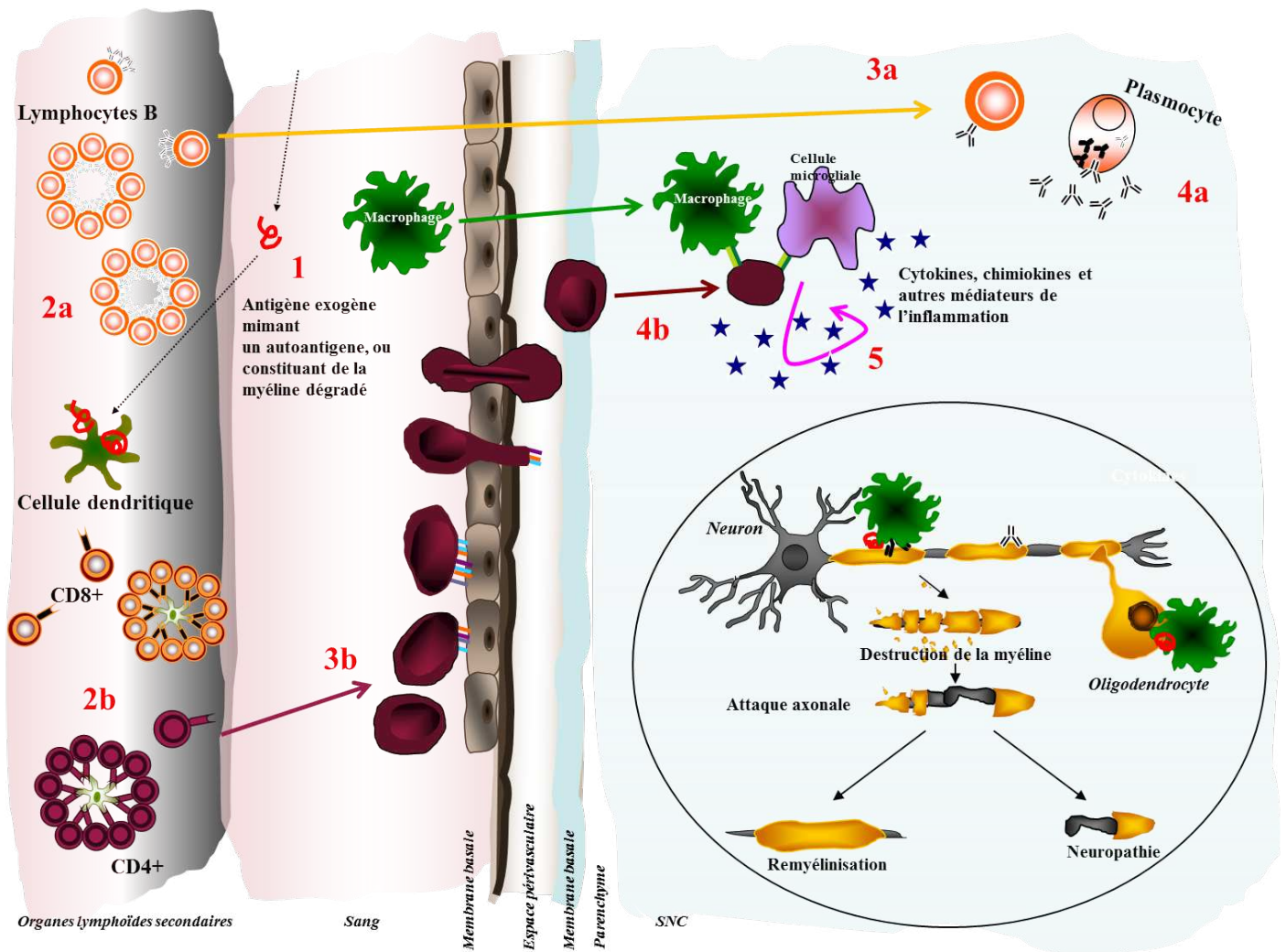
#### 1.4.3. La chronicité

La chronicité de la SEP provient d'un emballement de la réponse immunitaire et d'une inflammation qui intervient de manière répétée et non contrôlée, correspondant aux phases dites de « poussées ».

Les lymphocytes CD4+ infiltrés dans le SNC sont réactivés par les CPA, ces dernières présentant les auto-antigènes, par l'intermédiaire de leur CMH de classe II. Cette réactivation déclenche la sécrétion de cytokines pro-inflammatoires comme le TNF- $\alpha$ , l'IFN- $\gamma$  ou la lymphotoxine- $\alpha$ . Ces molécules provoquent l'activation de la microglie et des macrophages résidents. Ces derniers vont s'attaquer à la myéline, provoquant ainsi sa destruction. De plus, les lymphocytes T activés par les CPA permettent un recrutement accru de lymphocytes B, provenant de la périphérie. Ceux-ci produisent des anticorps dirigés contre la myéline, contribuant ainsi à sa destruction (Sospedra and Martin 2005, Gold and Wolinsky 2011, Rodgers and Miller 2012) (**Figure 11**, présentée page suivante)

Bien que les anomalies immunitaires observées dans la SEP soient décrites de plus en plus précisément, le caractère chronique et cyclique de l'inflammation caractéristique de cette maladie reste mal compris. Les gènes immunitaires de prédisposition identifiés codent pour la plupart pour des cytokines, des récepteurs aux cytokines et des facteurs de transcription. On peut imaginer d'une part que tous ces acteurs induisent une sur-activation des LT, permettant à ces derniers de se différencier et de produire toujours plus de cytokines pro-inflammatoires par l'intermédiaire de boucles de rétrocontrôles positives. D'autre part, il est aussi possible





**Figure 11. Mécanismes immunitaires mis en jeu dans la Sclérose En Plaques.**

Un antigène exogène mimant un autoantigène myélinique ou un constituant de la myéline dégradé est capté et présenté par la cellule dendritique dans les organes lymphoïdes secondaires (1). Des LB (2a) ou des LT (2b) reconnaissant cet antigène vont être alors activés et stimulés par la cellule dendritique. Les LB (3a) et les LT activés (3b) vont traverser la BHE et envahir le SNC. Les cellules B se différencient en plasmocytes qui vont produire des anticorps dirigés contre la myéline (4a). Au sein du SNC, les LT sont réactivés par les cellules microgliales résidentes ou les macrophages infiltrants (4b), ce qui provoque la sécrétion de cytokines pro-inflammatoires, de chimiokines et d'autres médiateurs cytotoxiques tels que des dérivés réactifs de l'oxygène. Il en résulte une inflammation accrue par l'intermédiaire des chimiokines qui vont attirer d'autres cellules immunitaires au sein du SNC ainsi qu'une destruction de la gaine de myéline par les lymphocytes et les macrophages.

**Figure 12. Lésions de démyélinisation caractérisées par IRM (Imagerie par Résonance Magnétique).**

Les lésions sont désignées par les flèches rouges.

- A. Lésion dans le cerveau.
- B. Lésion dans le nerf optique.
- C. Lésion dans la moelle épinière.

D'après (Compston and Coles 2002)

que ces gènes immunitaires influent sur les mécanismes de contrôles de la réponse immunitaires les rendant défailants.

## 1.5. Manifestations cliniques des mécanismes biologiques

### *1.5.1. Démyélinisation*

Les régions de démyélinisation dans le cerveau, la moelle épinière ou les nerfs optiques, encore appelée lésions ou « plaques » (**Figure 12**) entraînent différents symptômes qui sont à la fois épisodiques et très variés. Ceux-ci sont associés à des troubles affectant les systèmes moteurs, sensoriels et visuels. Les plus courants sont des engourdissements, des faiblesses musculaires, des troubles oculaires, des problèmes de coordination et des vertiges. Peuvent s'ajouter à cela des troubles divers comme la fatigue, des troubles cognitifs et des troubles sphinctériens (Compston and Coles 2008).

La symptomatologie est directement liée à la topographie des lésions, et témoigne de la dissémination spatiale et temporelle de celles-ci.

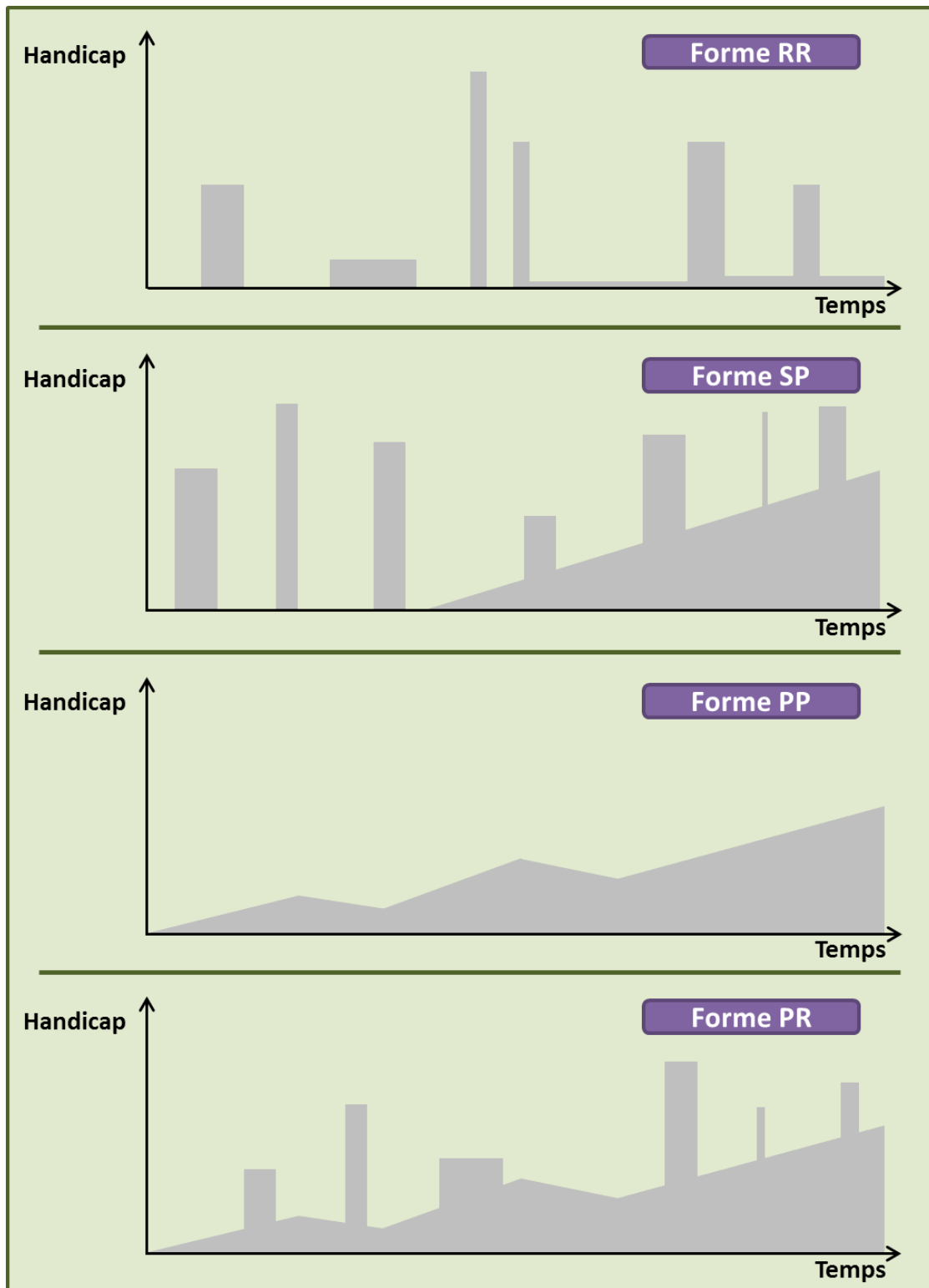
La visualisation par IRM (Imagerie par Résonance Magnétique) de la dynamique de ces lésions cérébrales et médullaires constitue aujourd'hui l'outil diagnostique essentiel de la SEP, permettant d'une part, la mise en évidence des plaques actives, témoins de l'inflammation et de la démyélinisation et d'autre part, de trous noirs, témoins de la destruction matricielle et de la perte axonale (Tourdias, Brochet et al. 2009).

En effet, alors qu'en 1965 les critères diagnostiques de Schumacker reposaient essentiellement sur des critères cliniques (Schumacker, Beebe et al. 1965), Poser introduit en 1983 de nouveaux critères, alliant données cliniques et paracliniques (Poser, Paty et al. 1983). En 2001, McDonald introduit les données IRM dans le diagnostic de la SEP (McDonald, Compston et al. 2001).

Aujourd'hui, le diagnostic d'une SEP repose sur 3 critères, appelés « critères de McDonald » (McDonald, Compston et al. 2001, Polman, Reingold et al. 2011):

- une multiplicité des lésions dans le système nerveux central (on parle alors de « **dissémination dans l'espace** »)





**Figure 13. Les formes d'évolution de la Sclérose en Plaques.**

Forme RR : forme rémittente

Forme SP : forme secondairement progressive SP

Forme PP : forme progressive primaire ou progressive d'emblée, sans poussées

Forme PR : forme progressive primaire avec poussées

- une évolution des lésions dans le temps (on parle alors de « **dissémination dans le temps** ») évaluée sur l'histoire clinique ou l'IRM cérébrale qui permet de visualiser les plaques

- l'élimination d'autres causes potentielles.

Du fait de l'absence de biomarqueur spécifique, le diagnostic de sclérose en plaques est donc porté sur un faisceau d'arguments cliniques, paracliniques mais aussi évolutifs. L'ensemble des arguments cumulés doivent conforter l'élimination du diagnostic différentiel d'une autre pathologie aux symptômes similaires.

Un patient ne présentant qu'un épisode lésionnel sera classé dans la catégorie des syndromes cliniquement isolés (CIS, Clinically Isolated Syndrome). Seule une partie des patients (environ 60-65%, 2 ans après le premier épisode lésionnel) développeront une SEP cliniquement définie après un CIS, dont le diagnostic ne sera porté qu'à la survenue d'un deuxième épisode symptomatique.

#### 1.5.2. L'inflammation chronique

La sclérose en plaques est une maladie chronique et évolutive, l'évolution clinique résultant de la combinaison de deux événements de base (Confavreux and Vukusic 2006):

- **la poussée**, source de handicap transitoire, qui est l'expression du processus inflammatoire, multifocal, aléatoire et récurrent de la maladie. Elle se définit comme l'apparition de nouveaux symptômes, la réapparition d'anciens symptômes ou l'aggravation de symptômes préexistants s'installant de manière rapide, en quelques heures à quelques jours, et récupérant de manière plus ou moins complète.

- **l'accumulation progressive du handicap** (ou progression) est quant à elle le reflet du processus chronique de perte axonale. Une fois débuté, ce processus se poursuit au cours de la maladie et constitue la source majeure de handicap chez les patients.

La combinaison de ces deux événements de base a amené les neurologues à distinguer 4 formes majeures d'évolution clinique de la SEP: les formes rémittentes RR, les formes secondairement progressives SP, les formes progressives primaires PP sans poussées et les formes progressives primaires PR avec poussées (**Figure 13**) (Lublin and Reingold 1996):

**Figure 14. Échelle EDSS (Extended Disability Status Scale) simplifiée.**  
Cette échelle simplifiée ne représente pas les valeurs de ½ point.

- la **forme rémittente RR**, la plus courante (85% des SEP), se caractérise par une succession de poussées-rémissions. En début de maladie, le patient ne présente pas de séquelles après les poussées ce qui témoigne d'une réparation spontanée des gaines de myéline (appelée remyélinisation).

- la **forme secondairement progressive SP** survient en moyenne 10 ans après le début des SEP RR, chez 65% des patients environ. Elle correspond à l'installation et à la progression d'un handicap persistant entre les poussées. Elle reflète la diminution voire l'absence du processus de remyélinisation et l'atteinte axonale qui en résulte.

- la **forme progressive d'emblée, ou progressive primaire PP, sans poussées**, s'observe chez 10% des patients. Elle se caractérise par une aggravation progressive des symptômes et du handicap en résultant dès le début de la maladie.

- la **forme progressive rémittente PR** (5% des cas), se caractérise par une évolution semblable à la forme PP mais **avec des poussées**.

Les formes PP et PR constituent les formes les plus graves de SEP.

### 1.5.3. Installation du handicap

Le caractère chronique de la maladie conduit à l'installation d'un handicap qui doit être mesuré pour estimer l'évolution et la sévérité de la maladie.

#### *1.5.3.1. Score de handicap*

La mesure du handicap la plus utilisée a été introduite par JF Kurtzke en 1983 (Kurtzke 1983). Il s'agit de l'échelle EDSS (Extended Disability Status Scale) qui permet aux cliniciens d'attribuer un score de handicap aux patients, essentiellement fondé sur les atteintes motrices. Cette échelle se répartit sur 20 niveaux, allant de 0 à 10, par paliers de ½ points (**Figure 14**). Elle combine des critères d'examen clinique, d'incapacité de certaines fonctions neurologiques (la marche) et de handicap et n'est par conséquent pas linéaire.

Bien que ce score soit largement utilisé par tous les neurologues pour évaluer l'évolution du handicap, il ne permet néanmoins pas d'évaluer l'ensemble des déficits et ne reflète donc pas la totalité des lésions du SNC. D'autre part, certains scores sont définis par un périmètre de marche qu'il est difficile d'estimer lors d'une consultation, comme par exemple l'EDSS 4 qui correspond à un périmètre de marche limité mais restant supérieur à 500m.

**Figure 15. Multiple Sclerosis Severity Score (MSSS).**

Le MSSS pour un individu est défini en se rapportant à la colonne de sa valeur EDSS et à la ligne correspondant à la durée écoulée depuis le début de la maladie.

De (Roxburgh, Seaman et al. 2005)

### *1.5.3.2. Score de sévérité*

Jusqu'en 2005, la sévérité de la sclérose en plaques était estimée par le calcul d'un ratio, appelé Indice de Progression IP (ou PI pour Progression Index) qui consiste à rapporter le score EDSS (cf partie 1.5.3.1 Score de handicap) à la durée d'évolution de la maladie, c'est à dire le temps qui s'est écoulé depuis son déclenchement. Cependant, ce ratio est sujet à fluctuation, l'évolution du score EDSS n'étant pas linéaire. En 2005, Roxburgh et al. ont développé un nouveau score, appelé Score MSS ou MSSS (Multiple Sclerosis Severity Score) (Roxburgh, Seaman et al. 2005). Comme l'IP, ce score est basé sur l'EDSS et la durée d'évolution de la maladie au moment de la mesure du handicap. A la différence de l'IP, le calcul du MSSS est corrigé par un algorithme qui compare le handicap d'un individu avec les scores EDSS des patients ayant la même durée d'évolution.

Ceci a permis de créer une table de référence (**Figure 15**) basée sur le calcul du MSSS d'une cohorte de 9892 patients qui permet d'attribuer un score MSSS à chaque patient qui présente une durée d'évolution comprise entre 1 et 30 ans. Ce score est compris entre 0 et 10 : Plus le score est fort, plus le patient est atteint d'une forme sévère.

Le suivi longitudinal d'une cohorte de 2429 patients a montré que le MSSS est globalement stable au cours du temps au sein d'une cohorte, mais ne peut pas être utilisé comme outil prédictif de la progression du handicap à l'échelle individuelle (Roxburgh, Seaman et al. 2005).

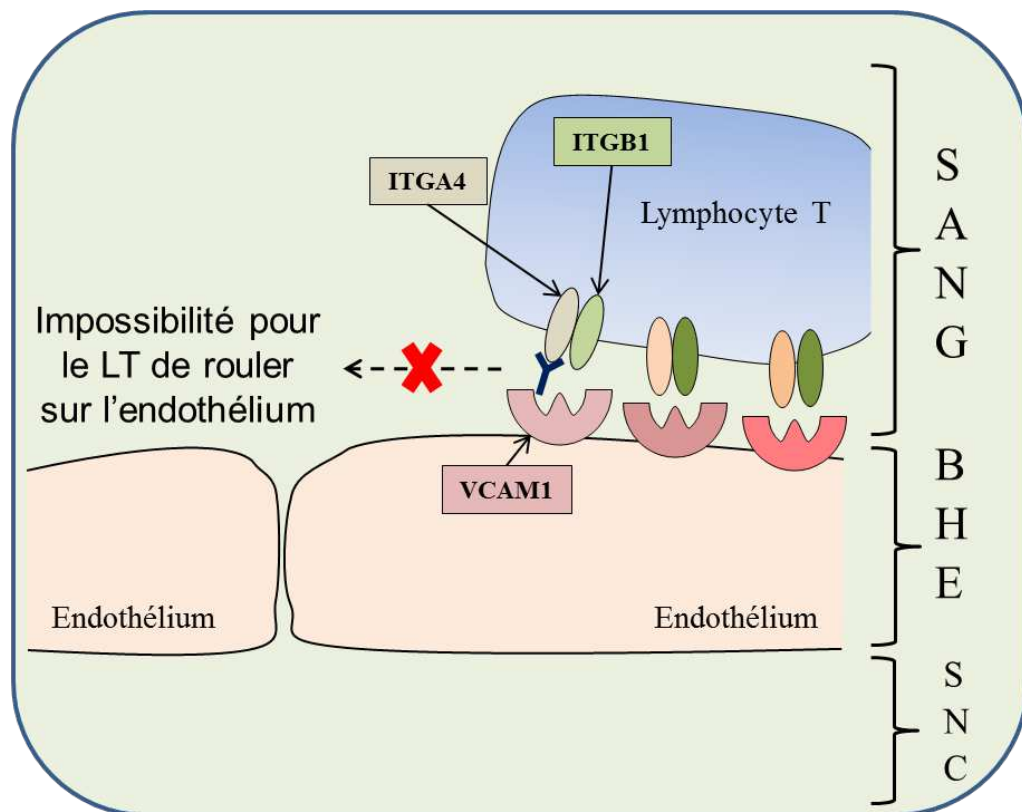
Le MSSS est aujourd'hui largement utilisé dans les études génétiques sur la sévérité de la maladie.

### *1.5.4. Les traitements de la Sclérose En Plaques*

Les différents traitements utilisés dans la sclérose en plaques ont pour objectifs de limiter la fréquence et l'intensité des poussées et d'améliorer le handicap permanent lié aux poussées. A ce jour, il n'existe aucun traitement curatif. La poussée, manifestation clinique d'une inflammation aiguë dans le SNC, est généralement traitée par corticoïdes. Les corticoïdes permettent de réduire la durée de la poussée mais n'ont pas d'effet ni sur les récurrences ni sur le handicap résiduel. Des traitements de fond sont donc proposés aux patients.

#### *1.5.4.1. Traitements de fond de première intention*

Pour les formes rémittentes classiques, les traitements de première intention sont dans la majorité des cas des ImmunoModulateurs (IM), à savoir l'IFN $\beta$ -1a sous-cutané (sc),



**Figure 16. Mode d'action du Natalizumab dans la Sclérose En Plaques.**

La fixation du Natalizumab sur la sous-unité ITGA4 de la molécule VLA-4 empêche l'interaction de cette dernière avec son récepteur VCAM1. Les lymphocytes T (LT) ne peuvent plus interagir avec les cellules endothéliales de la Barrière Hémato-Encéphalique (BHE), l'étape d'adhésion ne peut donc plus se faire. S'ensuit une impossibilité pour le LT de rouler sur l'endothélium et donc une impossibilité d'infiltrer le Système Nerveux Central (SNC).

l'IFN $\beta$ -1b sc, l'IFN $\beta$ -1a intra-musculaire ou l'acétate de glatiramère (encore appelé copolymère) (Vermersch 2009).

L'efficacité de l'interféron  $\beta$  dans la SEP est due à son effet immunomodulateur cytokinique. Il inhibe l'action de l'IFN- $\gamma$ , ce dernier semblant contribuer au déclenchement des poussées et à la progression de la maladie.

Le mode d'action de l'acétate de glutarimère (ou copolymère) n'est quant-à-lui pas connu avec précision. Il s'agit d'un polymère de 4 acides aminés (L-alanine, L-acide glutamique, L-lysine, L-tyrosine) qui a une action immuno-modulatrice sur les différentes étapes physiopathologiques lésionnelles de la SEP et pourrait aussi agir comme un « leurre antigénique », » du fait de son homologie de structure avec une des protéines de la myéline (Johnson, Brooks et al. 1995).

Les formes secondairement progressives sont elles traitées par IFN $\beta$ -1b sc ou l'IFN $\beta$ -1a sc, traitement probablement actif chez les patients ayant des poussées résiduelles. En revanche, il n'existe pour le moment aucun traitement pour les patients présentant une forme progressive primaire (Vermersch 2009).

#### *1.5.4.2. Traitements de fond de seconde intention*

Aujourd'hui, trois traitements de seconde intention existent : la Mitoxantrone, le Natalizumab et le Fingolimod, qui appartiennent à la classe des Immunosuppresseurs (IS) (Vermersch 2009).

La Mitoxantrone possède un effet cytotoxique. Elle réduit de manière très importante le taux de poussées et l'évolution des lésions IRM. Cependant, du fait du risque cardiotoxique et de leucémie qu'elle présente, elle est réservée aux formes très actives de SEP.

Le **Natalizumab** est un anticorps monoclonal humanisé capable de se fixer sur la sous-unité ITGA4 de l'intégrine VLA-4 (composée de la sous-unité ITGA4 et d'une sous-unité ITGB1) (Yednock, Cannon et al. 1992). De ce fait, il bloque la liaison du ligand VLA-4 se trouvant à la surface des lymphocytes à son récepteur VCAM-1 localisé à la surface des cellules endothéliales de la BHE. Les lymphocytes ne peuvent donc plus adhérer aux cellules endothéliales, étape nécessaire à leur migration à travers la BHE du SNC (**Figure 16**). Il en résulte une diminution de l'activité inflammatoire dans le SNC et donc une diminution des lésions SEP. En effet, une étude réalisée chez des patients atteints de SEP rémittente a montré, à l'issue des 2 ans d'étude, une diminution de 42% du risque de progression du handicap (EDSS), une réduction de 68% du taux annuel de poussées et une réduction de 83% du





nombre de nouvelles lésions (Polman, O'Connor et al. 2006). A ce jour, le Natalizumab constitue le traitement le plus spectaculairement efficace dans la SEP.

Le Fingolimod a un effet antagoniste sur les récepteurs sphingosine-1-phosphate (S1P) 1, 3 et 4 présents sur les lymphocytes. Sa fixation aboutit à une séquestration des lymphocytes dans les ganglions lymphatiques, induisant ainsi une lymphopénie. Il a aussi la faculté de traverser la BHE et de se fixer sur les récepteurs S1P présents sur les neurones et les astrocytes. En modulant les récepteurs S1P, il atténuerait la neurogliose. Cependant, du fait de la lymphopénie, les patients traités par Fingolimod sont immunodéprimés.

Une étude réalisée chez des patients atteints de SEP rémittente a permis de prouver l'efficacité de ce traitement en montrant, à l'issue des 2 ans d'étude, une diminution de 54 à 60% du taux annuel de poussées et une réduction du risque de progression de 30 à 32% (Kappos, Radue et al. 2010).

Il est à noter que la Mitoxantrone et le Natalizumab peuvent être prescrits en première intention dans les cas de SEP rémittentes très actives (Vermersch 2009).



## 2. Sclérose En Plaques et recherche génétique actuelle

Le GWAS publié par l'IMSGC en Août 2011 a constitué une réelle avancée dans la compréhension de la prédisposition génétique à la sclérose en plaques. Aujourd'hui 52 marqueurs génétiques de prédisposition à la maladie sont connus et unanimement reconnus, ainsi que 5 potentiels, hors HLA (**Tableau** pages 52-53). Cependant, les résultats obtenus n'expliquent que 17% de la part génétique, bien que le projet ait été conçu pour en identifier 80%. Ceci n'est pas exclusif à la SEP. En effet, depuis de nombreuses années, des études GWAS ont été entreprises dans différentes maladies et traits phénotypiques et bien qu'elles aient toutes été conçues pour identifier une grande partie de la part génétique des maladies, les résultats obtenus ne permettent d'en expliquer que de 1,5% à 50%. Ainsi les 32 loci identifiés dans la maladie de Crohn et les 18 loci identifiés dans le Diabète de Type 2 expliquent respectivement 20% et 6% de la part génétique de ces maladies (Manolio, Collins et al. 2009).

### ➤ Calcul de l'héritabilité :

Le calcul de l'héritabilité d'un risque ou d'une prédisposition est estimé à partir des corrélations génétiques entre germains.

Les 61 variants identifiés (57 non-HLA et 4 HLA) par le GWAS de l'IMSGC confèrent aux apparentés au 1<sup>er</sup> degré d'un patient un risque relatif de développer la maladie de 1,58. Dans une étude de la population suédoise (Hemminki, Li et al. 2009), le risque global pour les germains est de 6,3. Les 61 variants expliquent donc environ 25% ( $1,58/6,3 \times 100$ ) du risque relatif d'un apparenté de développer une SEP.

Le calcul d'héritabilité est fondé sur le modèle de seuil de prédisposition selon lequel la variance totale sur l'échelle de risque est de 1 dans la population. Sous la loi normale, en prenant en compte la prévalence de la maladie dans la population générale ( $K=0,001$ ), ce modèle identifie les patients comme les individus ayant une prédisposition plus grande que  $T=3,09$ . Considérant un risque relatif pour les germains de 1,58 conférés par les 61 variants et sous l'hypothèse d'un effet nul de l'environnement, l'héritabilité  $h$  (calculée selon la formule 3 p.166 de l'article de Reich et al. (Reich, James et al. 1972)), est de 38%.

Chaque variant identifié explique un risque relatif individuel ( $RR_{ind}$ ) de développer la maladie calculé en fonction de son OR ( $OR$ ), de la fréquence de l'allèle à risque ( $f$ ) et de l'héritabilité ( $h$ ) calculée à partir de la corrélation entre germains soit 38% (IMSGC, WTCCC2 et al. 2011).

$$RR_{ind} = 0.5 \times f(1-f) \times OR^2 / h^2$$



Le risque relatif global expliqué par l'ensemble des variants identifiés est alors la somme des risques relatifs individuels.

A partir de ces calculs, on estime que les 4 variants HLA identifiés par le GWAS de l'IMSGC expliquent 10.5% de la part génétique de la maladie et que les 48 variants non-HLA en expliquent 6.5%. Les résultats du GWAS expliquent donc 17% de la composante génétique de la sclérose en plaques.

C'est dans ce contexte qu'est apparu la notion d'« **héritabilité manquante** » ou « **missing heritability** » (Maher 2008). Cette héritabilité manquante correspond ainsi à la part génétique qu'il reste à identifier, les variants de risques mis en évidence jusqu'à présent et considérer individuellement ne permettant pas d'expliquer toute l'héritabilité de la maladie.

Telle que décrite par Maher (Maher 2008), Manolio et al. (Manolio, Collins et al. 2009) et Marian (Marian 2012), et outre l'identification d'effets individuels de variants, l'héritabilité manquante pourrait aussi s'expliquer par :

- des **interactions géniques** (aussi appelées phénomènes d'**épistasie**) non identifiées. L'importance des interactions entre les gènes est aujourd'hui de plus en plus mise en avant. L'épistasie correspond au phénomène par lequel l'effet d'un variant est affecté par un autre variant. Contrairement aux traits quantitatifs qui seraient principalement dus à des effets additifs plutôt qu'à des effets épistatiques (non-additifs), les maladies complexes seraient quant-à-elles dues à une combinaison de ces deux types d'effets. Ainsi, l'existence d'interaction entre des variants de prédisposition mis en évidence par les GWAS pourrait expliquer une partie de l'héritabilité manquante des maladies complexes. Dans la SEP, une analyse d'épistasie des variants de HLA-DRB1 a permis de mettre en évidence des combinaisons plus ou moins à risques pour la SEP. Ainsi, bien que l'allèle HLA-DRB1\*15 soit communément considéré comme l'allèle majeur de risque à la SEP, avec un risque allélique de 3, sa combinaison avec l'allèle HLA-DRB1\*14 diminue le risque associé au seul allèle HLA-DRB1\*15. En revanche, l'association de HLA-DRB1\*15 avec HLA-DRB1\*08 confère un risque génotypique relatif de 6,5, soit presque autant que l'homozygote HLA-DRB1\*15 (voir **Figure 6** page 44) (Ramagopalan and Ebers 2009)

- des **interactions gènes x environnements** (GxE). Une interaction biologique est définie comme l'effet conjoint de deux facteurs qui agissent ensemble dans une réaction physique ou chimique et la participation commune d'au moins deux facteurs dans le même mécanisme entraînant le développement de la maladie. Les facteurs de risques sont dits « interagissant » si l'effet d'un facteur de risque dépend de l'autre. La prise en



compte des interactions GxE pour l'identification de facteurs génétiques de risques pourrait augmenter la puissance de détection de gènes ayant de faibles effets individuels, d'autant plus si le gène n'a d'effet que dans un sous-groupe de patients donné, défini par une exposition environnementale précise (Dempfle, Scherag et al. 2008). Dans le cas de la SEP, une interaction taux de vitamine D/locus HLA-DRB1 a été identifiée comme associée à la maladie. En effet, HLA-DRB1 est le seul locus HLA pouvant posséder dans son promoteur un élément de réponse à la vitamine D (VDRE, Vitamin D Response Element) (voir partie 1.1.2.1 pour le rôle potentiel de la vitamine D dans la SEP). De manière intéressante, tous les haplotypes HLA-DRB1\*15 possèdent le VDRE dans le promoteur d'HLA-DRB1, alors qu'il est moins conservé dans les autres haplotypes non associés à la SEP. Des études fonctionnelles ont montré que la présence d'un VDRE dans le promoteur de HLA-DRB1 influençait l'expression de ce gène (Ramagopalan, Maugeri et al. 2009).

- des **phénomènes d'épigénétique**. L'épigénétique consiste en la régulation de la transcription par des modifications transmissibles et réversibles de l'ADN ou des histones, sans altération de la séquence d'ADN. Ces modifications se font soit par l'ajout ou la suppression de groupements méthyls, acétyls, phosphates, ribosyls et ubiquitines des histones ou par l'ajout de groupements méthyls sur les cytosines des îlots CpG des promoteurs. Ces ajouts/suppressions se font par l'intermédiaire d'histones acétyltransférases/déacétylases ou des ADN méthyltransférases (DNMT, DNA méthyltransférases). La régulation épigénétique est un processus dynamique qui peut changer et s'adapter en réponse à des signaux environnementaux ou développementaux. L'influence de facteurs épigénétiques dans la SEP pourrait être confortée par le faible taux de concordance observé chez les jumeaux monozygotes (26%). Cela suggère entre autres que des facteurs environnementaux ou l'effet de ces facteurs sur la transcription de l'ADN jouerait un rôle dans la SEP. Ainsi, l'inactivation du chromosome X, conséquence de mécanismes épigénétiques, pourrait expliquer le sex ratio en défaveur des femmes observé dans la SEP (voir le préambule). De plus, de manière générale, il a été montré que les lymphocytes T naïfs, lorsqu'ils se différencient en lymphocytes de type  $T_H1$  et  $T_H2$ , présentent une méthylation de l'ADN. Une faible méthylation de l'ADN des cellules T a été associée avec l'induction de lymphocytes T auto-réactifs. Par exemple, il a été montré que les cellules T de patients atteints de polyarthrite rhumatoïde présentaient une diminution de la méthylation de l'ADN due à une activité réduite de la DNMT (Kurtuncu and Tuzun 2008). Dans le domaine de la SEP, une étude de l'épigénome des lymphocytes T issus de 3 paires de jumeaux monozygotes discordants pour





la maladie n'a pas identifié de différences significatives de la méthylation des îlots CpG (Baranzini, Mudge et al. 2010).

Depuis quelques années, des études plus ciblées, **restreintes à des gènes dont les effets fonctionnels sont connus dans des modèles expérimentaux**, ont également permis la mise en évidence de gènes de prédisposition. C'est ainsi qu'un haplotype du gène VAV1 a été mis en évidence comme étant associé à la prédisposition à la SEP. VAV1 code pour une protéine transductrice du signal dans les lymphocytes et est localisé dans une région génomique connue comme régulant l'EAE chez le rat. L'analyse de 7 cohortes représentant 12735 individus a permis d'identifier un haplotype de deux SNPs, localisé dans le premier intron du gène VAV1, associé à la prédisposition à la SEP. La combinaison à risque est associée avec une plus forte expression de VAV1, observée chez les patients SEP. L'expression de VAV1 est aussi corrélée à l'expression du TNF et de l'IFN- $\gamma$  dans les cellules du sang périphérique et du liquide céphalo-rachidien (Jagodic, Colacios et al. 2009). Ces **études fonctionnelles** de gènes candidats par fonction constituent une approche complémentaire aux GWAS et peuvent, comme cela a été le cas pour les gènes IL7R et IL2R, permettre l'identification de nouveaux facteurs de prédisposition à la SEP.

Au cours de mon doctorat, je ne me suis intéressé qu'à la part purement génétique de l'héritabilité manquante et je me suis donc attaché à comprendre pourquoi les études antérieures de génétique de la SEP n'ont pas abouti à l'identification de la totalité des variants génétiques de prédisposition à la maladie (Manolio, Collins et al. 2009, Wang, Li et al. 2010, Marian 2012). Dans les paragraphes suivants seront expliquées et développées plusieurs des hypothèses avancées ces dernières années pour identifier les 83% d'héritabilité manquants.



## **2.1. Hypothèse 1 : Manque de puissance statistique liée à la taille des échantillons étudiés**

La première hypothèse, qui est la plus simple à formuler, postule que la plupart des GWAS publiés jusqu'à présent manquent de puissance statistique pour identifier tous les variants génétiques associés aux maladies complexes. Ceci est très bien illustré dans la SEP, où les GWAS publiés entre 2007 et 2010 n'ont permis chacun d'identifier qu'un petit nombre de marqueurs. Le fait que l'IMSGC entreprenne un GWAS sur près de 10 000 patients a considérablement augmenté la puissance statistique de l'étude et permis la confirmation et/ou l'identification de 52 marqueurs de prédisposition à la SEP. Il est donc logique de penser qu'augmenter de nouveau la taille des échantillons nous permettrait d'identifier de nouveaux marqueurs génétiques de prédisposition à la maladie (Polychronakos 2011).

**Cette hypothèse a été explorée au cours de mon doctorat à travers deux projets qui ont abouti à deux publications:**

- le premier a consisté à valider les 5 SNPs décrits comme potentiellement associés à la prédisposition à la SEP dans le GWAS de l'IMSGC de 2011 mais qui n'avaient pas atteint le seuil de significativité *genome-wide* de  $5 \times 10^{-8}$ . Ces polymorphismes sont localisés dans les régions des gènes MANBA/NFKB1, CXCR5, SOX8, RPS6KB1, ZBTB46/TNFRSF6B (*Publication IMSGC, Brain 2013 présentée page 85*)

- le second a porté sur l'étude du gène ANKRD55, associé à la prédisposition à une autre maladie auto-immune, la polyarthrite rhumatoïde, mais qui n'avait pas atteint le seuil d'association significatif pour la SEP dans une précédente étude gène-candidat espagnole (Alloza, Otaegui et al. 2012) (*Publication Lill et al., J Med Genet 2013 présentée page 95*)



**Article (co-auteur):**

*MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci  
for multiple sclerosis*

International Multiple Sclerosis Genetics Consortium, Lill CM, Schjeide BM, Graetz C, Ban M, Alcina A, Ortiz MA, Pérez J, **Damotte V**, Booth D, Lopez de Lapuente A, Broer L, Schilling M, Akkad DA, Aktas O, Alloza I, Antigüedad A, Arroyo R, Blaschke P, Buttmann M, Chan A, Compston A, Cournu-Rebeix I, Dörner T, Epplen JT, Fernández O, Gerdes LA, Guillot-Noël L, Hartung HP, Hoffjan S, Izquierdo G, Kemppinen A, Kroner A, Kubisch C, Kümpfel T, Li SC, Lindenberger U, Lohse P, Lubetzki C, Luessi F, Malhotra S, Mescheriakova J, Montalban X, Papeix C, Paredes LF, Rieckmann P, Steinhagen-Thiessen E, Winkelmann A, Zettl UK, Hintzen R, Vandenbroeck K, Stewart G, Fontaine B, Comabella M, Urcelay E, Matesanz F, Sawcer S, Bertram L, Zipp F

*Brain* 2013 Jun ;136(Pt 6) :1778 :82



**Introduction :** Le GWAS publié par l'IMSGC en 2011 (IMSGC, WTCCC2 et al. 2011) a permis la confirmation de 23 SNPs précédemment publiés comme étant associés à la SEP et l'identification de 29 nouveaux SNPs. De plus, 5 SNPs (rs228614 [*MANBA/NFKBI*], chromosome 4 ; rs630923 [*CXCR5*], chromosome 11 ; rs2744148 [*SOX8*], chromosome 16 ; rs180515 [*RPS6KBI*], chromosome 17 ; rs6062314 [*ZBTB46/TNFRSF6B*], chromosome 20) semblaient montrer une association avec la prédisposition à la SEP mais n'avaient pas atteint le seuil de significativité de  $5 \times 10^{-8}$ . L'étude menée en collaboration avec une équipe allemande avait pour but de valider l'association entre ces 5 SNPs et la SEP.

**Matériel et Méthodes :** 8499 patients, 8765 contrôles et 958 familles trios provenant de cohortes françaises, allemandes, australiennes, espagnoles et anglaises ont été génotypés pour les 5 SNPs précédemment cités.

Pour chaque SNP, une méta-analyse a été conduite pour combiner les résultats d'associations avec ceux du GWAS de l'IMSGC de 2011.

**Résultats :** Nous avons validé l'association des 5 SNPs avec la prédisposition à la SEP, les résultats d'association étant en dessous du seuil de  $5 \times 10^{-8}$  (rs228614 [*MANBA/NFKBI*],  $p=3,4 \times 10^{-12}$ ; rs630923 [*CXCR5*],  $p=4,7 \times 10^{-10}$ ; rs2744148 [*SOX8*],  $p=1,6 \times 10^{-12}$ ; rs180515 [*RPS6KBI*],  $p=2,3 \times 10^{-13}$ ; rs6062314 [*ZBTB46/TNFRSF6B*],  $p=2,3 \times 10^{-8}$ )

**Conclusions :** Cette étude a permis de confirmer l'association des 5 SNPs qui semblaient être associés à la prédisposition à la SEP suite au GWAS de l'IMSGC. Cette étude augmente donc à 57 le nombre de variants génétiques associés à la prédisposition à la SEP.

Des études de cartographies fines sont maintenant nécessaires pour identifier le ou les variant(s) causal(aux). De plus, des études fonctionnelles sont nécessaires pour déterminer les rôles exacts de ces variants dans la physiopathologie de la SEP.

**Ma contribution :** Ma participation à ce projet a été de génotyper la cohorte française (846 patients et 266 contrôles provenant de la cohorte REFGENSEP) sur la plateforme OpenArray® (Applied Biosystems). J'ai également effectué les contrôles qualités sur ces échantillons et les ai corrélés avec leurs données cliniques. J'ai contribué à l'écriture et à la relecture de l'article.





## MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis

International Multiple Sclerosis Genetics Consortium<sup>#</sup>

<sup>#</sup>A list of all authors and their affiliations can be found in Appendix 1;

Correspondence to: Prof. Dr. Frauke Zipp,  
Rhine Main Neuroscience Network (rmn2),  
Johannes Gutenberg University Centre Mainz,  
Neurology Department,  
Langenbeckstr. 1,  
55131 Mainz, Germany,  
E-mail: frauke.zipp@unimedizin-mainz.de

Correspondence may also be addressed to: Dr. Christina M. Lill, Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Neuropsychiatric Genetics Group, Ihnestr. 63-73, 14195 Berlin, Germany, E-mail: lill@molgen.mpg.de

A recent genome-wide association study reported five loci for which there was strong, but sub-genome-wide significant evidence for association with multiple sclerosis risk. The aim of this study was to evaluate the role of these potential risk loci in a large and independent data set of ~20 000 subjects. We tested five single nucleotide polymorphisms rs228614 (MANBA), rs630923 (CXCR5), rs2744148 (SOX8), rs180515 (RPS6KB1), and rs6062314 (ZBTB46) for association with multiple sclerosis risk in a total of 8499 cases with multiple sclerosis, 8765 unrelated control subjects and 958 trios of European descent. In addition, we assessed the overall evidence for association by combining these newly generated data with the results from the original genome-wide association study by meta-analysis. All five tested single nucleotide polymorphisms showed consistent and statistically significant evidence for association with multiple sclerosis in our validation data sets (rs228614: odds ratio = 0.91,  $P = 2.4 \times 10^{-6}$ ; rs630923: odds ratio = 0.89,  $P = 1.2 \times 10^{-4}$ ; rs2744148: odds ratio = 1.14,  $P = 1.8 \times 10^{-6}$ ; rs180515: odds ratio = 1.12,  $P = 5.2 \times 10^{-7}$ ; rs6062314: odds ratio = 0.90,  $P = 4.3 \times 10^{-3}$ ). Combining our data with results from the previous genome-wide association study by meta-analysis, the evidence for association was strengthened further, surpassing the threshold for genome-wide significance ( $P < 5 \times 10^{-8}$ ) in each case. Our study provides compelling evidence that these five loci are genuine multiple sclerosis susceptibility loci. These results may eventually lead to a better understanding of the underlying disease pathophysiology.

**Keywords:** multiple sclerosis; complex genetics; genetic risk; immunogenetics; genetic association

**Abbreviations:** GWAS = genome-wide association study; SNP = single nucleotide polymorphism

### Introduction

Multiple sclerosis is the most common inflammatory demyelinating disease of the CNS that is likely caused by an interplay of genetic and environmental risk factors. Apart from several independent association signals in the *HLA* (human leukocyte antigen) region

on chromosome 6p21, a recent genome-wide association study (GWAS) in multiple sclerosis has reported 52 loci exerting small to moderate risk effects (IMSGC and WTCCC2, 2011). In addition, five additional loci provided strong support for association ( $P < 5 \times 10^{-7}$ ) in that GWAS, but failed to meet current criteria for genome-wide significance ( $P < 5 \times 10^{-8}$ ). The most

significantly associated single-nucleotide polymorphisms (SNPs) in these regions were rs228614 in *MANBA* (mannosidase, beta A, lysosomal), rs630923 upstream of *CXCR5* (chemokine C-X-C motif receptor 5), rs2744148 downstream of *SOX8* (sex determining region Y-box 8), rs180515 downstream of *RPS6KB1* (ribosomal protein S6 kinase, 70 kDa, polypeptide 1), and rs6062314 in *ZBTB46* (zinc finger and BTB domain containing 46) (IMSGC and WTCCC2, 2011). Given the lack of genome-wide significance, independent validation efforts are needed to further discern the putative role of these loci in multiple sclerosis risk. To this end, we have tested these five SNPs for association with multiple sclerosis risk in a multicentric study comprising 20 138 subjects of European descent who were independent of the original GWAS sample (IMSGC and WTCCC2, 2011).

## Materials and methods

### Power analysis

Power was estimated using the Genetic Power Calculator (Purcell *et al.*, 2003) assuming a one-sided  $\alpha$  of 0.01 and a disease prevalence of 0.1%.

### Data sets

The current study included a total of 8805 multiple sclerosis cases and 8981 unrelated control subjects of self-reported European descent from Germany, Spain, France, The Netherlands, and Australia, as well as 963 trios from the UK. Subjects were selected specifically to be non-overlapping with the original study (IMSGC and WTCCC2, 2011). Diagnosis of multiple sclerosis was established according to standard diagnostic criteria (Poser *et al.*, 1983; McDonald *et al.*, 2001). All samples were collected with informed written consent and appropriate ethical approval at the respective sites. The effective sample size after quality control comprised 8499 multiple sclerosis cases, 8765 unrelated control subjects, and 958 trios (see below and Supplementary Table 1).

### Genotyping and quality control

Genotyping for the German, Spanish and British samples was performed at the individual sites using single-assay allelic discrimination assays based on TaqMan<sup>®</sup> chemistry following the manufacturer's instructions (Applied Biosystems, Inc.). The French subjects were TaqMan<sup>®</sup> genotyped using the multiplex 'OpenArray' platform

(Applied Biosystems, Inc.), the Australian subjects were genotyped using the MassARRAY iPLEX system (Sequenom, Inc.), and the Dutch genotypes were generated on the Human610-Quad Bead GWAS array (Illumina, Inc.). Samples with missing genotypes for more than two SNPs were excluded before analysis [applicable to a total of 115 samples (0.5%) across all data sets]. Information on sex and/or age at examination was available for >90% of subjects in all case-control data sets except in the sample from Central Spain. Samples with missing information in these categories ( $n = 407$ ) were excluded. The threshold for genotyping efficiency per SNP and data set was set to >95%. Hardy–Weinberg equilibrium was assessed in control subjects and in unaffected founders of the nuclear families. Deviations from Hardy–Weinberg equilibrium were defined as  $P < 0.05$  based on Pearson's  $\chi^2$  as implemented in PLINK v1.07 (Purcell *et al.*, 2007).

### Association analyses

All association analyses were performed using PLINK. For the case-unrelated control data sets, logistic regression with adjustment for age at examination and/or sex was performed where available (Supplementary Table 1) using an additive transmission model. Transmission equilibrium testing was applied to the UK trio data set. Odds ratios (OR) are displayed for the allele dosage of the minor allele as defined by the frequency in the overall data set. Meta-analyses across all validation data sets were based on fixed-effect models. The threshold for nominal significance was set to  $P < 0.01$  (i.e. applying a conservative Bonferroni correction for five tests). All  $P$ -values are one-sided with regard to the expected direction of effect based on the original study (IMSGC and WTCCC2, 2011). Between-study heterogeneity was quantified using the  $I^2$  metric, and statistical significance was assessed by the Q-test statistic. Forest plots were generated using a customized version of the 'rmeta' package in R language (Lill *et al.*, 2012). Two-sided unweighted  $P$ -values of the original GWAS (IMSGC and WTCCC2, 2011) and of this study were combined using METAL (Willer *et al.*, 2010).

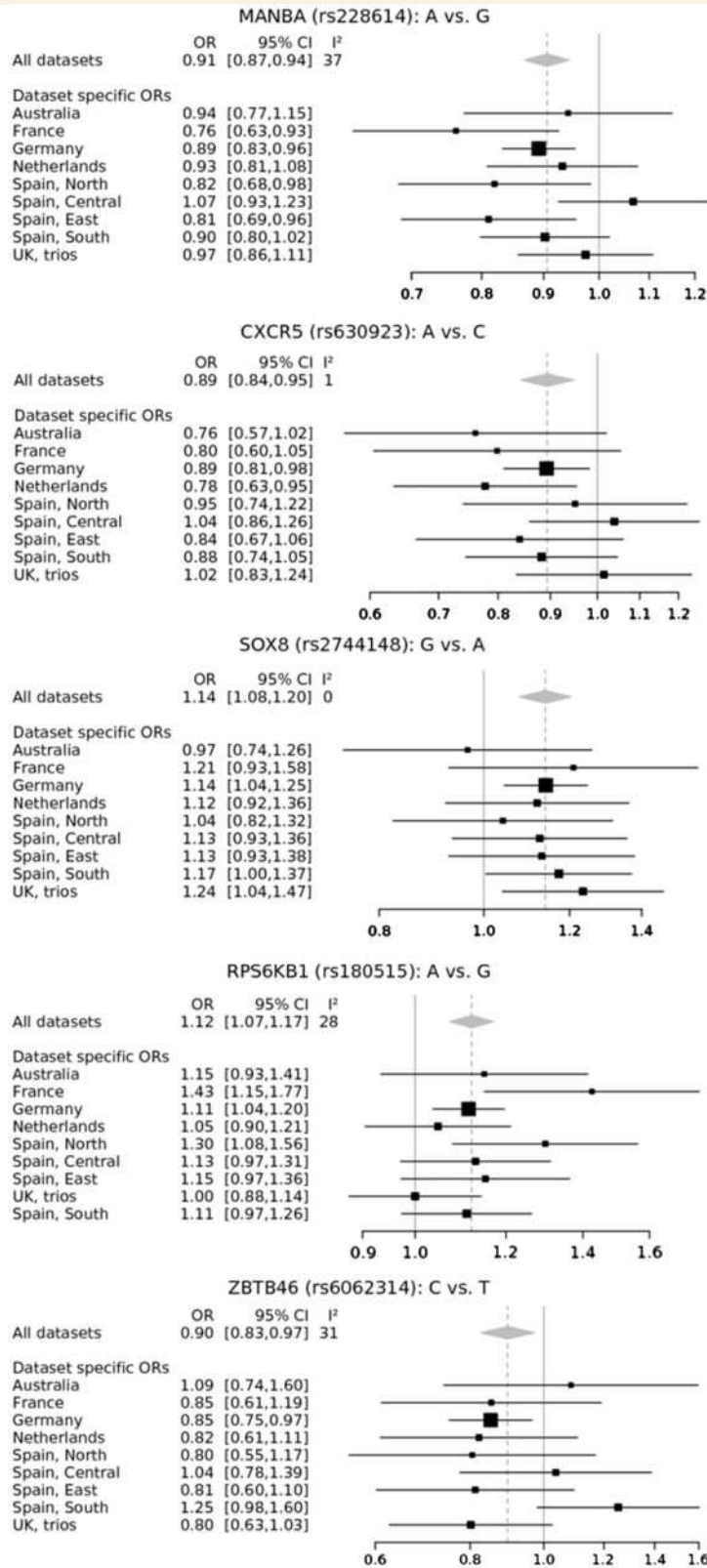
## Results

The combined effective replication data sets of 8499 cases, 8765 unrelated controls, and 958 trios had ~80% power to detect an odds ratio of 1.10 down to allele frequencies of 0.13. Control genotypes in all data sets were distributed according to Hardy–Weinberg equilibrium ( $P > 0.05$ ) for all SNPs. Total genotyping efficiency was >98% for each SNP (Table 1).

**Table 1 Association results for the five loci and multiple sclerosis assessed in 20 138 subjects of European descent**

SNP	Location (hg19)	Nearest gene	Validation data sets							Original study		Combined P**
			Eff.	MAF	OR (95% CI)	P*	I <sup>2</sup> (95% CI)	P <sub>Q</sub>	OR	P**		
rs228614 (A/G)	chr4:103,578,637	<i>MANBA</i> (intronic)	99.0	47.6	0.91 (0.87–0.95)	$2.4 \times 10^{-6}$	37 (0–71)	0.120	0.92	$1.4 \times 10^{-7}$	$3.4 \times 10^{-12}$	
rs630923 (A/C)	chr11:118,754,353	<i>CXCR5</i> (122 bp 5')	98.3	15.7	0.89 (0.84–0.95)	$1.2 \times 10^{-4}$	1 (0–65)	0.425	0.89	$2.8 \times 10^{-7}$	$4.7 \times 10^{-10}$	
rs2744148 (G/A)	chr16:1,073,552	<i>SOX8</i> (36,573 bp 3')	99.5	16.8	1.14 (1.08–1.20)	$1.8 \times 10^{-6}$	0 (0–14)	0.915	1.12	$8.4 \times 10^{-8}$	$1.6 \times 10^{-12}$	
rs180515 (G/A)	chr17:58,024,275	<i>RPS6KB1</i> (3' UTR)	99.1	35.5	1.12 (1.07–1.17)	$5.2 \times 10^{-7}$	28 (0–66)	0.198	1.09	$8.8 \times 10^{-8}$	$2.3 \times 10^{-13}$	
rs6062314 (C/T)	chr20:62,409,713	<i>ZBTB46</i> (intronic)	99.1	7.9	0.90 (0.83–0.97)	$4.3 \times 10^{-3}$	31 (0–68)	0.169	0.86	$1.3 \times 10^{-7}$	$2.3 \times 10^{-8}$	

Fixed effect meta-analysis results for the SNPs tested across all validation data sets were performed using PLINK. The association results from the original study (IMSGC and WTCCC2, 2011) and this study were combined using METAL. Allele names are displayed as minor/major allele based on frequencies in the entire validation data set. Brackets following the gene name list the location of the SNP relative to the gene, base pairs (bp) indicate upstream (5') or downstream (3') distance to the primary transcript (as annotated on the UCSC Genome Browser). CI = confidence interval; Eff. = genotyping efficiency (in %); hg19 = human genome build 19; MAF = minor allele frequency in controls (in %); OR = odds ratio; UTR = untranslated region; \* = one-sided; \*\* = two-sided.



**Figure 1** Meta-analysis of validation data sets assessing the association between the *MANBA*, *CXCR5*, *SOX8*, *RPS6KB1* and *ZBTB46* loci and multiple sclerosis risk in populations of European descent. The x-axis depicts the odds ratio (OR). Study-specific odds ratios (black squares) and 95% confidence intervals (CIs, lines) were calculated using an additive model. The summary odds ratios and 95% confidence intervals (grey diamonds) were calculated based on fixed-effect meta-analysis.

Fixed-effect meta-analysis across all validation data sets revealed highly significant associations of all five tested SNPs and multiple sclerosis risk in the validation data sets, i.e. rs228614 (*MANBA*, OR = 0.91,  $P = 2.4 \times 10^{-6}$ ), rs630923 (*CXCR5*, OR = 0.89,  $P = 1.2 \times 10^{-4}$ ), rs2744148 (*SOX8*, OR = 1.14,  $P = 1.8 \times 10^{-6}$ ), rs180515 (*RPS6KB1*, OR = 1.12,  $P = 5.2 \times 10^{-7}$ ), and rs6062314 (*ZBTB46*, OR = 0.90,  $P = 4.3 \times 10^{-3}$ ). Effect estimates were similar to those originally reported (IMSGC and WTCCC2, 2011). There was no evidence for substantial between-study heterogeneity for any of the five SNPs (Fig. 1 and Table 1). Combining our results with *P*-values from the original GWAS (IMSGC and WTCCC2, 2011) increased the statistical support of our findings to genome-wide significance for each of the five tested SNPs: rs228614 (*MANBA*),  $P = 3.4 \times 10^{-12}$ , rs630923 (*CXCR5*),  $P = 4.7 \times 10^{-10}$ , rs2744148 (*SOX8*),  $P = 1.6 \times 10^{-12}$ , rs180515 (*RPS6KB1*),  $P = 2.3 \times 10^{-13}$ , and rs6062314 (*ZBTB46*),  $P = 2.3 \times 10^{-8}$  (Table 1).

## Discussion

Our study shows that common genetic variants in or near *MANBA*, *CXCR5*, *SOX8*, *RPS6KB1*, and *ZBTB46*, are associated with multiple sclerosis risk at genome-wide significance. Our results, thus, provide compelling evidence that these loci represent genuine genetic risk factors for multiple sclerosis.

As is the case for the majority of genetic associations, the precise molecular genetic mechanisms underlying these results still remain to be assessed. That is, future studies need to clarify whether the SNPs tested here are directly involved in altering gene expression/protein function or whether such effects are exerted by other correlated variants, possibly located in neighbouring genes. For instance, SNP rs180515 in the 3' UTR of *RPS6KB1* is located in the seed region of a predicted micro-RNA binding site for hsa-miR-3616-5p and hsa-miR-573 and may thus directly alter *RPS6KB1* translation (Supplementary Fig. 1) (Schilling, 2012). The intronic SNP rs228614 in *MANBA* is in substantial linkage disequilibrium with two non-synonymous SNPs in the same gene [rs2866413 (p.Thr701Met),  $r^2 = 0.87$ , and rs227368 (p.Val253Leu),  $r^2 = 0.74$ , based on 1000 Genomes Pilot 1 CEU data (1000 Genomes Project Consortium, 2010)], which may affect protein function. However, and possibly more importantly, of all five loci tested here *MANBA* is the only one to contain SNPs (including rs228614) showing strong *cis* effects on messenger RNA expression based on recently published data (Yang *et al.*, 2010) (Supplementary Fig. 2). The intronic SNP rs6062314 in *ZBTB46* displays only moderate linkage disequilibrium to potentially functional variants, i.e. a non-synonymous SNP in *LIME1* [Lck interacting transmembrane adaptor 1, SNP rs1151625 (p.Pro211Leu),  $r^2 = 0.39$ ], and in *ZGPAT* [zinc finger, CCCH-type with G patch domain, SNP rs1291212 (p.Ser61Arg),  $r^2 = 0.31$ ]. In addition, *TNFRSF6B* (tumour necrosis factor receptor superfamily, member 6b, decoy) is also located in this chromosomal region and would represent a compelling candidate based on its implications on T cell function (e.g. Zhang *et al.*, 2001). However, the only coding sequence SNP displaying noteworthy linkage disequilibrium with rs6062314 in this gene does not invoke an amino acid change (rs2738787,  $r^2 = 0.36$ ). Finally,

rs630923 maps into a potential *CXCR5* transcription factor binding site for the nuclear factor of kappa light polypeptide gene enhancer in B-cells (NFkB) in a region of DNase I hypersensitivity (ENCODE Project Consortium, 2012). Rs630923 is predicted to alter NFkB binding (Boyle *et al.*, 2012) and could thus, potentially affect *CXCR5* transcription.

It should be emphasized that the abovementioned potential functional consequences are based on *in silico* assessments and require experimental testing and validation. It is also possible that hitherto unknown, rare DNA sequence variants underlie or contribute to the observed association signals.

In summary, our study provides compelling evidence that the list of established multiple sclerosis risk genes can now be extended by five additional loci, all of which show genome-wide significant association with disease risk. Further fine-mapping and functional studies are required to elucidate the biochemical and pathophysiological mechanisms underlying these associations.

## Acknowledgements

We are grateful to the individuals participating in this study. We acknowledge use of the cohort of the CRB-REFGENSEP and thank ICM, CIC Pitié-Salpêtrière, Généthron and REFGENSEP's members for their help and support.

## Funding

This project was funded by grants from the German Ministry for Education and Research (BMBF) and German Research Foundation (DFG; to F.Z.), the BMBF (grant 16SV5538) and the Cure Alzheimer's Fund (to L.B.), the Walter- and Ilse-Rose-Stiftung and the BMBF (grant 01GM1203A; to H.-P.H. and O.A.), the BMBF (grant NBL3 to U.K.Z.; grant 01UW0808 to U.L., and E.S.-T.), the Innovation Fund of the Max Planck Society (M.FE.A.LD0002 to U.L.), the Fondo de Investigaciones Sanitarias FEDER-FIS PI10/1985 and Fundación Alicia Koplowitz (to E.U.), and the FEDER-FIS grant number PI12/00555 and MINECO-FEDER grant number SAF2009-11491 (to F.M. and A.A.). The research leading to these results has received funding from the program 'Investissements d'avenir' ANR-10-IAIHU-06. This project was supported by INSERM, ARSEP, AFM, GIS-IBISA, the National Institutes of Health (grant RO1NS049477) and the Cambridge NIHR Biomedical Research Centre. C.M.L. was supported by the Fidelity Biosciences Research Initiative. E.U. works for the Fundación de Investigación Biomédica del Hospital Clínico San Carlos-IdISSC. S.M. is an early stage researcher funded by the European Community's Seventh Framework Programme ([FP7/2007-2013] under grant agreement n° 212877 (UEPHA\*multiple sclerosis).

## Supplementary material

Supplementary material is available at *Brain* online.

## References

- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; 22: 1790–7.
- ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57–74.
- IMSGC, WTCCC2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; 476: 214–19.
- Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet* 2012; 8: e1002548.
- McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; 50: 121–7.
- Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983; 13: 227–31.
- Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19: 149–50.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559–75.
- Schilling M. In silico assessment of the effects of single nucleotide polymorphisms on miRNA-mRNA interactions. Bachelor of Science in Bioinformatics, 2012. Max Planck Institute for Molecular Genetics, Free University Berlin.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–1.
- Yang T-P, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010; 26: 2474–6.
- Zhang J, Salcedo TW, Wan X, Ullrich S, Hu B, Gregorio T, et al. Modulation of T-cell responses to alloantigens by TR6/DcR3. *J Clin Invest* 2001; 107: 1459–68.
- 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467: 1061–73.
- Catherine Lubetzki,<sup>7,18</sup> Felix Luessi,<sup>2</sup> Sunny Malhotra,<sup>6</sup> Julia Mescheriakova,<sup>10</sup> Xavier Montalban,<sup>6</sup> Caroline Papeix,<sup>7,18</sup> Lidia F. Paredes,<sup>5</sup> Peter Rieckmann,<sup>16,†</sup> Elisabeth Steinhagen-Thiessen,<sup>27</sup> Alexander Winkelmann,<sup>15</sup> Uwe K. Zettl,<sup>15</sup> Rogier Hintzen,<sup>10</sup> Koen Vandenbroeck,<sup>9,28</sup> Graeme Stewart,<sup>8</sup> Bertrand Fontaine,<sup>7,18</sup> Manuel Comabella,<sup>6</sup> Elena Urcelay,<sup>5</sup> Fuencisla Matesanz,<sup>4</sup> Stephen Sawcer,<sup>3</sup> Lars Bertram<sup>1,‡</sup> and Frauke Zipp<sup>2,‡</sup>, on behalf of the International Multiple Sclerosis Genetics Consortium.
- <sup>1</sup>Neuropsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; <sup>2</sup>Department of Neurology, Focus Program Translational Neuroscience, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; <sup>3</sup>University of Cambridge, Department of Clinical Neurosciences, Addenbrookes Hospital, Cambridge, UK; <sup>4</sup>Instituto de Parasitología y Biomedicina 'López Neyra', Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), Granada, Spain; <sup>5</sup>Immunology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos- IdISSC, Madrid, Spain; <sup>6</sup>Department of Neurology-Neuroimmunology, Centre d'Esclerosi Múltiple de Catalunya, Cemcat, Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Spain; <sup>7</sup>UPMC-INSERM-CNRS-ICM, UMR 975-7225, Institut Cerveau Moelle Epinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France; <sup>8</sup>Institute for Immunology and Allergy Research, Westmead Millennium Institute, University of Sydney, Sydney, Australia; <sup>9</sup>Neurogenomik Laboratory, Department of Neuroscience, University of the Basque Country UPV/EHU, Leioa, Spain; <sup>10</sup>MS Centre ErasMS, Erasmus MC, University Medical Center, Rotterdam, The Netherlands; <sup>11</sup>Department of Human Genetics, Ruhr University, Bochum, Germany; <sup>12</sup>Department of Neurology, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany; <sup>13</sup>Servicio de Neurología, Hospital de Basurto, Bilbao, Spain; <sup>14</sup>Multiple Sclerosis Unit, Neurology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos- IdISSC, Madrid, Spain; <sup>15</sup>Department of Neurology, University of Rostock, Rostock, Germany; <sup>16</sup>Department of Neurology, University of Würzburg, Würzburg, Germany; <sup>17</sup>Department of Neurology, St. Josef-Hospital, Ruhr-University, Bochum, Germany; <sup>18</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Département de Neurologie, Hôpital Pitié-Salpêtrière, Paris, France; <sup>19</sup>Department of Medicine, Rheumatology, and Clinical Immunology & DRFZ, Charité University Medicine, Berlin, Germany; <sup>20</sup>Servicio de Neurología, Instituto de Neurociencias Clínicas, Hospital Regional Universitario Carlos Haya, Málaga, Spain; <sup>21</sup>Institute for Clinical Neuroimmunology, Ludwig Maximilian University, Munich, Germany; <sup>22</sup>Unidad de Esclerosis Múltiple, Hospital Virgen Macarena, Sevilla, Spain; <sup>23</sup>Institute of Human Genetics, University of Ulm, Ulm, Germany; <sup>24</sup>Max Planck Institute for Human Development, Berlin, Germany; <sup>25</sup>Department of Psychology, Technische Universität Dresden, Dresden, Germany; <sup>26</sup>Department of Clinical Chemistry, Ludwig Maximilian University, Munich, Germany; <sup>27</sup>Interdisciplinary Metabolic Center, Lipids Clinic, Charité University Medicine, Berlin, Germany; <sup>28</sup>IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
- <sup>†</sup>Current address: Centre for Research in Neuroscience at McGill University, Montreal, Canada (Antje Kroner), Institute of Laboratory Medicine and Human Genetics, Singen, Germany (Peter Lohse), Department of Neurology, Sozialstiftung Bamberg Hospital, Bamberg, Germany (Peter Rieckmann)
- <sup>‡</sup>These authors contributed equally to this work

## Appendix 1

### List of authors

Christina M. Lill,<sup>1,2</sup> Brit-Maren M. Schjeide,<sup>1</sup> Christiane Graetz,<sup>1,2</sup> Maria Ban,<sup>3</sup> Antonio Alcina,<sup>4</sup> Miguel A. Ortiz,<sup>5</sup> Jennifer Pérez,<sup>6</sup> Vincent Damotte,<sup>7</sup> David Booth,<sup>8</sup> Aitzkoa Lopez de Lapuente,<sup>9</sup> Linda Broer,<sup>10</sup> Marcel Schilling,<sup>1</sup> Denis A. Akkad,<sup>11</sup> Orhan Aktas,<sup>12</sup> Iraide Alloza,<sup>9</sup> Alfredo Antigüedad,<sup>13</sup> Rafa Arroyo,<sup>14</sup> Paul Blaschke,<sup>15</sup> Mathias Buttman,<sup>16</sup> Andrew Chan,<sup>17</sup> Alastair Compston,<sup>3</sup> Isabelle Cournu-Rebeix,<sup>7,18</sup> Thomas Dörner,<sup>19</sup> Joerg T. Epplen,<sup>11</sup> Óscar Fernández,<sup>20</sup> Lisa-Ann Gerdes,<sup>21</sup> Léna Guillot-Noël,<sup>7</sup> Hans-Peter Hartung,<sup>12</sup> Sabine Hoffman,<sup>11</sup> Guillermo Izquierdo,<sup>22</sup> Anu Kempainen,<sup>3</sup> Antje Kroner,<sup>16,†</sup> Christian Kubisch,<sup>23</sup> Tania Kümpfel,<sup>21</sup> Shu-Chen Li,<sup>24,25</sup> Ulman Lindenberger,<sup>24</sup> Peter Lohse,<sup>26,†</sup>



**Article (co-auteur):**

Genome-wide significant association of *ANKRD55* rs6859219  
and multiple sclerosis risk

Lill CM, Schjeide BM, Graetz C, Liu T, **Damotte V**, Akkad DA, Blaschke P, Gerdes LA, Kroner A, Luessi F, Cournu-Rebeix I, Hoffjan S, Winkelmann A, Touze E, Pico F, Corcia P, Otaegui D, Antigüedad A, Alcina A, Comabella M, Montalban X, Olascoaga J, Matesanz F, Dörner T, Li SC, Steinhagen-Thiessen E, Lindenberger U, Chan A, Rieckmann P, Hartung HP, Aktas O, Lohse P, Buttmann M, Kümpfel T, Kubisch C, Zettl UK, Epplen JT, Fontaine B, Zipp F, Vandenbroeck K, Bertram L

*J Med Genet.* 2013 Mar;50(3):140-3.





**Introduction** : Des GWAS ont été menés au cours des dernières années pour plusieurs maladies autoimmunes, mettant en évidence des facteurs de risques communs à plusieurs de ces maladies. Une étude espagnole a étudié l'association à la SEP de dix variants de risques dits auto-immuns, dont le SNP rs6859219, appartenant au gène *ANKRD55*, identifié comme associé à la polyarthrite rhumatoïde (Alloza, Otaegui et al. 2012). Cette étude, portant sur 2895 patients SEP espagnols et 2942 contrôles, suggérait l'association de ce polymorphisme avec la SEP. Cependant le résultat d'association ( $p=1,9 \times 10^{-7}$ ) était légèrement supérieur au seuil de significativité *genome-wide* de  $5 \times 10^{-8}$ .

Nous avons initié une collaboration Franco-Allemande dans le but de valider l'association du gène *ANKRD55* avec la SEP dans des cohortes indépendantes allemandes et françaises.

**Matériel et Méthodes** : 1344 patients et 768 contrôles français ainsi que 3762 patients et 2972 contrôles allemands ont été génotypés pour le SNP rs6859219.

Les résultats d'association dans la cohorte française et allemande ont été combinés au cours d'une méta-analyse avec ceux obtenus par l'étude espagnole ainsi qu'avec les résultats du GWAS IMSGC de 2007 (IMSGC, Hafler et al. 2007) et du GWAS GeneMSA (Baranzini, Galwey et al. 2009).

**Résultats** : L'analyse d'association valide le SNP rs6859219 comme étant un facteur de risque associé à la prédisposition à la SEP ( $p=3,1 \times 10^{-11}$  ; OR=1,19).

**Conclusions** : Cette étude a permis de mettre en évidence un nouveau facteur de risque génétique associé à la prédisposition à la SEP, portant à 58 le nombre de marqueurs associés à la maladie. Elle met également une fois de plus en exergue l'architecture génétique commune des maladies auto-immunes.

Cependant, le rôle que pourrait avoir le gène *ANKRD55* dans la prédisposition à la SEP est inconnu. En effet, tout d'abord, bien que le gène *ANKRD55* soit fortement exprimé dans les cellules effectrices CD4 mémoires, la fonction de la protéine Ankyrin repeat domain-containing protein 55 reste aujourd'hui inconnue. De plus, le SNP rs6859219 se trouve dans un intron du gène et il ne semble être en déséquilibre de liaison qu'avec d'autres variants introniques. Cependant, il semblerait que le SNP rs6859219 appartienne à un site de fixation de plusieurs facteurs de transcription dont l'activating enhancer binding protein 2 et l'early B cell factor 1.



**Ma contribution :** Ma participation à ce projet à consister à sélectionner les échantillons français et à compiler les données cliniques associées. J'ai également contribué à la rédaction et à la relecture de l'article.



## COMMUNICATIONS

Genome-wide significant association of *ANKRD55* rs6859219 and multiple sclerosis risk

Christina M Lill,<sup>1,2</sup> Brit-Maren M Schjeide,<sup>1</sup> Christiane Graetz,<sup>1,2</sup> Tian Liu,<sup>3</sup> Vincent Damotte,<sup>4</sup> Denis A Akkad,<sup>5</sup> Paul Blaschke,<sup>6</sup> Lisa-Ann Gerdes,<sup>7</sup> Antje Kroner,<sup>8,9</sup> Felix Luessi,<sup>2</sup> Isabelle Cournu-Rebeix,<sup>4,10</sup> Sabine Hoffjan,<sup>5</sup> Alexander Winkelmann,<sup>6</sup> Emmanuel Touze,<sup>11</sup> Fernando Pico,<sup>12</sup> Philippe Corcia,<sup>13</sup> David Otaegui,<sup>14</sup> Alfredo Antigüedad,<sup>15</sup> Antonio Alcina,<sup>16</sup> Manuel Comabella,<sup>17</sup> Xavier Montalban,<sup>17</sup> Javier Olascoaga,<sup>18</sup> Fuencisla Matesanz,<sup>16</sup> Thomas Dörner,<sup>19</sup> Shu-Chen Li,<sup>3,20</sup> Elisabeth Steinhagen-Thiessen,<sup>21</sup> Ulman Lindenberger,<sup>3</sup> Andrew Chan,<sup>22</sup> Peter Rieckmann,<sup>8,23</sup> Hans-Peter Hartung,<sup>24</sup> Orhan Aktas,<sup>24</sup> Peter Lohse,<sup>25</sup> Mathias Buttmann,<sup>8</sup> Tania Kümpfel,<sup>7</sup> Christian Kubisch,<sup>26</sup> Uwe K Zettl,<sup>6</sup> Joerg T Epplen,<sup>5</sup> Bertrand Fontaine,<sup>4,10</sup> Frauke Zipp,<sup>2</sup> Koen Vandenberg,<sup>27,28</sup> Lars Bertram<sup>1</sup>

For numbered affiliations see end of article.

**Correspondence to**

Dr Christina M Lill, Neuropsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany; lill@molgen.mpg.de

Received 9 November 2012  
Revised 10 December 2012  
Accepted 12 December 2012  
Published Online First  
12 January 2013

Multiple sclerosis (MS) is a genetically complex disease that shares a substantial proportion of risk loci with other autoimmune diseases.<sup>1</sup> Along these lines, *ANKRD55*, originally implicated in rheumatoid arthritis, was recently reported as a potential novel MS risk gene (rs6859219,  $p=1.9\times 10^{-7}$ ).<sup>2</sup> Here, we comprehensively validated this effect in independent datasets comprising 8846 newly genotyped subjects from Germany and France as well as 5003 subjects from two genome-wide association studies (GWAS). Upon meta-analysis of all available data (19 686 subjects), *ANKRD55* rs6859219 now shows compelling evidence for association with MS at genome-wide significance (OR=1.19,  $p=3.1\times 10^{-11}$ ). Our study adds *ANKRD55* to the list of established MS risk loci and extends previous evidence suggesting an overlapping genetic foundation across autoimmune diseases.

Ankyrin repeats are abundant in a large number of different proteins in humans and mediate protein–protein interactions. DNA-sequence variants in ankyrin repeat domain-containing proteins have been linked to a wide range of diseases; for example, *KRIT1* mutations causative for cerebral cavernous malformations,<sup>3</sup> *NOTCH3* mutations in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy, and *RFXANK* mutations in the bare lymphocyte syndrome.<sup>4</sup> *ANKRD55* (located on chromosome 5q11.2) encodes the ‘ankyrin repeat domain-containing protein 55’ the function of which is currently unknown. Single nucleotide polymorphism rs6859219 in *ANKRD55* was implicated in a recent GWAS meta-analysis on rheumatoid arthritis.<sup>5</sup> Furthermore, a joint analysis of datasets on rheumatoid arthritis and coeliac disease also indicated a role of *ANKRD55* in the latter.<sup>6</sup> Given the augmenting evidence suggesting an overlap in the genetic architecture of autoimmune diseases including MS, we have previously investigated 10

‘autoimmune loci’ in 2895 Spanish MS cases and 2942 controls.<sup>2</sup> In that study, rs6859219 emerged as a putative new MS locus albeit at subgenome-wide significance ( $p=1.9\times 10^{-7}$ ).<sup>2</sup> Our failure to establish genome-wide significance was likely owing to the comparatively small sample size; thus, we set out to corroborate our initial association finding in additional independent datasets and to assess the overall evidence for association by meta-analysis.

We genotyped rs6859219 in 5106 MS cases and 3740 healthy control subjects of self-reported European descent from Germany and France<sup>7–8</sup> (table 1) using a commercially available assay (‘TaqMan’, Applied Biosystems, Inc.). Furthermore, we obtained, reanalysed and included data on 1868 cases and 3135 controls for rs6859219 from two publicly available GWAS (‘IMSGC’<sup>9</sup> and ‘GeneMSA’;<sup>10</sup> in the latter, rs6859219 was analysed following imputation). GWAS quality control, imputation and analysis protocols were followed as described previously.<sup>8</sup> Combined, these replication datasets comprised 6974 cases and 6875 controls and had ~94% power to detect an OR of 1.20 at  $\alpha=1\times 10^{-4}$ . Power to detect association at genome-wide significance ( $\alpha=5\times 10^{-8}$ ) using all available data (9869 cases and 9817 controls, ie, after including the Spanish datasets of the original study) was ~96%.

Genotyping efficiency and accuracy (based on 5% duplicate samples) in the newly genotyped datasets were 99.0% and 100%, respectively. Genotypes in controls were distributed according to Hardy–Weinberg equilibrium ( $p=0.209$  using Pearson’s  $\chi^2$ ). Logistic regression analyses based on an additive model were adjusted for age and sex in the German and French datasets, and for principal components (PC 1–3) in IMSGC and GeneMSA to account for population substructure as previously described.<sup>8</sup> Fixed-effect meta-analysis revealed significant association of the *ANKRD55* rs6859219 C-allele with increased risk for MS across all

**To cite:** Lill CM, Schjeide B-MM, Graetz C, et al. *J Med Genet* 2013;**50**:140–143.

**Table 1** Demographic details of the German and French case-control datasets genotyped for ANKRD55 rs6859219

Sites	N cases (% females)	N controls (% females)	Mean AAE ( $\pm$ SD) cases	Mean AAE ( $\pm$ SD) controls	Mean AAO ( $\pm$ SD) cases
Germany	3762 (71)	2972 (60)	41 (11)	42 (17)	30 (10)
Bochum/Essen	1070 (71)	404 (43)	42 (11)	43 (12)	32 (10)
Duesseldorf/Koeln	257 (72)	829 (62)	39 (10)	44 (16)	–
Mainz/Berlin	787 (69)	869 (65)	38 (10)	34 (15)	29 (10)
Munich	595 (71)	400 (50)	42 (12)	42 (16)	30 (10)
Rostock	526 (74)	470 (70)	42 (12)	52 (19)	32 (11)
Wuerzburg	527 (69)	0 (0)	39 (11)	–	28 (10)
France (Paris)	1344 (72)	768 (60)	44 (12)	40 (13)	32 (10)
All	5106 (71)	3740 (60)	41 (12)	39 (16)	31 (10)

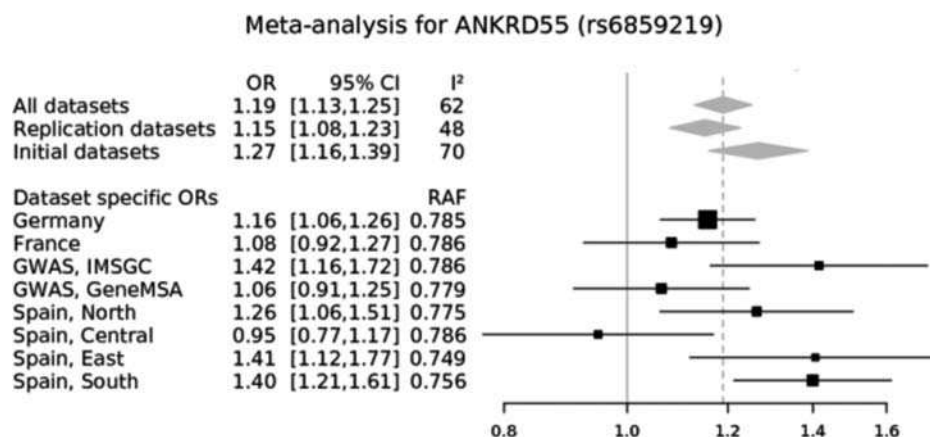
AAE, age at examination; AAO, age at onset; N, number.

replication datasets (OR=1.15, 95% CI 1.08 to 1.23,  $p=1.0\times 10^{-5}$ , figure 1). Inclusion of the Spanish case-control datasets now exceeded the threshold for genome-wide significance by more than three orders of magnitude (OR=1.19, 95% CI 1.13 to 1.25,  $p=3.1\times 10^{-11}$ , figure 1). While we found some evidence for heterogeneity of effect size estimates across datasets ( $I^2=62$ , 95% CI 20 to 83,  $p$  Q statistic=0.0093; figure 1), all dataset-specific ORs suggested a risk effect for the C-allele, except for Central Spain. This indicates that heterogeneity was nearly entirely due to variance of effect size estimates at the same side of the null (OR<1).

Despite the compelling evidence now adding ANKRD55 to the list of established MS risk loci, the following limitations should be considered when interpreting our results. First, since determination of ethnic origin was based on self-report in the German, French and Spanish datasets, the possibility exists that results in these samples are affected by more subtle population substructure. However, appropriately adjusting for potential substructure effects in the GWAS datasets did not show any substantial change in results as compared with the unadjusted datasets. Hence, it is unlikely that population substructure has had a notable influence on our association results with rs6859219. Second, not all MS GWAS datasets published to date are publicly available and could therefore not all be included in the

current study. This applies to two datasets in particular ('Australia and New Zealand Multiple Sclerosis Genetics Consortium' and 'Brigham and Women's Hospital') included in a recent GWAS meta-analysis that reported an association between rs6859219 and MS at nominal significance.<sup>11</sup> However, combining the summary results reported in that study<sup>11</sup> with our data (while excluding the GeneMSA and IMSGC results calculated here) does not appreciably change our overall meta-analysis results (OR=1.16, 95% CI 1.11 to 1.22,  $p=2.9\times 10^{-10}$ ). Finally, the pathophysiological mechanisms underlying the association between rs6859219 in ANKRD55 and MS remain elusive. Rs6859219 is located in intron 7 of the gene, which is highly expressed in CD4 effector memory cells<sup>12</sup> but whose function remains largely unknown. As described previously, the linkage disequilibrium pattern (LD) around ANKRD55 is rather narrow,<sup>2, 5</sup> and rs6859219 shows noteworthy LD ( $r^2\geq 0.3$ ) only with other intronic ANKRD55 variants (based on 1000G CEU data; assessed with the SNAP software, <http://www.broadinstitute.org/mpg/snap/>). The LD block does not appear to extend to the two neighbouring genes (*IL6ST* encoding interleukin 6 signal transducer and *IL31RA* encoding interleukin 31 receptor A), which may be other immediate candidates interfering with the underlying autoimmune process in MS. Interestingly, recent in vitro ChIP-seq data generated by the

**Figure 1** Meta-analysis of datasets assessing the association between ANKRD55 rs6859219 and multiple sclerosis susceptibility in populations of European descent. Study-specific ORs (black squares) and 95% CIs (lines) were calculated using an additive model. The x-axis depicts the OR with regard to the risk allele dosage, that is, the C-allele. The summary ORs and 95% CIs (grey diamonds) were calculated based on fixed-effect meta-analysis combining all datasets as well as after stratification for the initial datasets and the validation datasets as indicated. GeneMSA, Genetic Multiple Sclerosis Associations<sup>10</sup>; GWAS, genome-wide association studies; IMSGC, International Multiple Sclerosis Genetics Consortium<sup>9</sup>; RAF, risk allele frequency in controls in the individual datasets.



international ENCODE project indicate that rs6859219, located ~182 kb 5' of the *IL6ST* transcription start site, lies within a target site for several transcription factors including activating enhancer binding protein 2 and early B cell factor 1<sup>13</sup> (accessible via <http://genome.ucsc.edu/cgi-bin/hgGateway>). Thus, functional genetic studies have to assess whether the association between rs6859219 and risk for MS and other autoimmune diseases is due to a dysfunction of *ANKRD55*, for example, via affecting mRNA splicing, or whether the effects may be caused by an altered transcriptional regulation of *IL6ST* and/or *IL31RA*.

#### Author affiliations

- <sup>1</sup>Neuropsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany
- <sup>2</sup>Department of Neurology, Focus Program Translational Neuroscience, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
- <sup>3</sup>Max Planck Institute for Human Development, Berlin, Germany
- <sup>4</sup>UPMC-INSERM-CNRS-UPMC-ICM, UMR 975–7225, Institut Cerveau Moelle Epinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France
- <sup>5</sup>Department of Human Genetics, Ruhr University, Bochum, Germany
- <sup>6</sup>Department of Neurology, University of Rostock, Rostock, Germany
- <sup>7</sup>Institute for Clinical Neuroimmunology, Ludwig Maximilian University, Munich, Germany
- <sup>8</sup>Department of Neurology, University of Würzburg, Würzburg, Germany
- <sup>9</sup>Centre for Research in Neuroscience at McGill University, Montreal, Canada
- <sup>10</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Département de Neurologie, Hôpital Pitié-Salpêtrière, Paris, France
- <sup>11</sup>Department of Neurology, Hôpital Sainte-Anne, Paris, France
- <sup>12</sup>Department of Neurology, Centre Hospitalier de Versailles, Le Chesnay, France
- <sup>13</sup>Department of Neurology, Centre Hospitalier Regional Universitaire, Tours, France
- <sup>14</sup>Área de Neurociencias, Inst. Investigación Sanitaria Biodonostia, San Sebastián, Spain
- <sup>15</sup>Servicio de Neurología, Hospital de Basurto, Bilbao, Spain
- <sup>16</sup>Department of Biología Celular e Inmunología, Instituto de Parasitología y Biomedicina 'López Neyra' (IPBLN), CSIC, Granada, Spain
- <sup>17</sup>Centre d'Esclerosi Múltiple de Catalunya, CEMCat, Unitat de Neuroimmunologia Clínica, Hospital Universitari Vall d'Hebron, Barcelona, Spain
- <sup>18</sup>Servicio de Neurología, Unidad de Esclerosi Múltiple, Hospital Donostia, San Sebastián, Spain
- <sup>19</sup>Department of Medicine, Rheumatology, and Clinical Immunology & DRFZ, Charité University Medicine, Berlin, Germany
- <sup>20</sup>Department of Psychology, Technische Universität Dresden, Dresden, Germany
- <sup>21</sup>Interdisciplinary Metabolic Center, Lipids Clinic, Charité University Medicine, Berlin, Germany
- <sup>22</sup>Department of Neurology, St. Josef-Hospital, Ruhr-University, Bochum, Germany
- <sup>23</sup>Department of Neurology, Sozialstiftung Bamberg Hospital, Bamberg, Germany
- <sup>24</sup>Department of Neurology, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany
- <sup>25</sup>Department of Clinical Chemistry, Ludwig Maximilian University, Munich, Germany
- <sup>26</sup>Institute of Human Genetics, University of Ulm, Ulm, Germany
- <sup>27</sup>Neurogenomik Laboratory, Department of Neuroscience, University of the Basque Country UPV/EHU, Leioa, Spain
- <sup>28</sup>KERBASQUE, Basque Foundation for Science, Bilbao, Spain

**Acknowledgements** We are grateful to the individuals participating in this study. We would like to thank investigators from the International Multiple Sclerosis Genetics Consortium and from the GeneMSA project for making their GWAS data available via the dbGaP platform. We acknowledge use of the cohort of the CRB-REFGENSEP and thank ICM, CIC Pitié-Salpêtrière, Génomique and REFGENSEP's members for their help and support.

**Contributors** Study design and supervision: CML, FZ, KV and LB. Data acquisition and performing of experiments: CML, B-MMS, CG, VD, DAA, PB, L-AG, AK, FL, IC-R, SH, AW, ET, FP, PC, DO, AAn, AAI, MC, XM, JO, FM, TD, S-CL, ES-T, UL, AC, PR, H-PH, OA, PL, MB, TK, CK, UKZ, JTE and BF. Data analysis and interpretation of results: CML, TL, KV and LB. Writing of manuscript: CML and LB with the help of all coauthors.

**Funding** This project was funded by grants from the German Ministry for Education and Research (BMBF) and German Research Foundation (DFG; to FZ), the BMBF and the Cure Alzheimer's Fund (to LB), the Walter- and Ilse-Rose-Stiftung (to H-PH and OA), the BMBF (grant NBL3 to UKZ; grant 01UW0808 to UL and ES-T), and the Innovation Fund of the Max Planck Society (M.FE.A.BILD0002 to UL). This project was supported by INSERM, ARSEP, AFM and GIS-IBISA. CML was supported by the Fidelity Biosciences Research Initiative.

**Competing interests** LA Gerdes reports to have received travel expenses and personal compensation from Merck Serono, Teva Pharmaceutical Industries, Bayer

Schering Pharma, Novartis and Biogen Idec. T Kl reports to have received travel expenses and personal compensations from Bayer Schering Pharmacy, Teva, Merck-Serono, Novartis, Sanofi-Aventis and Biogen-Idec as well as grant support from Bayer-Schering AG. None of the other authors reports any disclosures.

**Provenance and peer review** Not commissioned; externally peer reviewed.

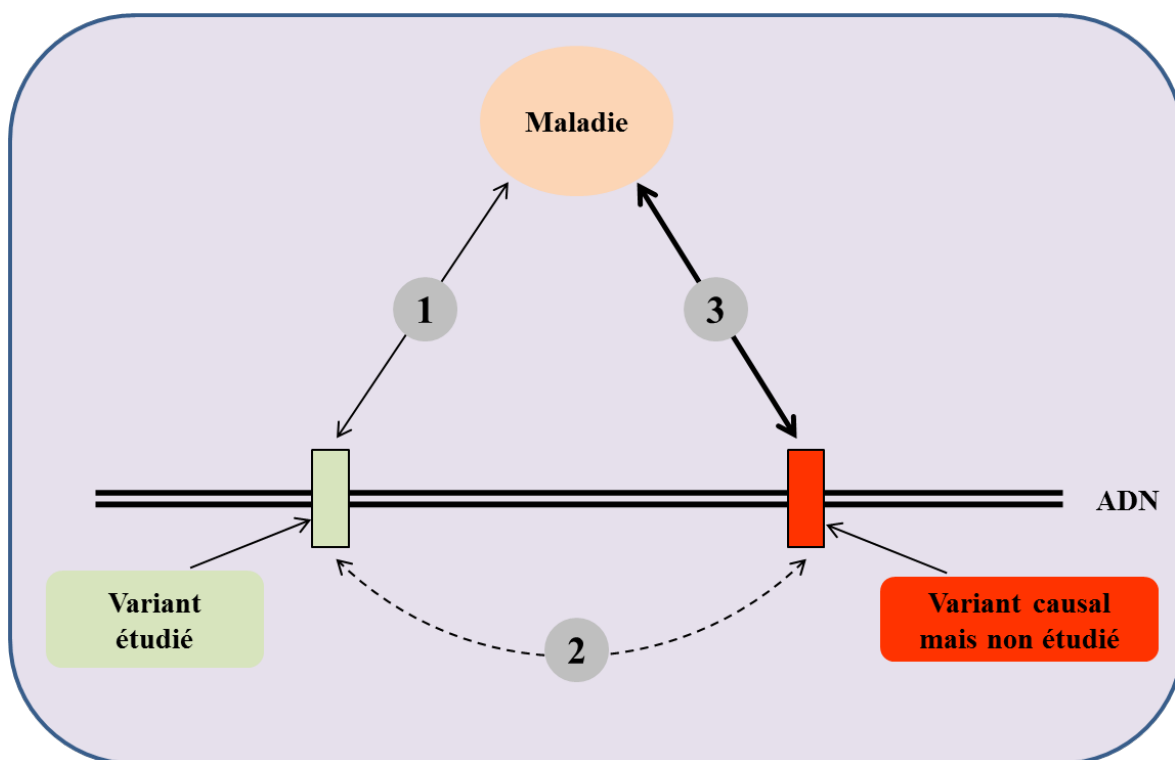
#### REFERENCES

- 1 Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitch KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ, FOCIS Network of Consortia. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011;7:e1002254.
- 2 Alloza I, Otaegui D, de Lapuente AL, Antigüedad A, Varadé J, Núñez C, Arroyo R, Urcelay E, Fernandez O, Leyva L, Fedetz M, Izquierdo G, Lucas M, Oliver-Martos B, Alcina A, Saiz A, Blanco Y, Comabella M, Montalbán X, Olascoaga J, Matesanz F, Vandenbroeck K. ANKRD55 and DHCR7 are novel multiple sclerosis risk loci. *Genes Immun* 2012;13:253–7.
- 3 Labauge P, Denier C, Bergametti F, Tournier-Lasserre E. Genetics of cavernous angiomas. *Lancet Neurol* 2007;6:237–44.
- 4 Mosavi LK, Cammett TJ, Desrosiers DC, Peng Z-Y. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* 2004;13:1435–48.
- 5 Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FA, Zhernakova A, Hinks A, Guiducci C, Chen R, Alfredsson L, Amos CI, Ardlie KG, Barton A, Bowes J, Brouwer E, Burt NP, Catanese JJ, Coblyn J, Coenen MJ, Costenbader KH, Criswell LA, Crusius JB, Cui J, de Bakker PI, De Jager PL, Ding B, Emery P, Flynn E, Harrison P, Hocking LJ, Huizinga TW, Kastner DL, Ke X, Lee AT, Liu X, Martin P, Morgan AW, Padyukov L, Posthumus MD, Radstake TR, Reid DM, Seielstad M, Seldin MF, Shadick NA, Steer S, Tak PP, Thomson W, van der Helm-van Mil AH, van der Horst-Bruinsma IE, van der Schoot CE, van Riel PL, Weinblatt ME, Wilson AG, Wolbink GJ, Wordsworth BP, Wijmenga C, Karlson EW, Toes RE, de Vries N, Begovich AB, Worthington J, Siminovitch KA, Gregersen PK, Klareskog L, Plenge RM, BIRAC Consortium, YEAR Consortium. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;42:508–14.
- 6 Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, Westra HJ, Fehrmann RS, Kurreeman FA, Thomson B, Gupta N, Romanos J, McManus R, Ryan AW, Turner G, Brouwer E, Posthumus MD, Remmers EF, Tucci F, Toes R, Grandone E, Mazzilli MC, Rybak A, Cukrowska B, Coenen MJ, Radstake TR, van Riel PL, Li Y, de Bakker PI, Gregersen PK, Worthington J, Siminovitch KA, Klareskog L, Huizinga TW, Wijmenga C, Plenge RM. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet* 2011;7:e1002004.
- 7 Lill CM, Schejide BM, Akkad DA, Blaschke P, Winkelmann A, Gerdes LA, Hoffjan S, Luessi F, Dörner T, Li SC, Steinhagen-Thiessen E, Lindenberger U, Chan A, Hartung HP, Aktas O, Lohse P, Kumpf T, Kubisch C, Epplen JT, Zettl UK, Bertram L, Zipp F. Independent replication of STAT3 association with multiple sclerosis risk in a large German case-control sample. *Neurogenetics* 2012;13:83–6.
- 8 Lill CM, Liu T, Schejide BM, Roehr JT, Akkad DA, Damotte V, Alcina A, Ortiz MA, Arroyo R, Lopez de Lapuente A, Blaschke P, Winkelmann A, Gerdes LA, Luessi F, Fernandez O, Izquierdo G, Antigüedad A, Hoffjan S, Courmu-Rebeix I, Gromöller S, Faber H, Liebsch M, Meissner E, Chanvillard C, Touze E, Pico F, Corcia P, Dörner T, Steinhagen-Thiessen E, Baeckman L, Heekeren HR, Li SC, Lindenberger U, Chan A, Hartung HP, Aktas O, Lohse P, Kumpf T, Kubisch C, Epplen JT, Zettl UK, Fontaine B, Vandenbroeck K, Matesanz F, Urcelay E, Bertram L, Zipp F, ANZgene Consortium. Closing the case of APOE in multiple sclerosis: no association with disease risk in over 29 000 subjects. *J Med Genet* 2012;49:558–62.
- 9 Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL, International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;357:851–62.
- 10 Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RL, Uitdehaag BM, Johnson MR, Angelakopoulou A, Hall L, Richardson JC, Prinjha RK, Gass A, Geurts JJ, Kragt J, Sombekke M, Vrenken H, Qualley P, Lincoln RR, Gomez R, Caillier SJ, George MF, Mousavi H, Guerrero R, Okuda DT, Cree BA, Green AJ, Waubant E, Goodin DS, Pelletier D, Matthews PM, Hauser SL, Kappos L, Polman CH, Oksenberg JR. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009;18:767–78.
- 11 Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, Sandbrink R, Pohl C, Edan G, Kappos L, Miller D, Montalbán J, Polman CH, Freedman MS, Hartung HP, Arnason BG, Comi G, Cook S, Filippi M, Goodin DS, Jeffery D, O'Connor P, Ebers GC, Langdon D, Reder AT, Traboulsee A, Zipp F, Schimrigk S, Hillert J, Bahl M, Booth DR, Broadley S, Brown MA, Browning BL, Browning SR, Butzkueven H, Carroll WM, Chapman C, Foote SJ, Griffiths L, Kermod AG, Kilpatrick TJ, Lechner-Scott J, Marriott M, Mason D, Moscato P, Heard RN, Pender MP, Perreau VM, Perera D, Rubio JP, Scott RJ, Slee M, Stankovich J, Stewart GJ,



- Taylor BV, Tubridy N, Willoughby E, Wiley J, Matthews P, Boneschi FM, Compston A, Haines J, Hauser SL, McCauley J, Ivinson A, Oksenberg JR, Pericak-Vance M, Sawcer SJ, De Jager PL, Hafler DA, de Bakker PI, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFN $\beta$ -1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011;70:897–912.
- 12 Hu X, Kim H, Stahl E, Plenge R, Daly M, Raychaudhuri S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* 2011;89:496–506.
- 13 Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altschuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elinitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altschuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Dutttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee BK, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhingee AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge EC, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry JS, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SC, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthraivadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanan E, Tress ML, van Baren MJ, Walters N, Washieti S, Wilming L, Zadissa A, Zhengdong Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Raymond A, Guigó R, Harrow J, Hubbard TJ, Landt SG, Frietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyenger S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Larnar-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenenbaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanavadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorson A, Yang X, Snyder M, White KP, Auer T, Centarin L, Eichenlaub M, Gruhl F, Heerman S, Hoekendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutayavin TM, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JA, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lusk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AO, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochoovsky L, Min R, Mu XJ, Rozowsky J, Yan KK, Yip KY, Birney E, ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.





**Figure 17. Association et variant causal.**

1. Le variant étudié montre une association avec la maladie.
2. Cependant, il peut ne pas être le variant causal. Il peut être lié génétiquement au variant causal, non génotypé dans l'étude.
3. Le variant causal, lorsqu'il est génotypé explique mieux l'association à la maladie et confère le risque de prédisposition maximal.

## 2.2. Hypothèse 2 : Les variants identifiés ne sont pas causaux

Les variants mis en évidence par les GWAS comme étant les variants de risque ne sont probablement pas les variants causaux. Les puces de génotypage utilisées dans les GWAS présentent une forte densité de SNPs (aujourd'hui, il existe des puces avec 5 millions de SNPs) et permettent ainsi de cartographier le génome plus ou moins finement. Cependant, dans la plupart des cas, elles ne permettent pas de distinguer un signal d'association dû à une association directe avec la maladie, d'un signal dû à une association indirecte, c'est à dire lorsque le SNP étudié est en fort déséquilibre de liaison avec le SNP causal (**Figure 17**) (Maher 2008, Manolio, Collins et al. 2009, Cortes and Brown 2011).

Dans le but d'identifier le variant causal localisé dans les régions des SNPs mis en évidence par les GWAS, l'analyse exhaustive de tous les variants fréquents, rares et non répertoriés dans les bases de données est indispensable. Il existe aujourd'hui deux moyens d'étude : le séquençage, qui a l'avantage d'être exhaustif et le *fine-mapping* (cartographie fine) par génotypage qui ne permet quant-à-lui que l'étude des SNPs déjà répertoriés dans les bases de données. Cependant, étant donné le nombre de régions à cartographier finement (58 dans le cas de la SEP) et la taille importante des cohortes nécessaires, le séquençage des régions d'intérêts se trouverait être onéreux. Le *fine-mapping* reposant sur du génotypage, le coût de l'étude serait majoritairement dû au prix de la puce de génotypage. Cependant, les puces de génotypage à façon, avec des SNPs d'intérêts pour une étude, sont généralement plus chères que les puces commerciales classiques.

Un des moyens permettant d'étudier finement ces régions par *fine-mapping* à un coût relativement faible est l'étude simultanée de plusieurs maladies ayant des caractéristiques physiopathologiques communes, avec par exemple, dans le cas de la SEP, d'autres maladies dites dys-immunes (Cortes and Brown 2011, Polychronakos 2011). En effet, beaucoup de régions d'intérêts sont communes à plusieurs maladies dys-immunes. Ainsi, près de la moitié des régions mises en évidence par le GWAS de l'IMSGC publié en 2011 avaient été aussi mises en évidence comme étant associées à une ou plusieurs autres maladies à composante dys-immune (**Tableau** pages 52-53) (Zhernakova, van Diemen et al. 2009, Cotsapas, Voight et al. 2011, IMSGC, WTCCC2 et al. 2011).

Dans le but d'étudier les régions communes d'intérêt aux maladies autoimmunes, un projet collaboratif a été initié entre plusieurs consortiums internationaux, aboutissant à la création de la puce « ImmunoChip ». Il s'agit d'une puce produite par Illumina, comportant



196.524 polymorphismes répartis dans 186 régions précédemment identifiées comme étant associées à une ou plusieurs maladies dys-immunes parmi les 12 suivantes : la SEP, le Diabète de type I, la maladie de Crohn, la Rectocolite Hémorragique, la Polyarthrite Rhumatoïde, la Spondylarthrite Ankylosante, la maladie Cœliaque, le Psoriasis, l'Hypothyroïdie, la Cirrhose Biliaire Primitive, le déficit en IgA et le Lupus Erythémateux Systémique. Les polymorphismes sélectionnés permettent de cartographier finement les régions d'intérêts avec une densité moyenne de 486 SNPs par régions (Cortes and Brown 2011, Polychronakos 2011, Trynka, Hunt et al. 2011). Les consortiums ayant garanti à Illumina une commande totale pour au minimum 150.000 individus, cette puce a pu être proposée aux équipes à un prix abordable.

Le projet ImmunoChip visait (a) à répliquer de manière définitive les résultats obtenus dans les GWAS antérieurs, (b) à identifier le ou les variants causaux au sein des régions identifiées par les GWAS, et (c) à étudier le rôle des variants rares présents dans ces mêmes régions. De plus, ce projet visait aussi à identifier des régions de prédisposition communes à ces maladies à composante dys-immunes.

Cependant, cette approche comporte quelques limites (Cortes and Brown 2011) :

- Beaucoup de SNPs rares ne sont pas encore répertoriés dans les bases de données et donc il n'est pas certain que les polymorphismes présents sur la puce permettent de capturer les effets de tous les SNPs rares.

- Le génotypage des SNPs rares présente quelques difficultés techniques, comme par exemple une mauvaise attribution de génotypes aux individus.

- Du fait de sa conception, la puce ImmunoChip ne couvre pas tout le génome. Les marqueurs sélectionnés dépendent en effet des résultats des GWAS antérieurs et par conséquent de leur puissance. Ainsi, si un signal n'avait pas été identifié dans les GWAS à cause d'un manque de puissance, il ne pourra pas non plus être identifié par cette puce ImmunoChip car il ne fera pas partie des régions cartographiées finement. Cela signifie que ce projet ne pourra pas résoudre le problème des « faux-négatifs », ces derniers n'étant pas présents sur la puce ImmunoChip car ils ne semblaient associés à aucune maladie à composante dys-immune.

**Le projet ImmunoChip a été effectué dans le cadre de l'IMSGC et a fait l'objet d'une publication, présentée ci-après. Il a d'une part permis d'étudier finement les**



**régions mises en évidence par le GWAS de l'IMSGC de 2011 (Hypothèse 4.2.) mais aussi d'identifier de nouvelles régions associées à la SEP grâce à une importante taille d'échantillons (Hypothèse 4.1) (*Publication IMSGC, Nat Genet 2013, présentée page 113*).**





## Article (co-auteur) :

### Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis

International Multiple Sclerosis Genetics Consortium, Beecham A, Patsopoulos N, Xifara D, Davis M, Kempainen A, Cotsapas C, Shah T, Spencer C, Booth D, Goris A, Oturai A, Saarela J, Fontaine B, Hemmer B, Martin C, Zipp F, D'Alfonso S, Martinelli-Boneschi F, Taylor B, Harbo H, Kockum I, Hillert J, Olsson T, Ban M, Oksenberg J, Agliardi C, Alfredsson L, Alizadeh M, Anderson C, Andrews R, Søndergaard H.B., Band G, Walton A, Baranzini S, Barizzone N, Barrett J, Bellenguez C, Bergamaschi L, Bernardinelli L, Berthele A, Biberacher V, Binder T, Blackburn H, Bomfim I, Brambilla P, Broadley S, Brochet B, Brundin L, Buck D, Butzkueven H, Caillier S, Camu W, Carpentier W, Cavalla P, Celius E, Coman I, Comi G, Corrado L, Cosemans L, Cournu-Rebeix I, Cree B, Cusi D, **Damotte V**, Defer G, Delgado S, Deloukas P, Di Sapio A, Dilthey A, Donnelly P, Dubois B, Duddy M, Edkins S, Elovaara I, Esposito F, Evangelou N, Fiddes B, Field J, Franke A, Freeman C, Galimberti D, Gieger C, Gourraud P-A, Graetz C, Graham A, Grummel V, Guaschino C, Hadjixenofontos A, Hakonarson H, Halfpenny C, Hall G, Hall P, Hamsten A, Harley J, Harrower T, Hawkins C, Hellenthal G, Hillier C, Hobart J, Hoshi M, Hunt S, Jagodic M, Jelčić I, Jochim A, Kendall B, Kermodé A, Kilpatrick T, Koivisto K, Konidari I, Korn T, Kronsbein H, Langford C, Larsson M, Lathrop M, Lebrun-Frenay C, Lechner-Scott J, Lee M, Leone M, Leppä V, Liberatore G, Lie B, Lill C, Lindén M, Link J, Luessi F, Lycke J, Macciardi F, Mannisto S, Manrique C, Martin R, Martinelli V, Mason D, Mazibrada G, McCabe C, Mero I-L, Moutsianas L, Myhr K-M, Nagels G, Nicholas R, Nilsson P, Piehl F, Pirinen M, Price S, Reunanen M, Robberecht W, Robertson N, Rodegher M, Rog D, Salvetti M, Schnetz-Boutaud N, Sellebjerg F, Selter R, Shaunak S, Shields S, Siffrin V, Slee M, Sorensen P.S., Sorosina M, Sospedra M, Spurkland A, Strange A, Sundqvist E, Thijs V, Thorpe J, Ticca A, Tienari P, Visser E, Vucic S, Westerlind H, Wiley J, Wilkins A, Wilson J, Winkelmann J, Frohlich I, Zajicek J, Zindler E, Ivinson A, Stewart G, Hafler D, Hauser S, Compston A, Haines J, Pericak-Vance M, De Jager P, McVean G, Sawcer S, McCauley J

*Nat Genet.* 2013 Sep 29. doi: 10.1038/ng.2770. [Epub ahead of print]



**Introduction** : Dans le contexte de la SEP, le projet « ImmunoChip » a été conduit par les équipes du consortium IMMSGC dans le but d'identifier de nouveaux variants de prédisposition à la maladie, de répliquer les associations potentielles et de rechercher les variants causaux au sein des régions d'intérêt par cartographie fine.

**Matériel et Méthodes** : Après contrôles qualités, 38 589 individus dont 14 498 patients SEP et 24 091 contrôles provenant de 11 pays différents (386 patients et 354 contrôles français provenant de la cohorte REFGENSEP) ont été typés par la technologie Illumina et inclus dans l'analyse.

Dans une première phase de screening à l'aide de la puce ImmunoChip, 135 signaux d'association inférieurs au seuil de  $p = 10^{-4}$  ont été sélectionnés pour réplification dans une cohorte indépendante de 14 802 patients et 26 703 contrôles provenant d'une méta-analyse de GWAS SEP publiés précédemment (meta v2.5 combinée (Patsopoulos, Esposito et al. 2011) au GWAS de l'IMMSGC de 2011 (IMMSGC, WTCCC2 et al. 2011)).

**Résultats** : Cette étude a permis d'identifier 97 SNPs associés à la SEP (93 primaires, c'est à dire ayant le seul ou le plus fort signal d'association significative dans la région identifiée et 4 secondaires, c'est à dire se trouvant dans la même région qu'un SNP primaire mais ayant un signal plus faible que le SNP primaire) correspondant à 48 nouveaux SNPs et 49 précédemment identifiés. De plus, 11 autres SNPs semblent montrer une association à la SEP mais ont besoin d'être confirmés car ils n'ont pas atteint le seuil de significativité de  $5 \times 10^{-8}$

De façon notable, pour les 4 régions d'intérêts des gènes *TNFSF14*, *IL2RA*, *TNFRSF1A* et *IL12A*, 50% de l'association avec la maladie s'explique par un unique variant.

**Conclusions** : Les résultats du projet « ImmunoChip » portent à 110 le nombre de marqueurs génétiques, hors *HLA*, associés à la prédisposition à la SEP. Il est estimé que ces 110 marqueurs n'expliquent que 12,8% du risque génétique, soit 23% en prenant en compte les SNPs de la région *HLA*.

Les 4 variants localisés dans les régions d'intérêts des gènes *TNFSF14*, *IL2RA*, *TNFRSF1A* et *IL12A* et expliquant plus de 50% de l'association de ces régions avec la maladie, pourraient correspondre au SNP causal de chacune de ces régions. Des études fonctionnelles sont maintenant nécessaires pour déterminer le rôle de ces variants dans le processus physiopathologique de la SEP.



**Ma contribution** : Ma participation à ce projet à consister à préparer et génotyper 500 patients et 500 contrôles français en utilisant la puce ImmunoChip. J'ai aussi été le représentant de l'équipe au cours des différentes conférences téléphoniques permettant de mettre en place et de suivre l'analyse statistique. J'ai participé à la relecture de l'article.



# Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis

International Multiple Sclerosis Genetics Consortium (IMSGC)\*

**Using the ImmunoChip custom genotyping array, we analyzed 14,498 subjects with multiple sclerosis and 24,091 healthy controls for 161,311 autosomal variants and identified 135 potentially associated regions ( $P < 1.0 \times 10^{-4}$ ). In a replication phase, we combined these data with previous genome-wide association study (GWAS) data from an independent 14,802 subjects with multiple sclerosis and 26,703 healthy controls. In these 80,094 individuals of European ancestry, we identified 48 new susceptibility variants ( $P < 5.0 \times 10^{-8}$ ), 3 of which we found after conditioning on previously identified variants. Thus, there are now 110 established multiple sclerosis risk variants at 103 discrete loci outside of the major histocompatibility complex. With high-resolution Bayesian fine mapping, we identified five regions where one variant accounted for more than 50% of the posterior probability of association. This study enhances the catalog of multiple sclerosis risk variants and illustrates the value of fine mapping in the resolution of GWAS signals.**

Multiple sclerosis (MIM 126200) is an inflammatory demyelinating disorder of the central nervous system that is a common cause of chronic neurological disability<sup>1,2</sup>. It shows greatest prevalence among individuals of northern European ancestry<sup>3</sup> and is moderately heritable<sup>4</sup>, with a sibling relative recurrence risk ( $\lambda_s$ ) of  $\sim 6.3$  (ref. 5). Aside from early success in demonstrating the important effects exerted by variants in the human leukocyte antigen (HLA) genes from the major histocompatibility complex (MHC)<sup>6</sup>, there was little progress in unraveling the genetic architecture underlying susceptibility to multiple sclerosis before the advent of GWAS technology. Over the last decade, our consortium has performed several GWAS and meta-analyses in large cohorts<sup>7–10</sup>, cumulatively identifying more than 50 non-MHC susceptibility alleles. As in other complex diseases, available data suggest that many additional susceptibility alleles remain to be identified<sup>11</sup>.

The striking overlap in the genetic architectures underlying susceptibility to multiple autoimmune diseases<sup>9,10,12,13</sup> prompted the collaborative construction of the ImmunoChip (see **Supplementary Figs. 1 and 2** and the **Supplementary Note** for details on IMSGC-nominated content), an efficient genotyping platform designed to deeply interrogate 184 non-MHC loci with genome-wide significant associations to at least 1 autoimmune disease and to provide lighter coverage of

other genomic regions with suggestive evidence of association<sup>14</sup>. Here we report a large-scale effort that leverages the ImmunoChip to detect association with multiple sclerosis susceptibility and to refine these associations via Bayesian fine mapping.

After performing stringent quality control, we report genotypes for 28,487 individuals of European ancestry (14,498 subjects with multiple sclerosis and 13,989 healthy controls) that are independent of the genotypes examined in previous GWAS efforts. We supplemented these data with genotypes from 10,102 independent control subjects provided by the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC)<sup>15</sup>, bringing the total number of subjects to 38,589 (14,498 subjects with multiple sclerosis and 24,091 healthy controls). We performed variant-level quality control, population outlier identification and subsequent case-control analysis in 11 country-organized strata. To account for within-stratum population stratification, we used the first five principal components as covariates in the association analysis. Per-stratum odds ratio (OR) and respective standard error (s.e.) estimates were then combined with an inverse variance meta-analysis under a fixed-effects model. In total, we tested 161,311 autosomal variants that passed quality control in at least 2 of the 11 strata (Online Methods). A Circos plot<sup>16</sup> summarizing the results from this discovery phase analysis is shown in **Figure 1**.

We defined an *a priori* significance threshold of  $P < 1 \times 10^{-4}$  for the discovery phase and identified 135 statistically independent primary association signals, comprising 67 in the designated fine-mapping regions and 68 in less densely covered regions selected for deep replication of earlier GWAS signals. Using forward stepwise logistic regression, we identified a second statistically independent signal in 13 of these regions (secondary signals) and a third statistically independent signal in 2 regions (tertiary signals). A total of 48 of the 150 statistically independent association signals (**Supplementary Table 1**) reached genome-wide significance of  $P < 5 \times 10^{-8}$  in the discovery phase alone. Next, we replicated our findings in 14,802 subjects with multiple sclerosis and 26,703 healthy controls with available GWAS data imputed using the 1000 Genomes Project European phase I (a) panel (Online Methods). Finally, we performed a joint analysis of the discovery and replication phases.

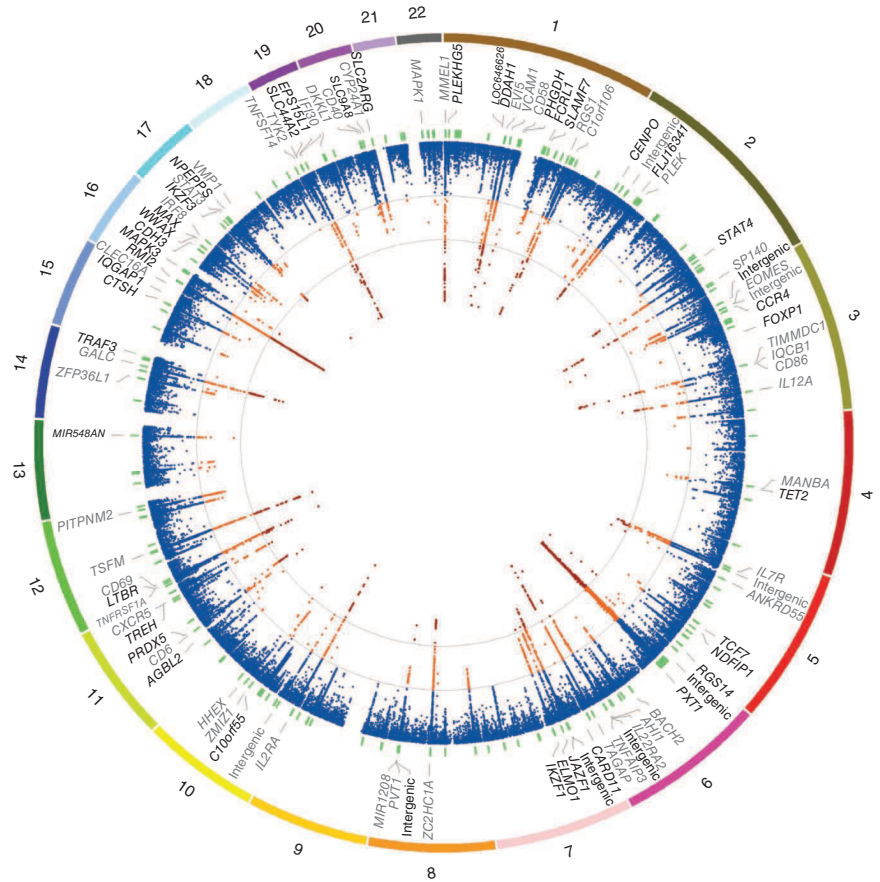
We identified 97 statistically independent SNPs meeting replication criteria (replication  $P < 0.05$ , joint  $P < 5 \times 10^{-8}$  and joint  $P < \text{discovery } P$ ), comprising 93 primary signals (**Supplementary Figs. 3–95**) and 4 secondary signals. Of these signals, 48 are new

\*Full lists of authors and affiliations appear at the end of the paper.

Received 24 April; accepted 3 September; published online 29 September 2013; doi:10.1038/ng.2770



**Figure 1** Discovery phase results. Circos plot showing primary association analysis of 161,311 autosomal variants in the discovery phase (14,498 cases and 24,091 healthy controls). The outermost track shows the numbered autosomal chromosomes. The second track indicates the gene closest to the most associated SNP meeting all replication criteria. Previously identified associations are indicated in gray. The third track indicates the physical position of the 184 fine-mapping intervals (green). The innermost track indicates  $-\log(P)$  (two-sided) for each SNP (scaled from 0–12, which truncates the signal in several regions; **Supplementary Table 1**). Additionally, contour lines are given at the *a priori* discovery ( $-\log(P) = 4$ ) and genome-wide significance ( $-\log(P) = 7.3$ ) thresholds. Orange indicates  $-\log(P) \geq 4$  and  $< 7.3$ , and red indicates  $-\log(P) \geq 7.3$ . Details of the full discovery phase results can be found in ImmunoBase (see URLs).



to multiple sclerosis (**Table 1**), and 49 correspond to previously identified susceptibility loci for multiple sclerosis (**Table 2**). An additional 11 independent SNPs showed suggestive evidence of association (joint  $P < 1 \times 10^{-6}$ ) (**Supplementary Table 2**).

The strongest newly associated SNP, rs12087340 (joint  $P = 1.1 \times 10^{-20}$ ; OR = 1.21), lies between *BCL10* (encoding B cell CLL/lymphoma 10) and *DDAH1* (encoding dimethylarginine dimethylaminohydrolase 1). The protein encoded by *BCL10* contains a caspase recruitment domain (CARD) and has been shown to activate nuclear factor (NF)- $\kappa$ B signaling<sup>17</sup>. This signaling molecule has an important role in controlling gene expression in inflammation, immunity, cell proliferation and apoptosis. It has been pursued as a potential therapeutic target for multiple sclerosis<sup>18</sup>. *BCL10* is also reported to interact with other CARD domain-containing proteins, including *CARD11* (ref. 19). We additionally identified a new association for rs1843938 (joint  $P = 1.2 \times 10^{-10}$ ; OR = 1.08), which is only 30 kb away from *CARD11*.

One newly associated SNP, rs2288904 (joint  $P = 1.6 \times 10^{-11}$ ; OR = 1.10), was found within an exon, representing a missense variant in *SLC44A2* (encoding solute carrier family 44, member 2). Notably, this variant is also reported as a monocyte-specific *cis*-acting expression quantitative trait locus (eQTL) for the antisense transcript of the nearby *ILF3* gene (encoding interleukin enhancer-binding factor 3)<sup>20</sup>. The *ILF3* protein was first discovered as a subunit of a nuclear factor found in activated T cells, which is required for T cell expression of *IL2*, an important molecule that regulates many aspects of inflammation.

Of the 49 previously identified risk loci<sup>9,10,21</sup>, 37 are in designated fine-mapping regions, and 23 of these 37 signals were localized to a single gene on the basis of genomic position (**Supplementary Table 3**). Although proximity does not necessarily indicate that these genes are functionally relevant, this observation nevertheless emphasizes the usefulness of dense mapping in localizing signals from a genome-wide screen. Our ImmunoChip analysis furthers the understanding of previously proposed secondary signals at three loci (**Supplementary Tables 4–6** and **Supplementary Note**); in particular, we showed that the effects of two previously proposed independent associations at the *IL2RA* locus<sup>7,22</sup> are driven by a single variant, rs2104286.

In an effort to define the functionally relevant variants underlying these associations, we further studied the regions surrounding the 97 associated SNPs using both Bayesian and frequentist approaches in 6,356 subjects with multiple sclerosis and 9,617 healthy controls from the UK (Online Methods)<sup>23</sup>. As determined by examining imputation quality, fine mapping was possible in 68 regions (**Supplementary Table 7**), including 66 of 93 primary (**Fig. 2a**) and 2 of 4 secondary signals. Eight of the 68 regions were fine mapped to high resolution (**Fig. 2b**, **Table 3** and **Supplementary Fig. 96**). One-third of the variants identified in these eight regions were imputed, indicating the value of imputation, even with dense genotyping coverage.

To assess whether functional annotation<sup>24</sup> provides clues about the molecular mechanisms associated with genetic risk, we considered the relationship of variants to annotated coding and regulatory features in these eight regions. Although we found no variants with missense or nonsense effects, there was a notable enrichment for variants with functional effects, including one variant known to affect splicing<sup>25</sup>, three variants known to correlate with RNA or serum protein levels<sup>22,26,27</sup> and several variants overlapping transcription factor binding sites and DNase I hypersensitive sites<sup>28,29</sup>. Four of the 18 variants in the fine-mapped regions are within conserved regions (GERP > 2)<sup>30</sup>. The overall lack of functional annotation likely reflects the limited repertoire of reference expression and epigenomic profiles and suggests that the function of the variants may be specific to cell type or cell state, as has been reported for many eQTLs in immune cell types<sup>20</sup>.

To determine the Gene Ontology (GO) processes of the 97 associated variants, we used MetaCore from Thomson Reuters (Online Methods). We found that the majority of the 97 variants lie within 50 kb of genes having immunological function. Of the 86 unique genes

**Table 1 Results for 48 new non-MHC variants associated with multiple sclerosis at a genome-wide significance level**

SNP	Chr.	Position <sup>a</sup>	RA	Discovery			Replication			Joint		Gene <sup>b</sup>	Function
				RAF	<i>P</i>	OR	RAF	<i>P</i>	OR	<i>P</i>	OR		
rs3007421	1	6530189	A	0.12	9.6 × 10 <sup>-7</sup>	1.12	0.13	8.8 × 10 <sup>-5</sup>	1.10	4.7 × 10 <sup>-10</sup>	1.11	<i>PLEKHG5</i>	Intronic
rs12087340	1	85746993	A	0.09	5.1 × 10 <sup>-12</sup>	1.22	0.09	2.9 × 10 <sup>-10</sup>	1.20	1.1 × 10 <sup>-20</sup>	1.21	<i>BCL10</i>	Intergenic
rs11587876	1	85915183	A	0.79	8.4 × 10 <sup>-8</sup>	1.12	0.81	2.9 × 10 <sup>-3</sup>	1.06	4.4 × 10 <sup>-9</sup>	1.09	<i>DDAH1</i>	Intronic
rs666930	1	120258970	G	0.53	7.5 × 10 <sup>-8</sup>	1.09	0.53	1.3 × 10 <sup>-5</sup>	1.07	6.0 × 10 <sup>-12</sup>	1.08	<i>PHGDH</i>	Intronic
rs2050568	1	157770241	G	0.53	1.3 × 10 <sup>-6</sup>	1.08	0.54	2.3 × 10 <sup>-5</sup>	1.07	1.5 × 10 <sup>-10</sup>	1.08	<i>FCRL1</i>	Intronic
rs35967351	1	160711804	A	0.67	1.7 × 10 <sup>-6</sup>	1.09	0.68	5.9 × 10 <sup>-6</sup>	1.09	4.4 × 10 <sup>-11</sup>	1.09	<i>SLAMF7</i>	Intronic
rs4665719	2	25017860	G	0.25	6.8 × 10 <sup>-6</sup>	1.09	0.25	1.1 × 10 <sup>-4</sup>	1.08	3.1 × 10 <sup>-9</sup>	1.08	<i>CENPO</i>	Intronic
rs842639	2	61095245	A	0.65	1.7 × 10 <sup>-9</sup>	1.11	0.67	1.4 × 10 <sup>-6</sup>	1.09	2.0 × 10 <sup>-14</sup>	1.10	<i>FLJ16341</i>	Noncoding RNA
rs9967792	2	191974435	G	0.62	1.8 × 10 <sup>-9</sup>	1.11	0.64	1.2 × 10 <sup>-4</sup>	1.07	3.5 × 10 <sup>-12</sup>	1.09	<i>STAT4</i>	Intronic
rs11719975	3	18785585	C	0.27	5.4 × 10 <sup>-6</sup>	1.09	0.28	4.1 × 10 <sup>-4</sup>	1.07	1.1 × 10 <sup>-8</sup>	1.08		Intergenic
rs4679081	3	33013483	G	0.52	1.2 × 10 <sup>-5</sup>	1.08	0.55	3.7 × 10 <sup>-4</sup>	1.07	2.2 × 10 <sup>-9</sup>	1.07	<i>CCR4</i>	Intergenic
rs9828629	3	71530346	G	0.62	5.5 × 10 <sup>-6</sup>	1.08	0.64	8.5 × 10 <sup>-6</sup>	1.08	1.9 × 10 <sup>-10</sup>	1.08	<i>FOX P1</i>	Intronic
rs2726518	4	106173199	C	0.55	1.2 × 10 <sup>-5</sup>	1.09	0.58	4.7 × 10 <sup>-4</sup>	1.06	3.9 × 10 <sup>-8</sup>	1.07	<i>TET2</i>	Intronic
rs756699	5	133446575	A	0.87	3.0 × 10 <sup>-6</sup>	1.12	0.88	6.5 × 10 <sup>-6</sup>	1.11	8.8 × 10 <sup>-11</sup>	1.12	<i>TCF7</i>	Intergenic
None <sup>c</sup>	5	141506564	C	0.61	6.0 × 10 <sup>-5</sup>	1.07	0.62	1.5 × 10 <sup>-5</sup>	1.08	3.6 × 10 <sup>-9</sup>	1.07	<i>NDFIP1</i>	Intronic
rs4976646	5	176788570	G	0.34	1.0 × 10 <sup>-12</sup>	1.13	0.36	5.0 × 10 <sup>-7</sup>	1.10	4.4 × 10 <sup>-18</sup>	1.12	<i>RGS14</i>	Intronic
rs17119	6	14719496	A	0.81	1.9 × 10 <sup>-6</sup>	1.11	0.80	1.2 × 10 <sup>-5</sup>	1.10	1.0 × 10 <sup>-10</sup>	1.10		Intergenic
rs941816	6	36375304	G	0.18	4.5 × 10 <sup>-9</sup>	1.13	0.20	8.3 × 10 <sup>-5</sup>	1.08	3.9 × 10 <sup>-12</sup>	1.11	<i>PXT1</i>	Intronic
rs1843938	7	3113034	A	0.44	2.2 × 10 <sup>-6</sup>	1.08	0.44	1.1 × 10 <sup>-5</sup>	1.08	1.2 × 10 <sup>-10</sup>	1.08	<i>CARD11</i>	Intergenic
rs706015	7	27014988	C	0.18	1.3 × 10 <sup>-9</sup>	1.14	0.18	9.9 × 10 <sup>-3</sup>	1.06	1.1 × 10 <sup>-9</sup>	1.10		Intergenic
rs917116	7	28172739	C	0.20	2.1 × 10 <sup>-8</sup>	1.12	0.21	5.8 × 10 <sup>-3</sup>	1.06	3.3 × 10 <sup>-9</sup>	1.09	<i>JAZF1</i>	Intronic
rs60600003	7	37382465	C	0.10	2.5 × 10 <sup>-8</sup>	1.16	0.10	4.2 × 10 <sup>-7</sup>	1.14	6.0 × 10 <sup>-14</sup>	1.15	<i>ELMO1</i>	Intronic
rs201847125 <sup>d</sup>	7	50325567	G	0.70	2.9 × 10 <sup>-8</sup>	1.11	0.70	6.7 × 10 <sup>-5</sup>	1.09	1.2 × 10 <sup>-11</sup>	1.10	<i>IKZF1</i>	Intergenic
rs2456449	8	128192981	G	0.36	2.2 × 10 <sup>-8</sup>	1.10	0.37	3.8 × 10 <sup>-3</sup>	1.05	1.8 × 10 <sup>-9</sup>	1.08		Intergenic
rs793108	10	31415106	A	0.50	5.6 × 10 <sup>-8</sup>	1.09	0.51	1.8 × 10 <sup>-5</sup>	1.07	6.1 × 10 <sup>-12</sup>	1.08		Intergenic
rs2688608	10	75658349	A	0.55	6.4 × 10 <sup>-5</sup>	1.07	0.56	2.0 × 10 <sup>-4</sup>	1.06	4.6 × 10 <sup>-8</sup>	1.07	<i>C10orf55</i>	Intergenic
rs7120737	11	47702395	G	0.15	7.6 × 10 <sup>-8</sup>	1.13	0.15	1.0 × 10 <sup>-3</sup>	1.08	1.0 × 10 <sup>-9</sup>	1.10	<i>AGBL2</i>	Intronic
rs694739	11	64097233	A	0.62	1.3 × 10 <sup>-5</sup>	1.08	0.62	3.8 × 10 <sup>-5</sup>	1.07	2.0 × 10 <sup>-9</sup>	1.07	<i>PRDX5</i>	Intergenic
rs9736016	11	118724894	T	0.63	2.2 × 10 <sup>-8</sup>	1.10	0.63	2.6 × 10 <sup>-8</sup>	1.10	3.0 × 10 <sup>-15</sup>	1.10	<i>CXCR5</i>	Intergenic
rs12296430	12	6503500	C	0.19	3.6 × 10 <sup>-10</sup>	1.14	0.21	1.7 × 10 <sup>-5</sup>	1.09	7.2 × 10 <sup>-14</sup>	1.12	<i>LTBR</i>	Intergenic
rs4772201	13	100086259	A	0.82	1.7 × 10 <sup>-7</sup>	1.12	0.83	1.1 × 10 <sup>-4</sup>	1.09	1.3 × 10 <sup>-10</sup>	1.10	<i>MIR548AN</i>	Intergenic
rs12148050	14	103263788	A	0.35	1.5 × 10 <sup>-5</sup>	1.08	0.36	4.3 × 10 <sup>-9</sup>	1.10	5.1 × 10 <sup>-13</sup>	1.09	<i>TRAF3</i>	Intronic
rs59772922	15	79207466	A	0.83	4.0 × 10 <sup>-6</sup>	1.11	0.83	5.4 × 10 <sup>-4</sup>	1.08	1.2 × 10 <sup>-8</sup>	1.09	<i>CTSH</i>	Intergenic
rs8042861	15	90977333	A	0.44	9.8 × 10 <sup>-7</sup>	1.08	0.45	3.4 × 10 <sup>-4</sup>	1.06	2.2 × 10 <sup>-9</sup>	1.07	<i>IQGAP1</i>	Intronic
rs6498184	16	11435990	G	0.81	2.1 × 10 <sup>-10</sup>	1.15	0.82	6.5 × 10 <sup>-9</sup>	1.14	7.4 × 10 <sup>-18</sup>	1.15	<i>RMI2</i>	Intergenic
rs7204270 <sup>e</sup>	16	30156963	G	0.50	9.3 × 10 <sup>-8</sup>	1.09	0.49	3.7 × 10 <sup>-5</sup>	1.08	1.6 × 10 <sup>-11</sup>	1.09	<i>MAPK3</i>	Intergenic
rs1886700	16	68685905	A	0.14	8.8 × 10 <sup>-6</sup>	1.11	0.14	3.2 × 10 <sup>-4</sup>	1.08	1.3 × 10 <sup>-8</sup>	1.10	<i>CDH3</i>	Intronic
rs12149527	16	79110596	A	0.47	1.7 × 10 <sup>-6</sup>	1.08	0.47	4.3 × 10 <sup>-6</sup>	1.08	3.3 × 10 <sup>-11</sup>	1.08	<i>WWOX</i>	Intronic
rs7196953	16	79649394	A	0.29	2.6 × 10 <sup>-5</sup>	1.08	0.30	7.9 × 10 <sup>-7</sup>	1.09	1.0 × 10 <sup>-10</sup>	1.09	<i>MAF</i>	Intergenic
rs12946510	17	37912377	A	0.47	8.5 × 10 <sup>-6</sup>	1.08	0.48	8.0 × 10 <sup>-5</sup>	1.07	2.9 × 10 <sup>-9</sup>	1.07	<i>IKZF3</i>	Intergenic
rs4794058	17	45597098	A	0.50	1.6 × 10 <sup>-5</sup>	1.07	0.52	3.5 × 10 <sup>-10</sup>	1.11	1.0 × 10 <sup>-13</sup>	1.09	<i>NPEPPS</i>	Intergenic
rs2288904	19	10742170	G	0.77	9.6 × 10 <sup>-10</sup>	1.14	0.78	5.4 × 10 <sup>-4</sup>	1.07	1.6 × 10 <sup>-11</sup>	1.10	<i>SLC44A2</i>	Exonic
rs1870071	19	16505106	G	0.29	5.7 × 10 <sup>-10</sup>	1.12	0.30	4.6 × 10 <sup>-7</sup>	1.09	2.0 × 10 <sup>-15</sup>	1.10	<i>EPS15L1</i>	Intronic
rs17785991	20	48438761	A	0.35	6.4 × 10 <sup>-7</sup>	1.09	0.34	5.9 × 10 <sup>-3</sup>	1.05	4.2 × 10 <sup>-8</sup>	1.07	<i>SLC9A8</i>	Intronic
rs2256814	20	62373983	A	0.19	8.3 × 10 <sup>-7</sup>	1.11	0.21	6.4 × 10 <sup>-4</sup>	1.08	3.5 × 10 <sup>-9</sup>	1.09	<i>SLC2A4RG</i>	Intronic
<b>Secondary</b>													
rs7769192 <sup>f</sup>	6	137962655	G	0.55	1.3 × 10 <sup>-5</sup>	1.08	0.54	5.1 × 10 <sup>-5</sup>	1.07	3.3 × 10 <sup>-9</sup>	1.08		Intergenic
rs533646 <sup>g</sup>	11	118566746	G	0.68	3.6 × 10 <sup>-7</sup>	1.10	0.68	3.9 × 10 <sup>-5</sup>	1.08	7.6 × 10 <sup>-11</sup>	1.09	<i>TREH</i>	Intergenic
rs4780346 <sup>h</sup>	16	11288806	A	0.23	6.8 × 10 <sup>-6</sup>	1.09	0.25	1.5 × 10 <sup>-5</sup>	1.09	4.4 × 10 <sup>-10</sup>	1.09	<i>CLEC16A</i>	Intergenic

All listed signals had discovery  $P \leq 1.0 \times 10^{-4}$ , replication  $P \leq 5.0 \times 10^{-2}$  and joint  $P \leq 5.0 \times 10^{-8}$ . All *P* values are two-sided. Chr., chromosome; RA, risk allele; RAF, risk allele frequency.

<sup>a</sup>Position is based on hg19 and dbSNP Build 137. <sup>b</sup>The nearest gene is listed if it lies within 50 kb of the signal. Bold font indicates genes that are part of the GO immune system process. <sup>c</sup>Proxy SNP rs1036207 ( $r^2 = 0.99$ ) was used in replication. <sup>d</sup>Proxy SNP rs1716719 ( $r^2 = 1.00$ ) was used in replication. <sup>e</sup>The primary SNP was rs11865086 ( $P = 1.77 \times 10^{-8}$ ) in the discovery phase, but this SNP (or a good proxy) was not available in replication, so the next best discovery SNP was used. <sup>f</sup>*P* values and OR values shown are after conditioning on rs67297943 (Table 2). <sup>g</sup>*P* values and OR values shown are after conditioning on rs9736016 (Table 2). <sup>h</sup>*P* values and OR values shown are after conditioning on rs12927355 (Table 2).

represented, 35 are linked to the GO immune system process (Tables 1 and 2). We did not see substantial over- or under-representation of certain GO processes when comparing the newly discovered and previously identified loci, but this may represent a limitation of the ImmunoChip in targeting genomic loci enriched for immunologically active genes, as more subtle distinctions between such loci might not be adequately captured by broad annotations such as GO processes.

Finally, we explored the overlap between our findings and those in other autoimmune diseases with reported ImmunoChip analyses. We calculated the percentage of multiple sclerosis signals (110 non-MHC; Supplementary Table 8) overlapping with those of other autoimmune diseases by requiring  $r^2 \geq 0.8$  between the best variants reported in each study using SNAP<sup>31</sup>. In total, we found that ~22% of our signals overlapped at least one other autoimmune disease signal.

Table 2 Results for 49 known non-MHC variants associated with multiple sclerosis at genome-wide significance

SNP	Chr.	Position <sup>a</sup>	RA	Discovery			Replication			Joint		Gene <sup>b</sup>	Function
				RAF	<i>P</i>	OR	RAF	<i>P</i>	OR	<i>P</i>	OR		
rs3748817	1	2525665	A	0.64	$1.3 \times 10^{-12}$	1.14	0.65	$1.2 \times 10^{-15}$	1.15	$1.3 \times 10^{-26}$	1.14	<i>MMEL1</i>	Intronic
rs41286801	1	92975464	A	0.14	$7.9 \times 10^{-16}$	1.20	0.16	$2.1 \times 10^{-12}$	1.17	$1.4 \times 10^{-26}$	1.19	<i>EVI5</i>	3' UTR
rs7552544 <sup>c</sup>	1	101240893	A	0.56	$3.7 \times 10^{-6}$	1.08	0.43	$3.3 \times 10^{-12}$	1.12	$1.9 \times 10^{-16}$	1.10	<i>VCAM1</i>	Intergenic
rs6677309	1	117080166	A	0.88	$1.5 \times 10^{-28}$	1.34	0.88	$4.1 \times 10^{-16}$	1.24	$5.4 \times 10^{-42}$	1.29	<i>CD58</i>	Intronic
rs1359062	1	192541472	C	0.82	$1.8 \times 10^{-13}$	1.18	0.83	$2.1 \times 10^{-8}$	1.13	$4.8 \times 10^{-20}$	1.15	<i>RGS1</i>	Intergenic
rs55838263	1	200874728	A	0.71	$1.4 \times 10^{-9}$	1.12	0.71	$3.9 \times 10^{-11}$	1.13	$4.0 \times 10^{-19}$	1.13	<i>C1orf106</i>	Intronic
rs2163226	2	43361256	A	0.71	$7.0 \times 10^{-8}$	1.10	0.73	$3.8 \times 10^{-10}$	1.14	$2.1 \times 10^{-16}$	1.12		Intergenic
rs7595717	2	68587477	A	0.26	$3.3 \times 10^{-7}$	1.10	0.27	$6.8 \times 10^{-8}$	1.10	$1.2 \times 10^{-13}$	1.10	<i>PLEK</i>	Intergenic
rs9989735	2	231115454	C	0.18	$7.8 \times 10^{-14}$	1.17	0.19	$6.8 \times 10^{-11}$	1.14	$4.2 \times 10^{-23}$	1.16	<i>SP140</i>	Intronic
rs2371108	3	27757018	A	0.38	$2.1 \times 10^{-6}$	1.08	0.39	$5.8 \times 10^{-11}$	1.12	$1.5 \times 10^{-15}$	1.10	<i>EDMES</i>	downstream
rs1813375	3	28078571	A	0.47	$5.7 \times 10^{-18}$	1.15	0.49	$4.4 \times 10^{-16}$	1.15	$1.9 \times 10^{-32}$	1.15		Intergenic
rs1131265	3	119222456	C	0.80	$2.0 \times 10^{-15}$	1.19	0.81	$4.8 \times 10^{-10}$	1.14	$1.4 \times 10^{-23}$	1.17	<i>TIMMDC1</i>	Exonic
rs1920296 <sup>c</sup>	3	121543577	C	0.64	$6.8 \times 10^{-15}$	1.14	0.64	$5.5 \times 10^{-9}$	1.10	$6.5 \times 10^{-22}$	1.12	<i>IQCB1</i>	Intronic
rs2255214 <sup>c</sup>	3	121770539	C	0.52	$5.3 \times 10^{-13}$	1.13	0.52	$3.3 \times 10^{-13}$	1.13	$1.2 \times 10^{-24}$	1.13	<i>CD86</i>	Intergenic
rs1014486	3	159691112	G	0.43	$1.2 \times 10^{-9}$	1.11	0.44	$1.4 \times 10^{-10}$	1.11	$1.1 \times 10^{-18}$	1.11	<i>IL12A</i>	Intergenic
rs7665090	4	103551603	G	0.52	$2.4 \times 10^{-6}$	1.08	0.53	$5.0 \times 10^{-4}$	1.13	$1.0 \times 10^{-8}$	1.09	<i>MANBA</i>	Intergenic
rs6881706	5	35879156	C	0.72	$4.9 \times 10^{-9}$	1.12	0.73	$1.7 \times 10^{-9}$	1.12	$4.3 \times 10^{-17}$	1.12	<i>IL7R</i>	Intergenic
rs6880778	5	40399096	G	0.60	$1.7 \times 10^{-8}$	1.10	0.61	$3.9 \times 10^{-13}$	1.13	$8.1 \times 10^{-20}$	1.12		Intergenic
rs71624119	5	55440730	G	0.76	$2.7 \times 10^{-9}$	1.12	0.76	$1.9 \times 10^{-5}$	1.09	$3.4 \times 10^{-13}$	1.11	<i>ANKRD55</i>	Intronic
rs72928038	6	90976768	A	0.17	$7.6 \times 10^{-7}$	1.11	0.19	$9.0 \times 10^{-11}$	1.17	$1.5 \times 10^{-15}$	1.14	<i>BACH2</i>	Intronic
rs11154801	6	135739355	A	0.37	$2.3 \times 10^{-9}$	1.11	0.37	$1.0 \times 10^{-12}$	1.13	$1.8 \times 10^{-20}$	1.12	<i>AHI1</i>	Intronic
rs17066096	6	137452908	G	0.23	$5.9 \times 10^{-12}$	1.14	0.25	$4.1 \times 10^{-13}$	1.15	$1.6 \times 10^{-23}$	1.14	<i>IL22RA2</i>	Intergenic
rs67297943	6	138244816	A	0.78	$4.8 \times 10^{-8}$	1.12	0.80	$2.5 \times 10^{-6}$	1.11	$5.5 \times 10^{-13}$	1.11	<i>TNFAIP3</i>	Intergenic
rs212405	6	159470559	T	0.62	$1.4 \times 10^{-15}$	1.15	0.64	$1.8 \times 10^{-7}$	1.10	$8.0 \times 10^{-21}$	1.12	<i>TAGAP</i>	Intergenic
rs1021156	8	79575804	A	0.24	$5.6 \times 10^{-10}$	1.12	0.26	$2.1 \times 10^{-8}$	1.11	$8.5 \times 10^{-17}$	1.11	<i>ZC2HC1A</i>	Intergenic
rs4410871	8	128815029	G	0.72	$2.0 \times 10^{-9}$	1.12	0.72	$3.4 \times 10^{-8}$	1.11	$4.3 \times 10^{-16}$	1.11	<i>MIR1204</i>	Intergenic
rs759648	8	129158945	C	0.31	$2.8 \times 10^{-6}$	1.09	0.31	$3.7 \times 10^{-5}$	1.08	$5.0 \times 10^{-10}$	1.08	<i>MIR1208</i>	Intergenic
rs2104286	10	6099045	A	0.72	$7.6 \times 10^{-23}$	1.21	0.73	$3.6 \times 10^{-26}$	1.23	$2.3 \times 10^{-47}$	1.22	<i>IL2RA</i>	Intronic
rs1782645	10	81048611	A	0.43	$4.3 \times 10^{-7}$	1.09	0.41	$6.2 \times 10^{-10}$	1.11	$2.5 \times 10^{-15}$	1.10	<i>ZMIZ1</i>	Intronic
rs7923837	10	94481917	G	0.61	$4.6 \times 10^{-9}$	1.11	0.62	$2.0 \times 10^{-9}$	1.11	$4.3 \times 10^{-17}$	1.11	<i>HHEX</i>	Intergenic
rs34383631	11	60793330	A	0.40	$5.7 \times 10^{-10}$	1.11	0.39	$4.5 \times 10^{-15}$	1.15	$3.7 \times 10^{-23}$	1.13	<i>CD6</i>	Intergenic
rs1800693	12	6440009	G	0.40	$6.9 \times 10^{-16}$	1.14	0.41	$1.0 \times 10^{-13}$	1.14	$6.7 \times 10^{-28}$	1.14	<i>TNFRSF1A</i>	Intronic
rs11052877	12	9905690	G	0.36	$5.4 \times 10^{-9}$	1.10	0.38	$1.2 \times 10^{-5}$	1.08	$5.6 \times 10^{-13}$	1.09	<i>CD69</i>	3' UTR
rs201202118 <sup>d</sup>	12	58182062	A	0.67	$7.4 \times 10^{-13}$	1.14	0.67	$1.6 \times 10^{-10}$	1.12	$9.0 \times 10^{-22}$	1.13	<i>TSFM</i>	Intronic
rs7132277	12	123593382	A	0.19	$1.9 \times 10^{-6}$	1.10	0.19	$1.4 \times 10^{-8}$	1.13	$1.9 \times 10^{-13}$	1.12	<i>PITPNM2</i>	Intronic
rs2236262	14	69261472	A	0.50	$1.2 \times 10^{-5}$	1.08	0.50	$3.8 \times 10^{-8}$	1.09	$2.5 \times 10^{-12}$	1.08	<i>ZFP36L1</i>	Intronic
rs74796499	14	88432328	C	0.95	$8.5 \times 10^{-11}$	1.31	0.95	$4.5 \times 10^{-11}$	1.33	$2.4 \times 10^{-20}$	1.32	<i>GALC</i>	Intronic
rs12927355	16	11194771	G	0.68	$8.2 \times 10^{-27}$	1.21	0.69	$4.3 \times 10^{-21}$	1.18	$6.4 \times 10^{-46}$	1.20	<i>CLEC16A</i>	Intronic
rs35929052	16	85994484	G	0.89	$3.3 \times 10^{-7}$	1.14	0.88	$3.6 \times 10^{-6}$	1.15	$5.9 \times 10^{-12}$	1.15	<i>IRF8</i>	Intergenic
rs4796791	17	40530763	A	0.36	$1.8 \times 10^{-8}$	1.10	0.36	$1.2 \times 10^{-13}$	1.14	$3.7 \times 10^{-20}$	1.12	<i>STAT3</i>	Intronic
rs8070345	17	57816757	A	0.45	$5.4 \times 10^{-16}$	1.14	0.46	$1.9 \times 10^{-9}$	1.10	$2.2 \times 10^{-23}$	1.12	<i>VMP1</i>	Intronic
rs1077667	19	6668972	G	0.79	$3.5 \times 10^{-13}$	1.16	0.79	$8.4 \times 10^{-13}$	1.16	$1.7 \times 10^{-24}$	1.16	<i>TNFSF14</i>	Intronic
rs34536443	19	10463118	C	0.95	$1.2 \times 10^{-8}$	1.28	0.96	$2.9 \times 10^{-7}$	1.30	$1.8 \times 10^{-14}$	1.29	<i>TYK2</i>	Exonic
rs11554159	19	18285944	G	0.73	$2.6 \times 10^{-13}$	1.15	0.74	$1.4 \times 10^{-12}$	1.15	$1.9 \times 10^{-24}$	1.15	<i>IFI30</i>	Exonic
rs8107548	19	49870643	G	0.25	$2.0 \times 10^{-6}$	1.09	0.26	$2.5 \times 10^{-10}$	1.13	$5.7 \times 10^{-15}$	1.11	<i>DKKL1</i>	Intronic
rs4810485	20	44747947	A	0.25	$1.8 \times 10^{-5}$	1.08	0.25	$1.4 \times 10^{-12}$	1.14	$7.7 \times 10^{-16}$	1.11	<i>CD40</i>	Intronic
rs2248359	20	52791518	G	0.60	$9.8 \times 10^{-5}$	1.07	0.62	$8.2 \times 10^{-11}$	1.12	$2.0 \times 10^{-13}$	1.09	<i>CYP24A1</i>	Intergenic
rs2283792	22	22131125	C	0.51	$1.1 \times 10^{-6}$	1.08	0.53	$5.4 \times 10^{-11}$	1.11	$5.5 \times 10^{-16}$	1.10	<i>MAPK1</i>	Intronic
<b>Secondary</b>													
rs523604 <sup>e</sup>	11	118755738	A	0.53	$2.5 \times 10^{-7}$	1.09	0.54	$4.0 \times 10^{-9}$	1.11	$6.2 \times 10^{-15}$	1.10	<i>CXCR5</i>	Intronic

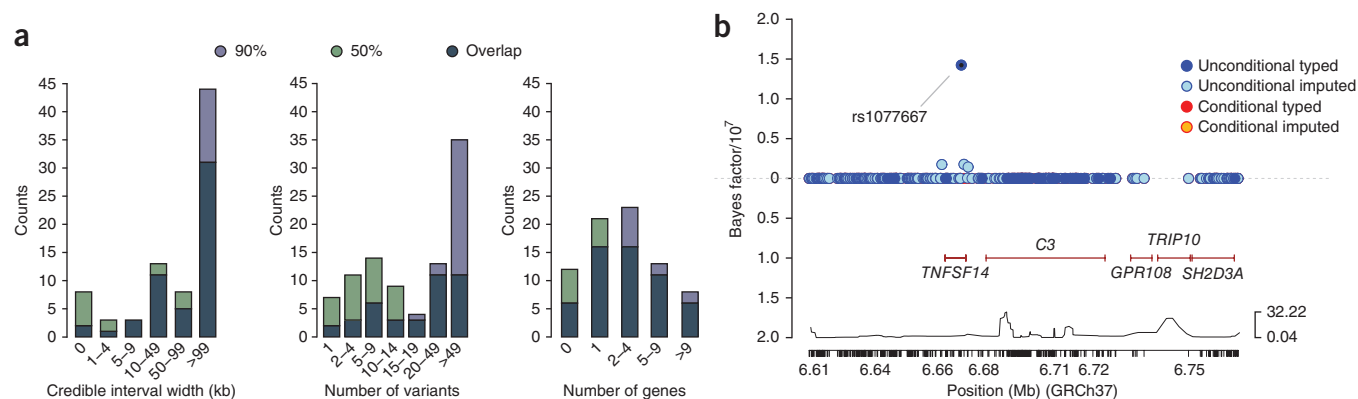
All listed signals had discovery  $P \leq 1.0 \times 10^{-4}$ , replication  $P \leq 5.0 \times 10^{-2}$  and joint  $P \leq 5.0 \times 10^{-8}$ . All *P* values are two-sided. Chr., chromosome; RA, risk allele; RAF, risk allele frequency.

<sup>a</sup>Position is based on hg19 and dbSNP Build 137. <sup>b</sup>The nearest gene is listed if it lies within 50 kb of the signal. Bold font indicates genes that are part of GO immune system process. <sup>c</sup>These three SNPs were not primary in the 2011 GWAS<sup>9</sup> (two were secondary, and the third was tertiary in that study). <sup>d</sup>A proxy SNP (rs10431552;  $r^2 = 0.99$ ) was used in replication. <sup>e</sup>The *P* values and OR values shown are after conditioning on rs9736016 and rs533646 (Table 1).

More specifically, ~9.1% overlapped signals in inflammatory bowel disease, ~7.3% overlapped signals in ulcerative colitis, ~9.1% overlapped signals in Crohn's disease<sup>15</sup>, ~9.1% overlapped signals in primary biliary cirrhosis<sup>32,33</sup>, ~4.5% overlapped signals in celiac disease<sup>34</sup>, ~4.5% overlapped signals in rheumatoid arthritis<sup>35</sup>, ~0.9% overlapped signals in psoriasis<sup>36</sup>, and ~2.7% overlapped signals in autoimmune thyroid disease<sup>37</sup>. We report at seven loci the same top variant seen in primary biliary cirrhosis. We also note that our best

*TYK2* variant (rs34536443)<sup>38</sup> is also the most associated *TYK2* variant for primary biliary cirrhosis, psoriasis and rheumatoid arthritis. Lastly, variants have been reported for autoimmune thyroid disease, celiac disease, primary biliary cirrhosis and rheumatoid arthritis that have pairwise  $r^2 \geq 0.8$  with the multiple sclerosis risk variant near *MMEL1* (Supplementary Table 8)<sup>39</sup>.

In summary, we have identified 48 new susceptibility variants for multiple sclerosis. These newly discovered loci expand our understanding



**Figure 2** Bayesian fine mapping within primary regions of association. (a) Summary of the extent of fine mapping across 66 regions in 6,356 subjects with multiple sclerosis and 9,617 healthy controls from the UK, showing the physical extent of the number of variants and the number of genes spanned by the posterior 90% and 50% credible sets. (b) Details of fine mapping in the region encompassing *TNFSF14*. Bayes factors summarizing evidence for association of the SNPs before conditioning on the lead SNP (rs1077667) are shown above the x axis (blue markers), and the Bayes factors after conditioning are shown below the x axis.

**Table 3 Results for 18 variants from the 8 regions with consistent high-resolution fine mapping**

Gene	SNP	Chr.	Position <sup>a</sup>	Posterior	GERP	Functional annotation <sup>b</sup>
<i>IL2RA</i>	rs2104286	10	6099045	0.93	-0.47	Intronic, correlates with soluble IL-2RA levels
<i>TNFSF14</i>	rs1077667	19	6668972	0.74	-3.89	Intronic, TFBS or DNase I peak, correlates with serum levels of <i>TNFSF14</i>
<i>TNFRSF1A</i>	rs1800693	12	6440009	0.69	2.53	Intronic, causes splicing defect and truncated soluble <i>TNFRSF1A</i>
	rs4149580 <sup>c</sup>	12	6446990	0.10	2.06	intronic
<i>IL12A</i>	rs1014486	3	159691112	0.67	0.24	-
<i>STAT4</i>	rs78712823	2	191958581	0.59	-3.98	Intronic
<i>TNFAIP3</i>	rs632574	6	137959118	0.27	-1.15	-
	rs498549 <sup>c</sup>	6	137984935	0.20	0.52	-
	rs651973	6	137996134	0.17	2.41	Downstream of RP11-95M15.1 lincRNA gene
	rs536331	6	137993049	0.15	0.19	Upstream of RP11-95M15.1 lincRNA gene
<i>CD58</i>	rs35275493 <sup>c</sup>	1	117095502	0.24	0.75	Intronic (insertion)
	rs10754324 <sup>c</sup>	1	117093035	0.22	0.32	Intronic
	rs6677309	1	117080166	0.21	-1.18	Intronic, TFBS or DNase I peak
	rs1335532	1	117100957	0.17	-1.32	Intronic
<i>CD6</i>	rs34383631	11	60793330	0.20	1.66	-
	rs4939490 <sup>c</sup>	11	60793651	0.14	-0.53	-
	rs4939491 <sup>c</sup>	11	60793722	0.14	-0.37	-
	rs4939489	11	60793648	0.10	3.25	-

All listed variants have posterior probability  $\geq 0.1$  in regions where  $\leq 5$  variants explain the top 50% of the posterior probability, and the top SNP from the frequentist analysis resides in the 90% confidence interval; SNPs are ordered by maximum posterior probability within each region. Posterior probability is the probability of any variant driving association<sup>23</sup>. GERP denotes Genomic Evolutionary Rate Profiling. Chr., chromosome; TFBS, transcription factor binding site; lincRNA, long noncoding RNA.

<sup>a</sup>Position is based on hg19 and dbSNP Build 137. <sup>b</sup>Functional data are from VEP, the eQTL browser, Fairfax *et al.*<sup>20</sup>, PubMed searches and the 1000 Genomes Project. Minus signs indicate intergenic signals with no additional annotation and no reported regulatory consequence. <sup>c</sup>Imputed variant.

of the immune system processes implicated in multiple sclerosis. We estimate that the 110 established non-MHC risk variants explain 20% of the sibling recurrence risk, 28% when including the already identified MHC effects<sup>9</sup> (**Supplementary Note**). Additionally, we have identified five previously associated regions (*TNFSF14*, *IL2RA*, *TNFRSF1A*, *IL12A* and *STAT4*) where consistent high-resolution fine mapping implicated one variant that accounted for more than 50% of the posterior probability of association. Our study further implicates NF- $\kappa$ B in multiple sclerosis pathobiology<sup>18</sup>, emphasizes the value of dense fine mapping in large follow-up data sets and exposes the urgent need for functional annotation in relevant tissues. Understanding the implicated networks and their relationship to environmental risk factors will promote the development of rational therapies and may enable the development of preventive strategies.

**URLs.** ImmunoBase, <http://www.immunobase.org/>; eQTL browser, <http://eqtl.uchicago.edu/>; MetaCore, <https://portal.genegoo.com/>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank the participants, the referring nurses, the physicians and the funders. Funding was provided by the US National Institutes of Health, the Wellcome Trust, the UK MS Society, the UK Medical Research Council, the US National MS Society, the Cambridge National Institute for Health Research (NIHR) Biomedical Research Centre, DeNDRon, the Bibbi and Niels Jensens Foundation, the Swedish Brain Foundation, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Foundation for Strategic Research, the Stockholm County Council, Karolinska Institutet, INSERM, Fondation d'Aide pour la Recherche sur la Sclérose en Plaques, Association Française contre les Myopathies, Infrastructures en Biologie Santé et Agronomie (GIS-IBISA), the German Ministry for Education and Research, the German Competence Network MS, Deutsche Forschungsgemeinschaft,

Munich Biotech Cluster M4, the Fidelity Biosciences Research Initiative, Research Foundation Flanders, Research Fund KU Leuven, the Belgian Charcot Foundation, Gemeinnützige Hertie Stiftung, University Zurich, the Danish MS Society, the Danish Council for Strategic Research, the Academy of Finland, the Sigrid Juselius Foundation, Helsinki University, the Italian MS Foundation, Fondazione Cariplo, the Italian Ministry of University and Research, the Torino Savings Bank Foundation, the Italian Ministry of Health, the Italian Institute of Experimental Neurology, the MS Association of Oslo, the Norwegian Research Council, the South-Eastern Norwegian Health Authorities, the Australian National Health and Medical Research Council, the Dutch MS Foundation and Kaiser Permanente. We acknowledge the British 1958 Birth Cohort, the UK National Blood Service, Vanderbilt University Medical Center's BioVU DNA Resources Core, Centre de Ressources Biologiques du Réseau Français d'Etude Génétique de la Sclérose en Plaques, the Norwegian Bone Marrow Registry, the Norwegian MS Registry and Biobank, the North American Research Committee on MS Registry, the Brigham and Women's Hospital PhenoGenetic Project and DILGOM, funded by the Academy of Finland. See the **Supplementary Note** for details.

#### AUTHOR CONTRIBUTIONS

M.F.D., D. Booth, A.O., J.S., B. Fontaine, B.H., C. Martin, F.Z., S.D., F.M.-B., B.T., H.F.H., I. Kockum, J. Hillert, T.O., J.R.O., R.H., L.F.B., C. Agliardi, L.A., L. Bernardinelli, V.B., S.B., B.B., L. Brundin, D. Buck, H. Butzkueven, W. Camu, P.C., E.G.C., I.C., G.C., I.C.-R., B.A.C.C., G.D., S.R.D., A.d.S., B.D., M.D., I.E., F.E., N.E., J.F., A.F., I.Y.F., D.G., C. Graetz, A. Graham, C. Guaschino, C. Halfpenny, G. Hall, J. Harley, T.H., C. Hawkins, C. Hillier, J. Hobart, M.H., I.J., A.J., B.K., A. Kermode, T. Kilpatrick, K.K., T. Korn, H.K., C.L.-F., J.L.-S., M.H.L., M.A.L., G.L., B.A.L., C.M.L., F.L., J. Lycke, S.M., C.P.M., R.M., V.M., D.M., G. Mazibrada, J.M., K.-M.M., G.N., R.N., P.N., F.P., S.E.P., H.Q., M. Reunanen, W.R., N.P.R., M. Rodegher, D.R., M. Salvetti, F.S., R.C.S., C. Schaefer, S. Shaunak, L.S., S. Shields, V.S., M. Slee, P.S.S., M. Sospedra, A. Spurkland, V.T., J.T., A.T., P.T., C.v.D., E.M.V., S.V., J.S.W., A.W., J.F.W., J.Z., E.Z., J.L.H., M.A.P.-V., G.S., D.H., S.L.H., A.C., P.D.J., S.J.S. and J.L.M. were involved with case ascertainment and phenotyping. A. Kempainen, D. Booth, A. Goris, A.O., B. Fontaine, S.D., F.M.-B., H.F.H., I. Kockum, M.B., J.R.O., L.F.B., IIBDGC, H.B.S., A. Baker, N.B., L. Bergamaschi, I.L.B., P.B., D. Buck, S.J.C., L. Corrado, L. Cosemans, I.C.-R., V.D., J.F., A.F., V.G., I.J., I. Konidari, V.L., C.M.L., M. Lindén, J. Link, C. McCabe, I.-L.M., H.Q., M. Sorosina, E.S., H.W., P.D.J., S.J.S. and J.L.M. processed the DNA. A. Kempainen, A.O., B. Fontaine, M.B., R.H., L.F.B., WTCCC2, IIBDGC, R.A., H.B.S., N.B., T.M.C.B., H. Blackburn, P.B., W. Carpentier, L. Corrado, I.C.-R., D.C., V.D., P. Deloukas, S.E., A.F., H.H., P.H., A. Hamsten, S.E.H., I.J., I. Konidari, C.L., M. Larsson, M. Lathrop, F.M., I.-L.M., J.M., H.Q., F.S., M. Sorosina, C.v.D., J.W., D.H., P.D.J., S.J.S. and J.L.M. conducted and supervised the genotyping of samples. A.H.B., N.A.P., D.K.X., M.F.D., A. Kempainen, C.C., T.S.S., C. Spencer, M.B., IIBDGC, C. Anderson, S.E.B., A.T.D., P. Donnelly, B. Fiddes, P.-A.G., G. Hellenthal, S.E.H., L.M., M.P., N.C.S.-B., J.L.H., M.A.P.-V., G. McVean, P.D.J., S.J.S. and J.L.M. performed the statistical analysis. A.H.B., N.A.P., D.K.X., M.F.D., A. Kempainen, C.C., T.S.S., C. Spencer, D. Booth, A. Goris, A.O., J.S., B. Fontaine, B.H., F.Z., S.D., F.M.-B., H.F.H., I. Kockum, M.B., R.H., L.F.B., C. Agliardi, M.A., C. Anderson, R.A., H.B.S., A. Baker, G.B., N.B., J.B., C.B., L. Bernardinelli, A. Berthele, V.B., T.M.C.B., H. Blackburn, I.L.B., B.B., D. Buck, S.J.C., W. Camu, P.C., E.G.C., I.C., G.C., L. Corrado, L. Cosemans, I.C.-R., B.A.C.C., D.C., G.D., S.R.D., P. Deloukas, A.d.S., A.T.D., P. Donnelly, B.D., M.D., S.E., F.E., N.E., B. Fiddes, J.F., A.F., C.F., D.G., C. Gieger, C. Graetz, A. Graham, V.G., C. Guaschino, A. Hadjixenofontos, H.H., C. Halfpenny, P.H., G. Hall, A. Hamsten, J. Harley, T.H., C. Hawkins, G. Hellenthal, C. Hillier, J. Hobart, M.H., S.E.H., I.J., A.J., B.K., I. Konidari, H.K., C.L., M. Larsson, M. Lathrop, C.L.-F., M.A.L., V.L., G.L., B.A.L., C.M.L., F.M., C.P.M., R.M., V.M., G. Mazibrada, C. McCabe, I.-L.M., L.M., K.-M.M., R.N., M.P., S.E.P., H.Q., N.P.R., M. Rodegher, D.R., M. Salvetti, N.C.S.-B., R.C.S., C. Schaefer, S. Shaunak, L.S., S. Shields, M. Sospedra, A. Strange, J.T., A.T., E.M.V., A.W., J.F.W., J.W., J.Z., J.L.H., A.J.I., G. McVean, P.D.J., S.J.S. and J.L.M. collected and managed the project data. A.H.B., N.A.P., M.F.D., A. Kempainen, C.C., T.S.S., C. Spencer, J.S., B.H., F.Z., S.D., F.M.-B., H.F.H., J. Hillert, T.O., M.B., J.R.O., R.H., L.F.B., L.A., C. Anderson, G.B., J.B., C.B., A. Berthele, E.G.C., G.C., P. Donnelly, F.E., C.F., C. Gieger, C. Graetz, G. Hellenthal, M.J., T. Korn, M.A.L., R.M., M.P., M. Sospedra, A. Spurkland, A. Strange, J.W., J.L.H., M.A.P.-V., A.J.I., G.S., D.H., S.L.H., A.C., G. McVean, P.D.J., S.J.S. and J.L.M. contributed to the study concept and design. A.H.B., N.A.P., D.K.X., G. McVean, P.D.J., S.J.S. and J.L.M. prepared the manuscript. All authors reviewed the final manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Gourraud, P.A., Harbo, H.F., Hauser, S.L. & Baranzini, S.E. The genetics of multiple sclerosis: an up-to-date review. *Immunol. Rev.* **248**, 87–103 (2012).
- Nylander, A. & Hafler, D.A. Multiple sclerosis. *J. Clin. Invest.* **122**, 1180–1188 (2012).
- Compston, A. *et al.* *McAlpine's Multiple Sclerosis* (Churchill Livingstone, London, 2006).
- Dyment, D.A., Yee, I.M., Ebers, G.C. & Sadovnick, A.D. Multiple sclerosis in stepsiblings: recurrence risk and ascertainment. *J. Neurol. Neurosurg. Psychiatry* **77**, 258–259 (2006).
- Hemminki, K., Li, X., Sundquist, J., Hillert, J. & Sundquist, K. Risk for multiple sclerosis in relatives and spouses of patients diagnosed with autoimmune and related conditions. *Neurogenetics* **10**, 5–11 (2009).
- Jersild, C., Svejgaard, A. & Fog, T. HL-A antigens and multiple sclerosis. *Lancet* **1**, 1240–1241 (1972).
- International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genome-wide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
- De Jager, P.L. *et al.* Meta-analysis of genome scans and replication identify *CD6*, *IRF8* and *TNFRSF1A* as new multiple sclerosis susceptibility loci. *Nat. Genet.* **41**, 776–782 (2009).
- International Multiple Sclerosis Genetics Consortium & Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
- Patsopoulos, N.A. *et al.* Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* **70**, 897–912 (2011).
- International Multiple Sclerosis Genetics Consortium. Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am. J. Hum. Genet.* **86**, 621–625 (2010).
- Baranzini, S.E. The genetics of autoimmune diseases: a networked perspective. *Curr. Opin. Immunol.* **21**, 596–605 (2009).
- Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254 (2011).
- Cortes, A. & Brown, M.A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
- Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
- Krzyszowski, M. *et al.* Circo: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
- Willis, T.G. *et al.* *Bcl10* is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. *Cell* **96**, 35–45 (1999).
- Yan, J. & Greer, J.M. NF- $\kappa$ B, a potential therapeutic target for the treatment of multiple sclerosis. *CNS Neurol. Disord. Drug Targets* **7**, 536–557 (2008).
- Wegener, E. & Krappmann, D. CARD-Bcl10-Malt1 signalosomes: missing link to NF- $\kappa$ B. *Sci. STKE* **2007**, pe21 (2007).
- Fairfax, B.P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
- Lill, C.M. *et al.* Genome-wide significant association of *ANKRD55* rs6859219 and multiple sclerosis risk. *J. Med. Genet.* **50**, 140–143 (2013).
- Maier, L.M. *et al.* *IL2RA* genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* **5**, e1000322 (2009).
- Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
- McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
- Gregory, A.P. *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* **488**, 508–511 (2012).
- De Jager, P.L. *et al.* The role of the *CD58* locus in multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **106**, 5264–5269 (2009).
- Malmström, C. *et al.* Serum levels of LIGHT in MS. *Mult. Scler.* **19**, 871–876 (2013).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).
- Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
- Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
- Juran, B.D. *et al.* Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum. Mol. Genet.* **21**, 5209–5221 (2012).
- Liu, J.Z. *et al.* Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **44**, 1137–1141 (2012).
- Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
- Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336–1340 (2012).
- Tsoi, L.C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341–1348 (2012).
- Cooper, J.D. *et al.* Seven newly identified loci for autoimmune thyroid disease. *Hum. Mol. Genet.* **21**, 5202–5208 (2012).
- Ban, M. *et al.* Replication analysis identifies *TYK2* as a multiple sclerosis susceptibility factor. *Eur. J. Hum. Genet.* **17**, 1309–1313 (2009).
- Ban, M. *et al.* A non-synonymous SNP within membrane metalloendopeptidase-like 1 (*MMEL1*) is associated with multiple sclerosis. *Genes Immun.* **11**, 660–664 (2010).

Ashley H Beecham<sup>1,127</sup>, Nikolaos A Patsopoulos<sup>2-6,127</sup>, Dionysia K Xifara<sup>7</sup>, Mary F Davis<sup>8</sup>, Anu Kempainen<sup>9</sup>, Chris Cotsapas<sup>10-12</sup>, Tejas S Shah<sup>13</sup>, Chris Spencer<sup>7</sup>, David Booth<sup>14</sup>, An Goris<sup>15</sup>, Annette Oturai<sup>16</sup>, Janna Saarela<sup>17</sup>, Bertrand Fontaine<sup>18</sup>, Bernhard Hemmer<sup>19-21</sup>, Claes Martin<sup>22</sup>, Frauke Zipp<sup>23</sup>, Sandra D'Alfonso<sup>24,25</sup>, Filippo Martinelli-Boneschi<sup>26,27</sup>, Bruce Taylor<sup>28</sup>, Hanne F Harbo<sup>29,30</sup>, Ingrid Kockum<sup>31</sup>, Jan Hillert<sup>31</sup>, Tomas Olsson<sup>31</sup>, Maria Ban<sup>9</sup>, Jorge R Oksenberg<sup>32</sup>, Rogier Hintzen<sup>33</sup>, Lisa F Barcellos<sup>34-36</sup>, Wellcome Trust Case Control Consortium 2 (WTCCC2)<sup>37</sup>, International IBD Genetics Consortium (IIBDGC)<sup>37</sup>, Cristina Agliardi<sup>38</sup>, Lars Alfredsson<sup>39</sup>, Mehdi Alizadeh<sup>40</sup>, Carl Anderson<sup>13</sup>, Robert Andrews<sup>13</sup>, Helle Bach Søndergaard<sup>16</sup>, Amie Baker<sup>9</sup>, Gavin Band<sup>7</sup>, Sergio E Baranzini<sup>32</sup>, Nadia Barizzone<sup>24,25</sup>, Jeffrey Barrett<sup>13</sup>, Céline Bellenguez<sup>7</sup>, Laura Bergamaschi<sup>24,25</sup>, Luisa Bernardinelli<sup>41</sup>, Achim Berthele<sup>19</sup>, Viola Biberacher<sup>19</sup>, Thomas M C Binder<sup>42</sup>, Hannah Blackburn<sup>13</sup>, Izaaura L Bomfim<sup>31</sup>, Paola Brambilla<sup>26</sup>, Simon Broadley<sup>43</sup>, Bruno Brochet<sup>44</sup>, Lou Brundin<sup>31</sup>, Dorothea Buck<sup>19</sup>, Helmut Butzkueven<sup>45,46</sup>, Stacy J Caillier<sup>32</sup>, William Camu<sup>47</sup>, Wassila Carpentier<sup>48</sup>, Paola Cavalla<sup>49,50</sup>, Elisabeth G Celius<sup>29</sup>, Irène Coman<sup>51</sup>, Giancarlo Comi<sup>26,27</sup>, Lucia Corrado<sup>24,25</sup>, Leentje Cosemans<sup>15</sup>, Isabelle Cournu-Rebeix<sup>18</sup>, Bruce A C Cree<sup>32</sup>, Daniele Cusi<sup>52</sup>, Vincent Damotte<sup>18</sup>, Gilles Defer<sup>53</sup>, Silvia R Delgado<sup>54</sup>, Panos Deloukas<sup>13</sup>, Alessia di Sapio<sup>55</sup>, Alexander T Dilthey<sup>7</sup>, Peter Donnelly<sup>7</sup>, Bénédicte Dubois<sup>15</sup>, Martin Duddy<sup>56</sup>, Sarah Edkins<sup>13</sup>, Irina Elovaara<sup>57</sup>, Federica Esposito<sup>26,27</sup>, Nikos Evangelou<sup>58</sup>, Barnaby Fiddes<sup>9</sup>, Judith Field<sup>59</sup>, Andre Franke<sup>60</sup>, Colin Freeman<sup>7</sup>, Irene Y Frohlich<sup>2</sup>, Daniela Galimberti<sup>61,62</sup>, Christian Gieger<sup>63</sup>, Pierre-Antoine Gourraud<sup>32</sup>, Christiane Graetz<sup>23</sup>, Andrew Graham<sup>64</sup>, Verena Grummel<sup>19</sup>, Clara Guaschino<sup>26,27</sup>, Athena Hadjixenofontos<sup>1</sup>, Hakon Hakonarson<sup>65,66</sup>, Christopher Halfpenny<sup>67</sup>, Gillian Hall<sup>68</sup>, Per Hall<sup>69</sup>, Anders Hamsten<sup>70</sup>, James Harley<sup>71</sup>, Timothy Harrower<sup>72</sup>, Clive Hawkins<sup>73</sup>, Garrett Hellenthal<sup>74</sup>, Charles Hillier<sup>75</sup>, Jeremy Hobart<sup>76</sup>, Muni Hoshi<sup>19</sup>, Sarah E Hunt<sup>13</sup>, Maja Jagodic<sup>31</sup>, Ilijas Jelčić<sup>77,78</sup>, Angela Jochim<sup>19</sup>, Brian Kendall<sup>79</sup>, Allan Kermodé<sup>80,81</sup>, Trevor Kilpatrick<sup>82</sup>, Keijo Koivisto<sup>83</sup>, Ioanna Konidari<sup>1</sup>, Thomas Korn<sup>19</sup>, Helena Kronsbein<sup>19</sup>, Cordelia Langford<sup>13</sup>, Malin Larsson<sup>84</sup>, Mark Lathrop<sup>85-87</sup>, Christine Lebrun-Frenay<sup>88</sup>, Jeannette Lechner-Scott<sup>89</sup>, Michelle H Lee<sup>2</sup>, Maurizio A Leone<sup>90</sup>, Virpi Leppä<sup>17</sup>, Giuseppe Liberatore<sup>26,27</sup>, Benedicte A Lie<sup>30,91</sup>, Christina M Lill<sup>23,92</sup>, Magdalena Lindén<sup>31</sup>, Jenny Link<sup>31</sup>, Felix Luessi<sup>23</sup>, Jan Lycke<sup>93</sup>, Fabio Macchiardi<sup>94,95</sup>, Satu Männistö<sup>96</sup>, Clara P Manrique<sup>1</sup>, Roland Martin<sup>77,78</sup>, Vittorio Martinelli<sup>27</sup>, Deborah Mason<sup>97</sup>, Gordon Mazibrada<sup>98</sup>, Cristin McCabe<sup>10</sup>, Inger-Lise Mero<sup>29,30,91</sup>, Julia Mescheriakova<sup>33</sup>, Loukas Moutsianas<sup>7</sup>, Kjell-Morten Myhr<sup>99</sup>, Guy Nagels<sup>100</sup>, Richard Nicholas<sup>101</sup>, Petra Nilsson<sup>102</sup>, Fredrik Piehl<sup>31</sup>, Matti Pirinen<sup>7</sup>, Siân E Price<sup>103</sup>, Hong Quach<sup>34</sup>, Mauri Reunanen<sup>104,105</sup>, Wim Robberecht<sup>106-108</sup>, Neil P Robertson<sup>109</sup>, Mariaemma Rodegher<sup>27</sup>, David Rog<sup>110</sup>, Marco Salvetti<sup>111</sup>, Nathalie C Schnetz-Boutaud<sup>8</sup>, Finn Sellebjerg<sup>16</sup>, Rebecca C Selter<sup>19</sup>, Catherine Schaefer<sup>36</sup>, Sandip Shaunak<sup>112</sup>, Ling Shen<sup>36</sup>, Simon Shields<sup>113</sup>, Volker Siffrin<sup>23</sup>, Mark Slee<sup>114</sup>, Per Soelberg Sorensen<sup>16</sup>, Melissa Sorosina<sup>26</sup>, Mireia Sospedra<sup>77,78</sup>, Anne Spurkland<sup>115</sup>, Amy Strange<sup>7</sup>, Emilie Sundqvist<sup>31</sup>, Vincent Thijs<sup>106-108</sup>, John Thorpe<sup>116</sup>, Anna Ticca<sup>117</sup>, Pentti Tienari<sup>118</sup>, Cornelia van Duijn<sup>119</sup>, Elizabeth M Visser<sup>120</sup>, Steve Vucic<sup>14</sup>, Helga Westerlind<sup>31</sup>, James S Wiley<sup>59</sup>, Alastair Wilkins<sup>121</sup>, James F Wilson<sup>122</sup>, Juliane Winkelmann<sup>19,20,123,124</sup>, John Zajicek<sup>76</sup>, Eva Zindler<sup>23</sup>, Jonathan L Haines<sup>8</sup>, Margaret A Pericak-Vance<sup>1</sup>, Adrian J Ivinson<sup>125</sup>, Graeme Stewart<sup>14</sup>, David Hafler<sup>10,11,126</sup>, Stephen L Hauser<sup>32</sup>, Alastair Compston<sup>9</sup>, Gil McVean<sup>7</sup>, Philip De Jager<sup>2,5,10,127,128</sup>, Stephen J Sawcer<sup>9,127,128</sup> & Jacob L McCauley<sup>1,127,128</sup>

<sup>1</sup>John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida, USA. <sup>2</sup>Department of Neurology, Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Brigham & Women's Hospital, Boston, Massachusetts, USA. <sup>3</sup>Department of Psychiatry, Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Brigham & Women's Hospital, Boston, Massachusetts, USA. <sup>4</sup>Department of Medicine, Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. <sup>5</sup>Harvard Medical School, Boston, Massachusetts, USA. <sup>6</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. <sup>7</sup>The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>8</sup>Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA. <sup>9</sup>Department of Clinical Neurosciences, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. <sup>10</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. <sup>11</sup>Department of Neurology, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>12</sup>Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>13</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>14</sup>Westmead Millennium Institute, University of Sydney, Westmead, New South Wales, Australia. <sup>15</sup>Section of Experimental Neurology, Laboratory for Neuroimmunology, KU Leuven, Leuven, Belgium. <sup>16</sup>Department of Neurology, Danish Multiple Sclerosis Center, Copenhagen University Hospital, Copenhagen, Denmark. <sup>17</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. <sup>18</sup>Département de Neurologie, INSERM UMR 975 Centre de Recherche de l'Institut du Cerveau et la Moelle Epinière (CRICM), Université Pierre et Marie Curie (UPMC), Pitié-Salpêtrière, Paris, France. <sup>19</sup>Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. <sup>20</sup>Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. <sup>21</sup>German Competence Network Multiple Sclerosis (KKNMS), Munich, Germany. <sup>22</sup>Department of Clinical Sciences, Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden. <sup>23</sup>Focus Program Translational Neuroscience (FTN), Rhine Main Neuroscience Network (rmn2), Johannes Gutenberg University Medical Center, Mainz, Germany. <sup>24</sup>Department of Health Sciences, University of Eastern Piedmont, Novara, Italy. <sup>25</sup>Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy. <sup>26</sup>Laboratory of Genetics of Neurological Complex Disorders, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele

Scientific Institute, Milan, Italy. <sup>27</sup>Department of Neurology, INSPE, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy. <sup>28</sup>Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia. <sup>29</sup>Department of Neurology, Oslo University Hospital, Ullevål, Oslo, Norway. <sup>30</sup>University of Oslo, Oslo, Norway. <sup>31</sup>Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. <sup>32</sup>Department of Neurology, University of California, San Francisco, Sandler Neurosciences Center, San Francisco, California, USA. <sup>33</sup>Department of Neurology, Erasmus MC, Erasmus University Medical Centre, Rotterdam, The Netherlands. <sup>34</sup>Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, Berkeley, California, USA. <sup>35</sup>California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, California, USA. <sup>36</sup>Kaiser Permanente Division of Research, Oakland, California, USA. <sup>37</sup>Full lists of members and affiliations appear in the **Supplementary Note**. <sup>38</sup>Laboratory of Molecular Medicine and Biotechnology, Don C. Gnocchi Foundation Organizzazione Non Lucrativa di Utilità Sociale (ONLUS), Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Santa Maria Nascente, Milan, Italy. <sup>39</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>40</sup>Laboratoire d'Immunologie, Université Rennes 1, Rennes, France. <sup>41</sup>Medical Research Council Biostatistics Unit, Cambridge, UK. <sup>42</sup>Department of Transfusion Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>43</sup>School of Medicine, Griffith University, Gold Coast, Queensland, Australia. <sup>44</sup>Centre Hospitalier Universitaire (CHU) Pellegrin, Université Bordeaux 2, Bordeaux, France. <sup>45</sup>Department of Medicine, University of Melbourne, Melbourne, Victoria, Australia. <sup>46</sup>Department of Neurology, Box Hill Hospital, Monash University, Box Hill, Victoria, Australia. <sup>47</sup>Service de Neurologie, Centre Hospitalier Universitaire Régional Montpellier, Montpellier, France. <sup>48</sup>Plateforme Post-Génomique P3S, Université Pierre et Marie Curie, INSERM, Paris, France. <sup>49</sup>Department of Neuroscience, Multiple Sclerosis Center, Azienda Ospedaliera Città della Salute e della Scienza di Torino, Turin, Italy. <sup>50</sup>Department of Neuroscience, University of Turin, Turin, Italy. <sup>51</sup>Service de Neurologie, Hôpital Avicenne, Bobigny, France. <sup>52</sup>Department of Health Sciences, San Paolo Hospital and Filarete Foundation, University of Milan, Milan, Italy. <sup>53</sup>Service de Neurologie, CHU de Caen and INSERM U919–Groupement d'Intérêt Public Cyceron, Caen, France. <sup>54</sup>Department of Neurology, Multiple Sclerosis Division, Miller School of Medicine, University of Miami, Miami, Florida, USA. <sup>55</sup>Neurologia 2–Regional Multiple Sclerosis Centre, Azienda Ospedaliera Universitaria San Luigi, Orbassano, Turin, Italy. <sup>56</sup>Department of Neurology, Royal Victoria Infirmary, Newcastle-upon-Tyne, UK. <sup>57</sup>Department of Neurology, University of Tampere, Medical School, Tampere, Finland. <sup>58</sup>Division of Clinical Neurology, Nottingham University Hospital, Nottingham, UK. <sup>59</sup>Florey Institute of Neuroscience and Mental Health, University of Melbourne, Melbourne, Victoria, Australia. <sup>60</sup>Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany. <sup>61</sup>Department of Pathophysiology and Transplantation, Neurology Unit, University of Milan, Milan, Italy. <sup>62</sup>Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy. <sup>63</sup>Kora-gen, Helmholtz Zentrum München–German Research Center for Environmental Health, Institute of Genetic Epidemiology, Neuherberg, Germany. <sup>64</sup>Department of Clinical Neurology, The Ipswich Hospital National Health Service (NHS) Trust, Ipswich, UK. <sup>65</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>66</sup>Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>67</sup>Wellcome Trust Clinical Research Facility, Southampton General Hospital, Southampton, UK. <sup>68</sup>Department of Neurology, Aberdeen Royal Infirmary, Aberdeen, UK. <sup>69</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>70</sup>Department of Medicine at Karolinska University Hospital Solna, Atherosclerosis Research Unit, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>71</sup>Department of Neurology, Hull Royal Infirmary, Hull, UK. <sup>72</sup>Department of Neurology, Royal Devon and Exeter Foundation Trust Hospital, Exeter, UK. <sup>73</sup>Keele University Medical School, University Hospital of North Staffordshire, Stoke-on-Trent, UK. <sup>74</sup>UCL Genetics Institute (UGI), University College London, London, UK. <sup>75</sup>Department of Neurology, Poole General Hospital, Poole, UK. <sup>76</sup>Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK. <sup>77</sup>Institute for Neuroimmunology and Clinical Multiple Sclerosis Research (inims), Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>78</sup>Department of Neuroimmunology and Multiple Sclerosis Research, Neurology Clinic, University Hospital Zürich, Zürich, Switzerland. <sup>79</sup>Department of Neurology, Division of Clinical Neurology, Leicester Royal Infirmary, Leicester, UK. <sup>80</sup>Australian Neuromuscular Research Institute, University of Western Australia, Nedlands, Western Australia, Australia. <sup>81</sup>Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia. <sup>82</sup>Melbourne Neuroscience Institute, University of Melbourne, Melbourne, Victoria, Australia. <sup>83</sup>Department of Neurology, Seinäjoki Central Hospital, Seinäjoki, Finland. <sup>84</sup>Department of Physics, Chemistry and Biology (IFM) Bioinformatics, Linköping University, Linköping, Sweden. <sup>85</sup>Fondation Jean Dausset–Centre d'Etude du Polymorphisme Humain, Paris, France. <sup>86</sup>Commissariat à l'Energie Atomique, Institut Génomique, Centre National de Génotypage, Evry, France. <sup>87</sup>McGill University and Génomique Québec Innovation Centre, Montreal, Quebec, Canada. <sup>88</sup>Service de Neurologie, Hôpital Pasteur, CHRU Nice, France. <sup>89</sup>Hunter Medical Research Institute, University of Newcastle, Newcastle, New South Wales, Australia. <sup>90</sup>Department of Neurology, Ospedale Maggiore, Novara, Italy. <sup>91</sup>Department of Medical Genetics, Oslo University Hospital, Ullevål, Oslo, Norway. <sup>92</sup>Department of Vertebrate Genomics, Neuropsychiatric Genetics Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>93</sup>Department of Clinical Neurosciences and Rehabilitation, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Göteborgs Universitet, Göteborg, Sweden. <sup>94</sup>Department of Psychiatry and Human Behavior, School of Medicine, University of California–Irvine, Irvine, California, USA. <sup>95</sup>Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy. <sup>96</sup>Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. <sup>97</sup>Canterbury District Health Board, Christchurch, New Zealand. <sup>98</sup>Department of Neurology, Queen Elizabeth Medical Centre, Edgbaston, Birmingham, UK. <sup>99</sup>Department of Neurology, The Norwegian Multiple Sclerosis Registry and Biobank, Haukeland University Hospital, Bergen, Norway. <sup>100</sup>National Multiple Sclerosis Center Melsbroek, Melsbroek, Belgium. <sup>101</sup>Neurology Department, Charing Cross Hospital, London, UK. <sup>102</sup>Department of Clinical Sciences, Lund University, Lund, Sweden. <sup>103</sup>Department of Neurology, Royal Hallamshire Hospital, Sheffield, UK. <sup>104</sup>Department of Neurology, University of Oulu, Oulu, Finland. <sup>105</sup>Department of Neurology, University Hospital of Oulu, Oulu, Finland. <sup>106</sup>Laboratory of Neurobiology, Vesalius Research Center, Leuven, Belgium. <sup>107</sup>Experimental Neurology, Leuven Research Institute for Neurodegenerative Diseases (LIND), University of Leuven (KU Leuven), Leuven, Belgium. <sup>108</sup>Department of Neurology, University Hospitals Leuven, Leuven, Belgium. <sup>109</sup>Institute of Psychological Medicine and Clinical Neuroscience, Cardiff University, University Hospital of Wales, Cardiff, UK. <sup>110</sup>Department of Neurology, Greater Manchester Neurosciences Centre, Salford Royal NHS Foundation Trust, Salford, UK. <sup>111</sup>Department of Neuroscience, Centre for Experimental Neurological Therapies, Mental Health and Sensory Organs, Sapienza Università di Roma, Rome, Italy. <sup>112</sup>Department of Neurology, Royal Preston Hospital, Preston, UK. <sup>113</sup>Department of Neurology, Norfolk and Norwich Hospital, Norwich, UK. <sup>114</sup>Department of Neurology, Flinders University, Adelaide, South Australia, Australia. <sup>115</sup>Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway. <sup>116</sup>Department of Neurology, Peterborough City Hospital, Peterborough, UK. <sup>117</sup>Neurology and Stroke Unit, San Francesco Hospital, Nuoro, Italy. <sup>118</sup>Department of Neurology, Helsinki University Central Hospital and Molecular Neurology Programme, Biomedicum, University of Helsinki, Helsinki, Finland. <sup>119</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>120</sup>Division of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK. <sup>121</sup>Institute of Clinical Neurosciences, University of Bristol, Frenchay Hospital, Bristol, UK. <sup>122</sup>Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK. <sup>123</sup>Institut für Humangenetik, Technische Universität München, Munich, Germany. <sup>124</sup>Institut für Humangenetik, Helmholtz Zentrum München–Munich, Germany. <sup>125</sup>Harvard NeuroDiscovery Center, Harvard Medical School, Boston, Massachusetts, USA. <sup>126</sup>Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>127</sup>These authors contributed equally to this work. <sup>128</sup>These authors jointly directed this work. Correspondence should be addressed to J.L.M. ([jmccauley@med.miami.edu](mailto:jmccauley@med.miami.edu)).

## ONLINE METHODS

**ImmunoChip data (discovery set).** Details of case ascertainment, processing and genotyping for the discovery phase are provided in the **Supplementary Note** (see also **Supplementary Table 9**). All cases and controls involved in this study gave valid informed consent in accordance with approval from the relevant local ethical committees or institutional review boards (IRBs). The investigators were not blinded to allocation during experiments and outcome assessment. Genotype calling for all samples was performed using Optical<sup>40</sup>. Samples that performed poorly or were determined to be related were removed (**Supplementary Table 10**). Data were organized into 11 country-level strata: ANZ (Australia and New Zealand), Belgium, Denmark, Finland, France, Germany, Italy, Norway, Sweden, the UK and the United States. SNP-level quality control (**Supplementary Table 11**) and population outlier identification using principal-components analysis (**Supplementary Fig. 97**) were carried out in each stratum separately. No statistical method was used to pre-determine sample size.

**Discovery set analysis.** We applied logistic regression, assuming a per-allelic genetic model for each data set, including the first five principal components as covariates to correct for population stratification (**Supplementary Table 12** lists the genomic inflation factors for each data set,  $\lambda$ ). We then performed an inverse variance meta-analysis of the 11 strata under a fixed-effects model, as implemented in PLINK<sup>41</sup>. To be more conservative and account for any residual inflation in the test statistic, we applied the genomic control equivalent to the per-SNP standard error in each stratum. Specifically, we corrected the standard errors for SNPs by multiplying them by the square root of the raw genomic inflation factor,  $\lambda$ , for each data set if  $\lambda$  was  $>1$ .

Within the designated fine-mapping intervals, we applied a forward stepwise logistic regression to identify statistically independent effects. The primary SNP in each interval was included as a covariate, and association analysis was repeated for the remaining SNPs. This process was repeated until no SNPs reached the minimum level of significance ( $P < 1 \times 10^{-4}$ ). Outside of the designated fine-mapping intervals, all SNPs having association  $P < 1 \times 10^{-4}$  were identified and grouped into sets on the basis of being separated by a physical distance of less than 2 Mb, and a similar stepwise regression model was applied. Any SNPs that entered the model with association  $P < 1 \times 10^{-4}$  after conditioning were considered to be statistically independent primary signals.

In addition, because of the close physical proximity of some fine-mapping intervals and SNP sets, independence was tested for all identified signals within 2 Mb of one another. Forest and cluster plots (**Supplementary Fig. 98**) of all independent SNPs were examined, and a SNP was excluded if cluster plots were unsatisfactory. If a SNP was excluded, forward stepwise logistic regression within the corresponding fine-mapping interval or SNP set was repeated after removal of the SNP. During this process, 17 additional SNPs were excluded on the basis of cluster or forest plot review.

**Replication set.** The replication phase included GWAS data organized into 15 strata. Within each stratum, poorly performing samples (call rate  $< 95\%$ , sex discordance or excess heterozygosity) and poorly performing SNPs (Hardy-Weinberg equilibrium  $P < 1 \times 10^{-6}$ , minor allele frequency (MAF)  $< 1\%$  or call rate  $< 95\%$ ) were removed. Principal-components analysis was performed to identify population outliers for each stratum, and the genomic control inflation factor was  $< 1.1$  for each. Data included in the final discovery and replication analyses are summarized in **Supplementary Tables 13** and **14**. All samples used in the replication set were unrelated to those in the discovery set, as verified by identity-by-descent analysis.

We attempted replication of all non-MHC independent signals that reached discovery  $P < 1 \times 10^{-4}$  in a meta-analysis set of GWAS. Each data set was imputed with the 1000 Genomes Project European phase I (a) panel using BEAGLE<sup>42</sup> to maximize the overlap between the ImmunoChip SNP content and the GWAS data. Post-imputation genotypic probabilities were used in a logistic regression model, for each stratum, to estimate SNP effect sizes and  $P$  values. By using the post-imputation genotypic probabilities, we penalized SNPs that did not have good imputation quality, thus ensuring a conservative analysis. Furthermore, we accounted for population stratification in each data set by including the first five principal components in the logistic model.

We then performed meta-analysis of the effect sizes and respective standard errors of the 15 strata using a fixed-effects, inverse variance method. We applied the genomic control equivalent to the per-SNP standard error in each stratum, controlling for the respective genomic inflation factor,  $\lambda$  (**Supplementary Table 14**).

To replicate the primary SNPs with identified signal in the discovery phase, we determined the replication effect sizes and respective standard errors. For the secondary and tertiary SNPs, we fitted the same exact models as in the discovery phase for each data set. We then performed fixed-effects meta-analysis to estimate an effect size that corresponded to the same logistic model. In the case that a SNP was not present in the replication set, we replaced it with a perfectly tagging SNP, i.e., a SNP that had  $r^2$  and  $D'$  values equal to 1. If a perfectly tagging SNP was not available, we selected a SNP that had equivalent MAF and the highest possible  $r^2$  and  $D'$  values. Estimation of  $r^2$  and  $D'$  values for this objective were based on data from ImmunoChip control samples.

**Joint analysis (discovery and replication sets).** We performed meta-analysis of the effect sizes and respective standard errors from the discovery and replication phases under a fixed-effects model. A SNP was considered to have replicated when all three of the following criteria were met: (i) replication  $P < 5.0 \times 10^{-2}$ ; (ii) joint  $P < 5 \times 10^{-8}$ ; and (iii) the joint  $P$  value was more statistically significant than the discovery  $P$  value. SNPs that reached  $P < 1 \times 10^{-6}$  but did not pass the genome-wide significance threshold were defined as having suggestive signals if criteria (i) and (iii) were met.

**Fine mapping of association signals.** To fine map association signals, we used a combination of imputation and Bayesian methodology<sup>23</sup>. Around each of the 97 associated SNPs, we isolated 2 Mb of sequence in the discovery and replication phase UK data as well as in the European samples from Phase 1 of the 1000 Genomes Project<sup>28</sup>. Forming the single largest cohort, only UK samples were considered to minimize the effects of differential imputation quality between populations of different ancestry. In addition to the previous quality control, SNPs with failed alignment or difference in MAF  $> 10\%$  between the typed cohorts and the 1000 Genomes Project samples, MAF  $< 1\%$  or Hardy-Weinberg equilibrium  $P < 1.0 \times 10^{-4}$  were removed.

Imputation was performed separately for the UK discovery and replication cohorts on each 2-Mb region using the default settings of IMPUTEv2 (refs. 43,44). Missing genotypes for the genotyped SNPs were not imputed, and any imputed SNP that failed to reach the Hardy-Weinberg equilibrium and MAF thresholds was subsequently removed. We carried out frequentist and Bayesian association tests on all SNPs in each cohort separately, assuming additivity, using the default settings of SNPTESTv2 (ref. 45). Frequentist fixed-effects meta-analysis was carried out using the software META<sup>46</sup>. Bayesian meta-analysis was carried out using an independence prior (near-identical results were obtained using a fixed-effects Bayesian meta-analysis).

To identify regions where reliable fine mapping could be achieved, we used the information score (INFO, obtained from IMPUTEv2) as determined from the 1000 Genomes Project samples. Specifically, we measured the fraction of variants having both  $r^2 > 0.5$  and  $r^2 > 0.8$  with the primary associated variant and having greater than 50% and 80% INFO scores, respectively. Regions where any SNP with  $r^2 > 0.5$  had INFO  $< 50\%$  were excluded. We also excluded regions where the top hit from imputation had an INFO score less than 80%. Regions were considered to be fine mapped with high quality when all variants with  $r^2 > 0.8$  had an INFO score of at least 80%. Within these regions, we excluded variants where the inferred direction of association was opposite between the UK discovery and replication cohorts.

To measure the posterior probability that any single variant drives association, we calculated the Bayes factor. Under the assumption that there is a single causal variant in the region, the Bayes factor is proportional to the probability that the variant drives the association<sup>23</sup>. We identified the smallest set of variants that contained 90% and 50% of the posterior probability. We called a region successfully and consistently fine mapped if there were at most five variants in the 50% confidence interval and the top SNP from the frequentist analysis was contained in the 90% confidence interval. For these regions, we annotated variants with information about evolutionary conservation, predicted coding consequences, regulation, published associations to expression



or DNase I hypersensitive sites using ANNOVAR<sup>47</sup>, VEP<sup>24</sup> and the eQTL browser, a recent study of expression in immune cells<sup>20</sup> and other literature.

**GO classification.** To determine the GO processes in which our 97 associated variants were involved, we used MetaCore from Thomson Reuters. We annotated the processes for the unique genes within 50 kb of the variants.

**Cross-disease comparisons.** To explore the potential overlap with variants identified across other autoimmune diseases, we calculated the percentage overlap of reported variants found in other ImmunoChip reports to our ImmunoChip results. The top variants reported as either newly discovered or previously known in other ImmunoChip reports were compared with the 110 variants representing both our new and previous discoveries in multiple sclerosis. For a signal to be considered to be overlapping, we required  $r^2 \geq 0.8$  using the Pairwise LD function of the SNAP tool in European samples<sup>31</sup>.

**Secondary analyses.** We performed a severity-based analysis of the Multiple Sclerosis Severity Score (MSSS) in cases only from the discovery phase

(**Supplementary Fig. 99**). In addition, a transmission disequilibrium test was performed in 633 trios to test for transmission of the 97 identified risk alleles (**Supplementary Fig. 100**). Details are given in the **Supplementary Note**.

40. Shah, T.S. *et al.* optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* **28**, 1598–1603 (2012).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
42. Browning, B.L. & Yu, Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* **85**, 847–861 (2009).
43. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
44. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
45. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
46. Liu, J.Z. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* **42**, 436–440 (2010).
47. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

### **2.3. Hypothèse 3 : Les régions génomiques non investiguées pourraient contenir des marqueurs de prédisposition**

Certaines régions du génome n'ont pas ou peu été analysées, à cause de contraintes techniques le plus souvent. En réalité, les SNPs sélectionnés pour les études pangénomiques ne couvrent qu'environ 80% du génome. 20% du génome n'ont donc jamais été étudiés:

- soit parce que certaines régions du génome n'ont pas du tout été analysées, aucun SNPs de ces régions n'étant présents sur les puces utilisées,

- soit parce que certaines régions n'étaient que peu couvertes, c'est à dire que le nombre de SNPs présents sur les puces ne permettait pas d'étudier entièrement une région donnée en termes de déséquilibre de liaison.

Il semble aujourd'hui important de se focaliser sur ces régions pas ou peu analysées. Ceci peut être réalisé par deux approches : soit par des études à moyenne échelle (en terme de couverture) nous permettant d'analyser tous les SNPs nécessaires à la couverture des 20% restant, soit par des études plus ciblées au cours desquelles seront analysées des gènes candidats se trouvant dans les régions non étudiées jusqu'à présent.

**Cette hypothèse a été explorée lors de mon doctorat par le projet gène-candidats *CLDN3*, *CLDN5* et *OCLN* présenté ci-après. Ce travail fera l'objet d'une publication, actuellement en cours de préparation (*Publication Damotte et al., en préparation, présentée page 131*).**



**Article (1<sup>er</sup> auteur):**

A pilot study on the role of *OCLN*, *CLDN3* and *CLDN5* genes in multiple sclerosis susceptibility and severity

**Vincent Damotte**, Pierre-Antoine Gourraud, Léna Guillot-Noel, Mohamed El Behi, Isabelle Cournu-Rebeix\*, Bertrand Fontaine\*

\* equal contribution

En préparation



**Introduction :** Les molécules d'adhésion Occludine, Claudine-3 et Claudine-5, respectivement codées par les gènes *OCN*, *CLDN3* et *CLDN5*, sont des constituants majeurs des jonctions serrées formées entre les cellules endothéliales de la BHE (cf **Figure 10**, page 58). Le franchissement de la BHE est une étape importante dans le déclenchement de la SEP car elle est à l'origine de la colonisation du système nerveux central par les cellules immunitaires activées. Ce processus initie le développement de l'inflammation et l'apparition de la démyélinisation. Ces 3 gènes sont localisés dans des régions génomiques qui n'ont pas ou peu été couvertes par les GWAS. Partant de cette constatation, nous avons initié une étude d'association de ces gènes hautement candidats par fonction avec la prédisposition à la SEP, ainsi qu'avec sa sévérité, son âge de début et sa forme d'évolution.

**Matériel et Méthodes :** Nous avons testé l'association de 31 SNPs appartenant aux gènes *OCN*, *CLDN3* et *CLDN5* avec la prédisposition, la sévérité, l'âge de début et la forme de la maladie. Une cohorte française de 810 patients et 626 contrôles a été utilisée pour la phase de découverte. La phase de réplication a été effectuée dans une cohorte américaine de 581 patients et 454 contrôles.

**Résultats :** Dans la population française, le SNP rs13447272 du gène *CLDN5* semblait associé à la prédisposition à la SEP ( $p = 4.889 \times 10^{-5}$ ) et le SNP rs6971390 du gène *CLDN3* semblait associé à la sévérité de la maladie ( $p = 0,0064$  en considérant le score de sévérité MSSS comme une variable continue et  $p = 0,0043$  en considérant deux groupes de patients, un groupe avec un faible score MSSS, et un groupe avec un fort score MSSS). L'étude de réplication dans la cohorte américaine n'a pas permis de conclure sur le rôle des gènes *CLDN3* et *CLDN5* dans la SEP.

**Conclusions :** Nous n'avons pas pu mettre en évidence d'association entre des SNPs des gènes *OCN*, *CLDN3* et *CLDN5* et la prédisposition ou la sévérité à/de la SEP. Cependant, l'étude des "zones d'ombres" du génome et des gènes « oubliés » est à prendre en considération pour les études ultérieures. Nos résultats soulignent l'importance de la taille de cohortes étudiées pour l'identification du rôle des variants rares.

**Ma contribution :** J'ai mené cette étude de sa conception aux résultats. J'ai génotypé les échantillons, réalisé les analyses statistiques et effectué les démarches nécessaires à



l'obtention de la cohorte américaine lors d'un séjour de 3 mois à l'Université de Californie, San Francisco (UCSF). J'ai écrit la première version de l'article et effectué toutes les corrections suggérées lors des relectures.





## **A pilot study on the role of OCLN, CLDN3 and CLDN5 genes in multiple sclerosis susceptibility and severity**

Vincent Damotte<sup>1</sup>, Pierre-Antoine Gourraud<sup>2</sup>, Léna Guillot-Noel<sup>1</sup>, Mohamed El Behi<sup>1,3</sup>, Isabelle Cournu-Rebeix<sup>1,4\*</sup> and Bertrand Fontaine<sup>1,4\*</sup>

\*Equal contributors

<sup>1</sup>INSERM, UMR S 975 CRICM UPMC, 75013 Paris, France

<sup>2</sup>University of California at San Francisco Department of Neurology, San Francisco, California, USA

<sup>3</sup>Institut du Cerveau et de la Moelle épinière, ICM, 75013 Paris, France

<sup>4</sup>Département de Neurologie, Pitié-Salpêtrière Hospital AP-HP , 75013 Paris, France

Corresponding author :

Isabelle COURNU-REBEIX

Hôpital Pitié-Salpêtrière

Bâtiment ICM

UMR\_S975

Equipe Bertrand Fontaine

4<sup>ème</sup> étage

47 boulevard de l'Hôpital

75013 PARIS, France

T : 01 57 27 44 08

F : 01 57 27 47 92

e-mail : [isabelle.rebeix@upmc.fr](mailto:isabelle.rebeix@upmc.fr)

Key Words: Multiple Sclerosis, complex trait, cell adhesion molecules, susceptibility, severity

Word count: 1923

## **ABSTRACT**

**Background:** Genome-Wide Association Studies performed in Multiple Sclerosis (MS) led to identifying 25% of the genetic component of host predisposition. Several hypothesis have been put forward to determine the so-called missing heritability. We legitimately formulated that genes localized in genomic regions that were insufficiently or not investigated by genome-wide studies could be good candidates, especially those which are biologically relevant in MS.

**Objective:** The objective of this study is to test the association of Occludin, Claudin3 and Claudin5 genes with MS susceptibility, severity, age at onset and disease form.

**Methods:** We performed association tests for MS susceptibility, severity, age at onset and disease form by analyzing 31 SNPs using 1391 cases and 1080 controls from two well-characterized cohorts.

**Results:** Although promising significant results were found in a discovery cohort for MS susceptibility and severity ( $p = 4.889 \times 10^{-5}$  and  $p = 0.0064$  respectively), these findings were not replicated in an independent cohort.

**Conclusions:** We were not able to decisively determine the role of OCLN, CLDN3 and CLDN5 genes in MS. Based on our results, we suggest that studies investigating some “forgotten” genes and analyzing rare variants will need larger cohorts than those currently available, with high quality clinical data.

**Keywords:** Multiple Sclerosis, complex trait, cell adhesion molecules, susceptibility, severity

## INTRODUCTION

Multiple Sclerosis (MS) is a chronic inflammatory disease of the central nervous system. It is widely accepted that MS results from interaction between the genetic background of the host and environmental factors. If environmental factors remain today poorly defined, genome-wide studies such as GWAS have proven successful in identifying genes implicated in MS susceptibility. To date, MHC genetic variants as well as about 100 non-MHC variants, most of them in close vicinity to immunological genes, have been associated with MS susceptibility. However, it has been estimated that the genes spotted to date by GWAS only account for 25% of the genetic component of MS susceptibility<sup>1</sup>. To unravel this missing heritability, some main hypotheses have been put forward such as the contribution of rare variants, genes of individually small effect, interaction between variants from the same biological pathway, the non-identification of causative variants in regions of interest<sup>2, 3</sup>.

Existing GWAS microarrays, are generally designed to cover about 80% of the genome, leading to an uncovered amount of up to 680Mbps, that we named “shadow regions”. This unrepresented part of the genome may be explained by low or null density of probes (SNPs, Single Nucleotide Polymorphisms) targeting these particular regions in commercially available arrays. This could be due to technical limitations of amplification. To explain some of the genetic missing part, we hypothesized that unidentified variants localized in “shadow regions” could have a role in MS susceptibility and severity. Therefore we sought to perform a candidate genes study focusing on genes localized within shadow regions and which could be good candidates because of their known biological function for disease susceptibility. Occludin (OCLN), Claudin-3 (CLDN-3) and Claudin-5 (CLDN-5) coding genes were chosen according to these criteria and our study is the first full investigation of these

genes in MS. These molecules are involved in tight junctions between endothelial cells of the Blood-Brain-Barrier (BBB)<sup>4</sup>. In the MS animal model Experimental Autoimmune Encephalomyelitis (EAE), BBB permeability increase is associated with inflammatory events and is directly correlated with disease severity during the acute phase of EAE<sup>5</sup>. The role of OCLN, CLDN-3 and CLDN-5 in inflammatory cells transmigration has been highlighted in several publications. It has been demonstrated that loss of BBB integrity was associated with a selective loss of Claudin-3<sup>6</sup>. Claudin-5-deficient mice present a severely compromised, leaky BBB<sup>7</sup>. Furthermore, it has been shown that BBB mature endothelial cells may need Occludin to regulate rather than establish their barrier properties<sup>4</sup>.

Based on these arguments, we genotyped 31 SNPs distributed on genes coding for Occludin, Claudin-3 and Claudin-5 and performed association analyses for MS susceptibility and severity. Although promising results were obtained for MS susceptibility and severity in our French MS cohort, we were not able to replicate our findings in an independent American cohort.

## **MATERIALS AND METHODS**

### **Cohorts**

All donors were of Caucasian origin and belong to two well-characterized different cohorts. French individuals used in the discovery phase were provided by the Biological Resources Center REFGENSEP (BRC-REFGENSEP). All French and US cases and controls signed an informed consent to participate in genetic research and the French collection is legally certified by the French Ministry of Research (N° DC-2008-539 and AC-2008-548). Samples used in this study belong to a larger collection of which hundreds of samples were successfully used in the IMSSC GWAS

published in 2011<sup>1</sup>. Furthermore, Cournu-Rebeix et al demonstrated that the collection was valuable for genetic research, and was representative of the natural history of MS<sup>8</sup>. US individuals included in the replication phase belong to the UCSF cohort. The study was approved by the appropriate US Institutional Review Board or Ethics Committee. Clinical information and blood sample were collected under the supervision of a trained physician. Overall 810 French cases and 626 French controls were available for analysis in the discovery phase after standard quality controls. The replication phase included 581 US cases and 454 US controls. Demographic data are presented in Table 1. Severity was assessed by a composite score relying mostly on locomotion and therefore involving only motor impairments due to MS (Extended Disability Severity Score: EDSS)<sup>9</sup> in a version adapted to genetic studies (Multiple Sclerosis Severity Score: MSSS)<sup>10</sup>.

### **SNPs selection**

We selected TagSNPs using the TagSNP Picker tool of HapMap website (<http://hapmap.ncbi.nlm.nih.gov/>). The TagSNP Picker has been applied to each of the 3 genes +/- 10kb in order to capture polymorphisms that could be involved in gene expression regulation. The TagSNP Picker ran on CEU population data from the HapMap release 27, with  $r^2 \geq 0.8$  and MAF (Minor Allele Frequency)  $> 0$ .

Thus, 18 TagSNPs in CLDN5, 4 TagSNPs in CLDN3 and 10 TagSNPs in OCLN were selected. Of the 18 TagSNPs of CLDN5 gene and the 10 TagSNPs of OCLN gene, two and one of them were impossible to design by Applied Biosystems, respectively. Moreover, since SNPs tagging CLDN5 (+/- 10kb) did not belong to CLDN5 gene itself, we decided to test two additional SNPs within the gene.

## **Genotyping**

Genotyping of French samples was performed using the Taqman<sup>®</sup> OpenArray<sup>®</sup> Genotyping Instrument Platform, as recommended by the manufacturer (Applied Biosystems) and genotypes were assigned using the Taqman<sup>®</sup> Genotyper software. Genotyping of UCSF samples was performed using predesigned Taqman SNP genotyping assays. Taqman SNP genotyping assays were conducted in 384-well plates using TaqMan Universal PCR Master Mix on the ABI 7900HT Sequence Detection system using SDS 2.3 software (Applied Biosystems). Standard quality controls were applied to samples and SNPs using Plink v1.07<sup>11</sup>. Criteria were as follows: less than 10% missing genotypes per individual, SNP call-rates superior to 90% and Hardy-Weinberg proportion test  $p > 0.001$  in controls.

## **Statistical analysis**

Statistical analysis was performed using Plink v1.07<sup>11</sup>. Association analysis was performed using Fisher allelic exact test for cases/controls and disease form analysis.

Severity analysis was conducted considering MSSS as a continuous variable or considering extreme groups of MSSS. Patients were divided into 7 groups according to their MSSS. Groups were defined as  $\pm 0.25$ ,  $\pm 0.75$  and  $\pm 1.25$  SD from the mean MSSS. Patients with MSSS higher than  $0.75 \times \text{SD}$  from the mean were considered as belonging to the high MSSS group (named [MSSS+] group), whereas patients with MSSS lower than  $-0.75 \times \text{SD}$  from the mean were considered as belonging to the low MSSS group (named [MSSS-] group). Linear regression was applied when MSSS was considered as a continuous variable and Fisher allelic exact test was applied when the 2 groups of patients were considered. We also performed

linear regression to test association with Age at Onset (AAO).

Corrections for multiple testing were performed using Bonferroni correction as implemented in Plink v.1.07<sup>11</sup>.

The combination of p-values between discovery and replication phases was performed using the basic meta-analysis command as implemented in Plink v.1.07<sup>11</sup>.

### **Statistical power computation**

To compute power of the susceptibility study, we used the Genetic Power Calculator tool<sup>12</sup> with disease prevalence of 0.001, Odds ratio of 2 and high risk allele frequency of 0.996 (Hapmap frequency for rs13447272-G = 0.004, <http://hapmap.ncbi.nlm.nih.gov/>)

## **RESULTS**

All SNPs met the quality control criteria and thus were available for analysis (Table 2). While SNP rs13447272 (CLDN5) reached the threshold for significance for association with MS susceptibility in the French cohort ( $p = 4.889 \times 10^{-5}$ ) (Table 2), this result was not replicated in the UCSF dataset ( $p = 0.4386$ ) (Table 3). We performed meta-analysis combining the 2 datasets with non-significant results ( $p$ -meta = 0.9363). Moreover, SNP rs6971390 (CLDN3) seemed to be associated with MS severity in the French cohort ( $p = 0.0064$  considering MSSS as a continuous variable,  $p = 0.0043$ , considering 2 groups of MS patients, with low or high MSSS). The Materials and Methods section describes the groups definition, and Supplementary Table 1 shows the results for association with MS severity of the 31 genotyped SNPs. However, these findings were not replicated in the UCSF replication dataset whether considering MSSS as a continuous variable ( $p = 0.5657$ )



or comparing the high MSSS and low MSSS groups ( $p = 0.5821$ ). See Table 3.

No significant results for association with disease age at onset or with disease form were obtained (see Supplementary Table 2 and Supplementary Table 3).

## **DISCUSSION**

Considering CLDN3, CLDN5 and OCLN genes as candidates for MS susceptibility and severity, we were not able to identify new genetic variants supporting evidence for a role of these genes in the complex genetic architecture of MS. These results raise two important questions in the genetic dissection of MS: 1) Study of a very low frequency variant probably requires very large dataset and is beyond the reach of current techniques as well and 2) Identification of MS severity variants is complex since it relies on the correlation with a clinical parameter which takes primarily into account only the neurological impairments involved in locomotion. It has indeed been estimated that sample sizes (cases + controls, with equal sized groups) required to obtain a power of 0.8 to detect an allelic odds ratio of 2 at an  $\alpha$  level of  $5 \times 10^{-8}$  increase from 12,000 to 117,000 samples, as the Minor Allele Frequency (MAF) decreases from 0.01 to 0.001, respectively<sup>13</sup>. With the 1391 cases and 1080 controls included in our study, power to detect an allelic odds ratio of 2 for SNP rs13447272 at an  $\alpha$  level of 0.001 was 0.023. Thus, the estimate of the sample size required to have 80% of the power to detect an allelic odds ratio of 2 of the SNP rs13447272 at an  $\alpha$  level of 0.001 is 12,806 cases (and equal size for controls). Our study did not reach the necessary sample size to detect rare variant association and therefore we cannot definitively conclude if the effect seen in the French cohort is a false positive or if the absence of effect seen in the UCSF cohort is a false negative. A larger study including at least 12,000 cases and 12,000 controls would have to be done to reach

a definitive conclusion on the role of CLDN5 gene in MS susceptibility. Based on our results, further studies of rare variants as MS susceptibility candidates should include the appropriate number of individuals in order to identify the missing heritability of MS<sup>2</sup>.

Identification of severity markers is very challenging in MS. Some large and small studies tried to identify such variants but they failed to achieve the threshold for significance<sup>1, 14-17</sup>. To date, only MGAT5 seems to be associated with MS severity in French, Swedish and Italian populations<sup>18</sup>. We failed in our study to identify severity variants in candidate gene analysis, although the results were promising in the French cohort. In our data, the MSSS mean in the French cohort equals 4.46 and the MSSS mean in the UCSF cohort equals 2.93. We performed a mean comparison using the Z-score and concluded that the 2 investigated populations were significantly different according to MSSS (Z-Score = 10.17). This could suggest a real difference of severity between French and American patients due to treatment and different recruitment criteria. Alternatively, it could reflect a difference in evaluation of disabilities by the French and US physicians due to cross-site variation. A severity algorithm based on well-defined criteria to evaluate inflammatory level, aggressiveness of the disease by MRI, cognitive deficits and motor disabilities has to be established in further investigations on genetics of MS severity.

Although GWAS have successfully identified susceptibility markers in complex diseases, there are still parts of heritability that remain unexplained. Our non-conclusive results lead us to suggest that further studies focusing on “shadow regions” to investigate some “forgotten” genes and using rare variants will require larger cohorts with high quality and well defined clinical data.

**Acknowledgments.**

We acknowledge use of the cohort of the CRB-REFGENSEP and the EPIC cohort. We thank ICM, CIC Pitié-Salpêtrière, Généthon, REFGENSEP's members and EPIC cohort's investigators for their help and support as well as Jorge Oksenberg and Sergio Baranzini. VD received a travel grant from the Fondation ARSEP and ICM Carnot Institute. PAG is a recipient of the Nancy Davis Young Investigators.

**Conflict of interest**

The authors declare no conflict of interest.

**Funding**

This study was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Fondation d'Aide pour la Recherche sur la Sclérose En Plaques (ARSEP), the Association Française contre les Myopathies (AFM) and GIS-IBISA. The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06.

## References

1. IMSGC, WTCCC2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011; 476: 214-9.
2. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461: 747-53.
3. Marian AJ. Elements of 'missing heritability'. *Current opinion in cardiology*. 2012; 27: 197-201.
4. Wolburg H and Lippoldt A. Tight junctions of the blood-brain barrier: development, composition and regulation. *Vascul Pharmacol*. 2002; 38: 323-37.
5. Fabis MJ, Scott GS, Kean RB, Koprowski H and Hooper DC. Loss of blood-brain barrier integrity in the spinal cord is common to experimental allergic encephalomyelitis in knockout mouse models. *Proc Natl Acad Sci U S A*. 2007; 104: 5656-61.
6. Wolburg H, Wolburg-Buchholz K, Kraus J, et al. Localization of claudin-3 in tight junctions of the blood-brain barrier is selectively lost during experimental autoimmune encephalomyelitis and human glioblastoma multiforme. *Acta Neuropathol*. 2003; 105: 586-92.
7. Nitta T, Hata M, Gotoh S, et al. Size-selective loosening of the blood-brain barrier in claudin-5-deficient mice. *The Journal of cell biology*. 2003; 161: 653-60.
8. Cournu-Rebeix I, Genin E, Leray E, et al. HLA-DRB1\*15 allele influences the later course of relapsing remitting multiple sclerosis. *Genes Immun*. 2008; 9: 570-4.
9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33: 1444-52.
10. Roxburgh RH, Seaman SR, Masterman T, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology*. 2005; 64: 1144-51.
11. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559-75.
12. Purcell S, Cherny SS and Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003; 19: 149-50.
13. Asimit J and Zeggini E. Rare variant association analysis methods for complex traits. *Annual review of genetics*. 2010; 44: 293-308.
14. Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet*. 2009; 18: 767-78.
15. Jensen CJ, Stankovich J, Van der Walt A, et al. Multiple sclerosis susceptibility-associated SNPs do not influence disease severity measures in a cohort of Australian MS patients. *PLoS One*. 2010; 5: e10003.
16. Gemmati D, Zeri G, Orioli E, et al. Polymorphisms in the genes coding for iron binding and transporting proteins are associated with disability, severity, and early progression in multiple sclerosis. *BMC Med Genet*. 2012; 13: 70.
17. Pauwels I, Cosmans L, Boonen S, Dubois B and Goris A. Rare MEFV variants are not associated with risk to develop multiple sclerosis and severity of disease. *Mult Scler*. 2013.
18. Brynedal B, Wojcik J, Esposito F, et al. MGAT5 alters the severity of multiple sclerosis. *J Neuroimmunol*. 2010; 220: 120-4.

## **Legend.**

### **Table 1. Demographic data of the French cohort (discovery dataset) and UCSF cohort (replication dataset).**

SD: Standard Deviation / RR: Relapsing Remitting / SP: Secondary Progressive / PP: Primary Progressive/ CIS: Clinically Isolated Syndrome

\* Z score = 10.17

### **Table 2. Results of association tests with MS susceptibility in the French discovery dataset for the 31 genotyped SNPs**

SNP rs13447272 (in bold) has reached the significant threshold for replication. Positions are on NCBI B36 assembly.

CHR: Chromosome / MAF: Minor Allele Frequency / P\_HWE: P-value of Hardy-Weinberg Equilibrium test/ OR: Odds Ratio

### **Table 3. Results of association tests with MS susceptibility (top) and MS severity (bottom) in the discovery and the replication datasets.**

MAF: Minor Allele frequency (see Table 2)

MSSS: Multiple Sclerosis Severity Score / Freq.: Frequency / SD: Standard Deviation / OR: Odds Ratio / CI: Confidence Interval (95%)

### **Supplementary Table 1. Results of association test with MS severity in the French discovery dataset for the 31 genotyped SNPs**

SNP rs6971390 (in bold) selected for the replication phase

Positions are on NCBI B36 assembly

MSSS: Multiple Sclerosis Severity Score / Freq.: Frequency / SD: Standard Deviation / OR: Odds Ratio / CI: Confidence Interval (95%)

### **Supplementary Table 2. Results of association test with MS Age at Onset in the French discovery dataset for the 31 genotyped SNPs**

Positions are on NCBI B36 assembly - AAO: Age At Onset / Freq.: Frequency / SD: Standard Deviation / CI: Confidence Interval (95%)

### **Supplementary Table 3. Results of association test with MS form in the French discovery dataset for the 31 genotyped SNPs**

Positions are on NCBI B36 assembly

PP: Primary Progressive form / RR: Relapsing-Remitting form / SP: Secondary Progressive form

Freq.: Frequency / OR: Odds Ratio / CI: Confidence Interval (95%)

	French		UCSF	
	Cases	Controls	Cases	Controls
<b>Total</b>	810	626	581	454
<b>Sex ratio (Female : Male)</b>	2.4 : 1	1.6 : 1	2.2 : 1	2.0 : 1
<b>Age At Onset Mean (SD)</b>	32.27 (10.45)	-	33.60 (9.69)	-
<b>MSSS Mean (SD)</b>	4.46 (2.79)*	-	2.93 (2.31)*	-
<b>Disease Form (RR+SP/PP/CIS)</b>	87%/13%/0%	-	80%/4%/16%	-

**Table 1. Demographic data of the French cohort (discovery dataset) and UCSF cohort (replication dataset).**

Gene	SNP	CHR	Position	Minor Allele A1	Major Allele A2	Call-rate (%)	P_HWE Controls	MAF Cases	MAF Controls	P-Value	OR [CI]
OCLN	rs28616120	5	68821054	G	A	99.79	0.4803	0.2252	0.2160	0.5861	1.06 [0.88;1.26]
OCLN	rs3733742	5	68824017	G	A	97.28	0.8744	0.1692	0.1498	0.1770	1.16 [0.94;1.42]
OCLN	rs34221525	5	68833874	A	G	99.93	1	0.1032	0.1070	0.7589	0.96 [0.76;1.22]
OCLN	rs28633210	5	68838122	C	T	98.96	0.3110	0.2642	0.2741	0.5786	0.95 [0.80;1.12]
OCLN	rs35107257	5	68841372	A	G	99.93	0.2269	0.0346	0.0384	0.6151	0.90 [0.61;1.33]
OCLN	rs28558979	5	68856104	G	T	99.79	0.5939	0.1368	0.1296	0.5805	1.06 [0.86;1.32]
OCLN	rs35628909	5	68860642	G	A	94.29	0.8658	0.1311	0.1403	0.4971	0.92 [0.74;1.15]
OCLN	rs28703202	5	68862544	A	G	97.35	1	0.0532	0.0584	0.5604	0.91 [0.65;1.25]
OCLN	rs256925	5	68887297	A	G	99.44	1	0.0471	0.0362	0.1608	1.31 [0.90;1.92]
CLDN3	rs12536661	7	72819953	A	G	99.16	0.4567	0.2101	0.2018	0.6078	1.05 [0.88;1.26]
CLDN3	rs6971390	7	72824566	A	C	99.72	0.3671	0.1006	0.0989	0.9000	1.02 [0.80;1.31]
CLDN3	rs6460055	7	72825453	C	T	99.72	0.5493	0.2791	0.2772	0.9330	1.01 [0.86;1.19]
CLDN3	rs4484568	7	72825587	A	G	99.23	0.8721	0.4666	0.4992	0.0754	0.87 [0.75;1.01]
CLDN5	rs2283655	22	17880657	G	C	99.86	0.4700	0.4740	0.4672	0.7343	1.03 [0.89;1.19]
<b>CLDN5</b>	<b>rs13447272</b>	<b>22</b>	<b>17881688</b>	<b>A</b>	<b>G</b>	<b>99.44</b>	<b>1</b>	<b>0</b>	<b>0.0096</b>	<b>4.889e-005</b>	<b>- [-;-]</b>
CLDN5	rs13447274	22	17881767	A	G	97.01	0.6576	0.0478	0.0502	0.7907	0.95 [0.67;1.34]
CLDN5	rs1548359	22	17883274	G	C	99.72	0.2140	0.1922	0.2014	0.5377	0.94 [0.78;1.14]
CLDN5	rs13447279	22	17883587	A	G	99.44	1	0.0286	0.0192	0.1137	1.51 [0.91;2.48]
CLDN5	rs2073732	22	17884470	T	C	99.72	0.8526	0.3007	0.3141	0.4620	0.94 [0.80;1.10]
CLDN5	rs13447284	22	17885531	T	C	99.37	0.7624	0.0742	0.0728	0.9424	1.02 [0.77;1.36]
CLDN5	rs9606044	22	17886560	A	T	99.79	0.0300	0.0624	0.0633	0.9382	0.99 [0.73;1.34]
CLDN5	rs9606045	22	17886880	A	G	100	0.6650	0.0488	0.0503	0.8625	0.97 [0.69;1.36]
CLDN5	rs13447298	22	17887447	C	A	97.15	1	0.0267	0.0239	0.7165	1.12 [0.69;1.81]
CLDN5	rs13447300	22	17887929	T	C	100	1	0.0031	0.0016	0.7061	1.94 [0.38;9.99]
CLDN5	rs885984	22	17891008	A	G	99.37	0.6420	0.0485	0.0474	0.9298	1.02 [0.72;1.45]
CLDN5	rs885985	22	17891925	G	A	97.63	0.5636	0.4324	0.4444	0.5385	0.95 [0.82;1.11]
CLDN5	rs2871029	22	17893930	A	G	99.93	0.0029	0.1360	0.1414	0.7028	0.96 [0.77;1.18]
CLDN5	rs739371	22	17893998	C	G	99.58	0.0026	0.2293	0.2375	0.6236	0.96 [0.80;1.14]
CLDN5	rs9604969	22	17894013	G	A	99.09	0.7759	0.0831	0.0770	0.5784	1.09 [0.83;1.43]
CLDN5	rs5748261	22	17898159	G	C	96.24	1	0.0173	0.0216	0.4847	0.80 [0.46;1.38]
CLDN5	rs7290191	22	17902510	G	T	98.68	0.1697	0.1352	0.1359	1	0.99 [0.80;1.24]

**Table 2. Results of association tests with MS susceptibility in the French discovery dataset for the 31 genotyped SNPs**

### ***Susceptibility***

<i>Gene</i>	<i>SNP</i>	<u>Discovery - French dataset</u>			<u>Replication - UCSF dataset</u>			<u>Meta-analysis</u>	
		<i>MAF Cases</i> <i>N=810</i>	<i>MAF Controls</i> <i>N=626</i>	<i>P-Value</i>	<i>MAF Cases</i> <i>N=581</i>	<i>MAF Controls</i> <i>N=454</i>	<i>P-Value</i>	<i>P-Value</i>	<i>OR</i>
CLDN5	rs13447272	0	0.0096	4.889e-005	0	0.0011	0.4386	0.9363	0.9924

### ***Severity***

#### *Continuous variable*

<i>Gene</i>	<i>SNP</i>	<u>Discovery - French dataset</u>		<u>Replication - UCSF dataset</u>	
		<i>P-Value</i>	<i>BETA [CI]</i>	<i>P-Value</i>	<i>BETA [CI]</i>
CLDN3	rs6971390	0.0064	-0.72 [-1.24;-0.20]	0.5657	-0.13 [-0.59;0.32]

#### *[MSSS+] vs [MSSS-] Groups*

<i>Gene</i>	<i>SNP</i>	<u>Discovery - French dataset</u>			<u>Replication - UCSF dataset</u>			<u>Meta-analysis</u>	
		<i>A1 Freq. in [MSSS +] group</i> <i>N=182</i>	<i>A1 Freq. in [MSSS -] group</i> <i>N=197</i>	<i>P-Value</i>	<i>A1 Freq. in [MSSS +] group</i> <i>N=131</i>	<i>A1 Freq. in [MSSS -] group</i> <i>N=176</i>	<i>P-Value</i>	<i>P-Value</i>	<i>OR</i>
CLDN3	rs6971390	0.0714	0.1371	0.0043	0.08779	0.1023	0.5821	0.0106	0.62

**Table 3. Results of association tests with MS susceptibility (top) and MS severity (bottom) in the discovery and the replication datasets.**



Gene	SNP	CHR	Position	A1	A2	MSSS as a Continuous Variable								[MSSS +] vs [MSSS -] Groups			
						A1A1 Freq.	A1A2 Freq.	A2A2 Freq.	A1A1 MSSS Mean (SD)	A1A2 MSSS Mean (SD)	A2A2 MSSS Mean (SD)	P-Value	BETA [CI]	A1 Freq. in [MSSS +] group	A1 Freq. in [MSSS -] group	P-Value	OR [CI]
OCLN	rs28616120	5	68821054	G	A	0.0495	0.3493	0.6012	4.10 (2.28)	4.37 (2.91)	4.54 (2.76)	0.3065	-0.19 [-0.55;0.17]	0.2044	0.2372	0.2930	0.83 [0.58;1.17]
OCLN	rs3733742	5	68824017	G	A	0.0324	0.2743	0.6934	4.00 (2.38)	4.39 (2.86)	4.53 (2.80)	0.3624	-0.19 [-0.59;0.21]	0.1469	0.1797	0.2344	0.79 [0.53;1.17]
OCLN	rs34221525	5	68833874	A	G	0.0150	0.1704	0.8146	4.10 (2.57)	4.61 (2.76)	4.44 (2.81)	0.7626	0.07 [-0.41;0.56]	0.1016	0.1041	1	0.97 [0.61;1.56]
OCLN	rs28633210	5	68838122	C	T	0.0659	0.3892	0.5449	4.32 (2.48)	4.48 (2.88)	4.48 (2.77)	0.8199	-0.04 [-0.38;0.30]	0.2527	0.2704	0.6196	0.91 [0.66;1.26]
OCLN	rs35107257	5	68841372	A	G	0.0045	0.0628	0.9327	4.01 (4.55)	4.38 (2.86)	4.47 (2.79)	0.7593	-0.12 [-0.89;0.65]	0.0385	0.0482	0.5944	0.79 [0.39;1.60]
OCLN	rs28558979	5	68856104	G	T	0.0240	0.2249	0.7511	3.88 (2.39)	4.45 (2.88)	4.49 (2.78)	0.5516	-0.13 [-0.56;0.30]	0.1236	0.1403	0.5206	0.86 [0.57;1.32]
OCLN	rs35628909	5	68860642	G	A	0.0224	0.2112	0.7664	4.76 (2.95)	4.52 (2.73)	4.42 (2.79)	0.6011	0.12 [-0.33;0.57]	0.1302	0.1317	1	0.99 [0.64;1.53]
OCLN	rs28703202	5	68862544	A	G	0.0015	0.1041	0.8943	8.64 (0)	3.95 (2.67)	4.52 (2.79)	0.2260	-0.42 [-1.10;0.26]	0.0393	0.0628	0.1825	0.61 [0.31;1.20]
OCLN	rs256925	5	68887297	A	G	0	0.0975	0.9025	- (-)	4.13 (2.83)	4.51 (2.78)	0.2954	-0.38 [-1.09;0.33]	0.0440	0.0513	0.7333	0.85 [0.43;1.67]
CLDN3	rs12536661	7	72819953	A	G	0.0347	0.3399	0.6254	3.48 (2.98)	4.25 (2.72)	4.63 (2.81)	0.02084	-0.45 [-0.83;-0.07]	0.1722	0.2345	0.0371	0.68 [0.47;0.97]
<b>CLDN3</b>	<b>rs6971390</b>	<b>7</b>	<b>72824566</b>	<b>A</b>	<b>C</b>	<b>0.0045</b>	<b>0.1868</b>	<b>0.8087</b>	<b>2.84 (4.27)</b>	<b>3.90 (2.66)</b>	<b>4.60 (2.80)</b>	<b>0.0064</b>	<b>-0.72 [-1.24;-0.20]</b>	<b>0.0714</b>	<b>0.1371</b>	<b>0.0043</b>	<b>0.48 [0.30;0.79]</b>
CLDN3	rs6460055	7	72825453	C	T	0.0735	0.4003	0.5262	3.88 (2.87)	4.29 (2.72)	4.67 (2.81)	0.0228	-0.39 [-0.73;-0.06]	0.2348	0.3061	0.0331	0.70 [0.50;0.96]
CLDN3	rs4484568	7	72825587	A	G	0.2335	0.4775	0.2889	4.77 (2.84)	4.37 (2.69)	4.37 (2.91)	0.2066	0.19 [-0.10;0.48]	0.5110	0.4362	0.0415	1.35 [1.01;1.80]
CLDN5	rs2283655	22	17880657	G	C	0.2365	0.4895	0.2740	4.30 (2.83)	4.55 (2.75)	4.45 (2.86)	0.6466	-0.07 [-0.37;0.23]	0.4615	0.5000	0.3087	0.86 [0.64;1.14]
CLDN5	rs13447272	22	17881688	A	G	0	0	1	- (-)	- (-)	4.45 (2.79)	-	- [-;-]	-	-	-	- [-;-]
CLDN5	rs13447274	22	17881767	A	G	0.0015	0.0878	0.9106	2.34 (0)	4.59 (2.54)	4.48 (2.82)	0.9505	0.02 [-0.71;0.76]	0.0393	0.0370	1	1.06 [0.50;2.27]
CLDN5	rs1548359	22	17883274	G	C	0.0344	0.2829	0.6826	4.07 (2.20)	4.41 (2.80)	4.49 (2.82)	0.4965	-0.13 [-0.52;0.25]	0.1602	0.1675	0.8443	0.95 [0.64;1.39]
CLDN5	rs13447279	22	17883587	A	G	0.0015	0.0572	0.9413	2.33 (0)	4.73 (2.77)	4.44 (2.80)	0.7299	0.15 [-0.72;1.03]	0.0276	0.0279	1	0.99 [0.42;2.36]
CLDN5	rs2073732	22	17884470	T	C	0.0915	0.3838	0.5247	4.37 (2.63)	4.55 (2.87)	4.41 (2.76)	0.7987	0.04 [-0.28;0.37]	0.2790	0.2730	0.8707	1.03 [0.75;1.42]
CLDN5	rs13447284	22	17885531	T	C	0.0045	0.1392	0.8563	2.35 (2.66)	4.95 (3.06)	4.40 (2.74)	0.2260	0.36 [-0.22;0.93]	0.0907	0.0791	0.6024	1.16 [0.70;1.94]
CLDN5	rs9606044	22	17886560	A	T	0.0090	0.1183	0.8728	5.96 (2.54)	4.85 (2.87)	4.39 (2.78)	0.0649	0.54 [-0.03;1.12]	0.0912	0.0533	0.0481	1.78 [1.01;3.14]
CLDN5	rs9606045	22	17886880	A	G	0.0075	0.0927	0.8999	5.30 (2.19)	4.56 (2.94)	4.44 (2.78)	0.5406	0.02 [-0.44;0.83]	0.0632	0.0533	0.6416	1.20 [0.65;2.20]
CLDN5	rs13447298	22	17887447	C	A	0.0015	0.0553	0.9432	2.29 (0)	4.10 (2.92)	4.51 (2.79)	0.2908	-0.48 [-1.37;0.41]	0.0253	0.0393	0.3069	0.63 [0.27;1.47]
CLDN5	rs13447300	22	17887929	T	C	0	0.0045	0.9955	- (-)	6.95 (2.71)	4.45 (2.79)	0.1224	2.50 [-0.67;5.66]	0.0055	0	0.2303	- [-;-]
CLDN5	rs885984	22	17891008	A	G	0.006	0.0932	0.9008	5.13 (2.49)	4.59 (2.94)	4.45 (2.78)	0.5767	0.19 [-0.47;0.84]	0.0635	0.0513	0.5307	1.26 [0.68;2.33]
CLDN5	rs885985	22	17891925	G	A	0.1866	0.4659	0.3475	4.31 (2.75)	4.61 (2.80)	4.39 (2.82)	0.9941	-0.0011 [-0.3006;0.2983]	0.4222	0.4016	0.6024	1.09 [0.81;1.46]
CLDN5	rs2871029	22	17893930	A	G	0.0180	0.2395	0.7425	4.61 (2.68)	4.39 (2.76)	4.49 (2.81)	0.7849	-0.06 [-0.50;0.38]	0.1429	0.1250	0.5213	1.17 [0.77;1.77]
CLDN5	rs739371	22	17893998	C	G	0.0569	0.3458	0.5973	4.21 (3.07)	4.73 (2.83)	4.32 (2.74)	0.3072	0.18 [-0.17;0.54]	0.2624	0.2183	0.1724	1.27 [0.91;1.78]
CLDN5	rs9604969	22	17894013	G	A	0.0120	0.1557	0.8323	4.49 (1.94)	4.37 (2.94)	4.49 (2.77)	0.7421	-0.09 [-0.60;0.43]	0.0934	0.0995	0.8063	0.93 [0.57;1.51]
CLDN5	rs5748261	22	17898159	G	C	0	0.0356	0.9644	- (-)	5.30 (3.07)	4.46 (2.78)	0.1572	0.84 [-0.32;2.00]	0.0281	0.0106	0.1070	2.69 [0.84;8.65]
CLDN5	rs7290191	22	17902510	G	T	0.0228	0.2264	0.7508	4.59 (2.79)	4.68 (2.65)	4.41 (2.83)	0.3348	0.21 [-0.22;0.64]	0.1528	0.1146	0.1317	1.39 [0.91;2.13]

Supplementary Table 1. Results of association test with MS severity in the French discovery dataset for the 31 genotyped SNPs

Gene	SNP	CHR	Position	A1	A2	A1A1 Freq.	A1A2 Freq.	A2A2 Freq.	A1A1 AAO Mean (SD)	A1A2 AAO Mean (SD)	A2A2 AAO Mean (SD)	P-Value	BETA [CI]
OCLN	rs28616120	5	68821054	G	A	0.0509	0.3481	0.6011	31.21 (9.58)	32.72 (10.53)	32.13 (10.51)	0.86	0.11 [-1.16;1.39]
OCLN	rs3733742	5	68824017	G	A	0.0303	0.2796	0.6901	31.50 (10.52)	32.45 (10.48)	32.41 (10.50)	0.8548	-0.13 [-1.56;1.30]
OCLN	rs34221525	5	68833874	A	G	0.0147	0.1725	0.8128	32.09 (7.66)	32.24 (10.64)	32.29 (10.48)	0.9388	-0.07 [-1.79;1.65]
OCLN	rs28633210	5	68838122	C	T	0.0710	0.3815	0.5475	30.55 (8.50)	32.37 (10.47)	32.46 (10.68)	0.3621	-0.56 [-1.76;0.64]
OCLN	rs35107257	5	68841372	A	G	0.0040	0.0641	0.9319	36.67 (6.66)	30.56 (9.58)	32.37 (10.52)	0.4768	-0.99 [-3.73;1.74]
OCLN	rs28558979	5	68856104	G	T	0.0228	0.2289	0.7483	29.76 (8.52)	31.92 (10.36)	32.48 (10.55)	0.2898	-0.82 [-2.34;0.70]
OCLN	rs35628909	5	68860642	G	A	0.0215	0.2132	0.7654	34.07 (8.23)	31.98 (10.30)	32.26 (10.56)	0.9049	0.1 [-1.51;1.70]
OCLN	rs28703202	5	68862544	A	G	0.0014	0.1015	0.8971	32 (0)	31.84 (10.25)	32.35 (10.37)	0.6905	-0.49 [-2.92;1.93]
OCLN	rs256925	5	68887297	A	G	0	0.0965	0.9035	NA (NA)	32.21 (11.47)	32.31 (10.36)	0.9385	-0.10 [-2.65;2.45]
CLDN3	rs12536661	7	72819953	A	G	0.0337	0.3401	0.6262	31.04 (9.07)	32.16 (10.13)	32.49 (10.74)	0.5039	-0.46 [-1.82;0.89]
CLDN3	rs6971390	7	72824566	A	C	0.0040	0.1883	0.8077	28.67 (8.51)	32.44 (10.16)	32.25 (10.54)	0.9879	0.01 [-1.83;1.86]
CLDN3	rs6460055	7	72825453	C	T	0.0736	0.3976	0.5288	30.85 (9.76)	31.92 (10.32)	32.77 (10.66)	0.1351	-0.91 [-2.10;0.28]
CLDN3	rs4484568	7	72825587	A	G	0.2369	0.4712	0.2918	32.77 (11.13)	32.26 (10.16)	31.87 (10.46)	0.3939	0.45 [-0.59;1.49]
CLDN5	rs2283655	22	17880657	G	C	0.2246	0.5	0.2754	32.92 (10.53)	32.38 (10.36)	31.59 (10.60)	0.216	0.67 [-0.39;1.73]
CLDN5	rs13447272	22	17881688	A	G	0	0	1	NA (NA)	NA (NA)	32.25 (10.42)	-	- [-;-]
CLDN5	rs13447274	22	17881767	A	G	0.0014	0.0966	0.9021	48 (0)	31.16 (10.82)	32.44 (10.42)	0.5491	-0.77 [-3.28;1.74]
CLDN5	rs1548359	22	17883274	G	C	0.0388	0.3008	0.6604	30.62 (8.95)	31.9 (10.77)	32.54 (10.41)	0.2615	-0.77 [-2.11;0.57]
CLDN5	rs13447279	22	17883587	A	G	0.0013	0.0566	0.942	48 (0)	31.21 (8.62)	32.35 (10.54)	0.8351	-0.33 [-3.44;2.78]
CLDN5	rs2073732	22	17884470	T	C	0.0937	0.4016	0.5047	30.73 (9.46)	32.16 (10.65)	32.70 (10.47)	0.1637	-0.81 [-1.96;0.33]
CLDN5	rs13447284	22	17885531	T	C	0.0040	0.1350	0.8610	31 (1)	32 (10.74)	32.34 (10.44)	0.7192	-0.38 [-2.46;1.69]
CLDN5	rs9606044	22	17886560	A	T	0.0080	0.1123	0.8797	40.33 (5.99)	33.42 (10.85)	32.02 (10.38)	0.05535	2.04 [-0.04;4.13]
CLDN5	rs9606045	22	17886880	A	G	0.0067	0.0881	0.9052	39 (5.61)	34.23 (10.80)	32.03 (10.43)	0.03388	2.51 [0.20;4.82]
CLDN5	rs13447298	22	17887447	C	A	0.0014	0.0523	0.9464	18 (0)	29.55 (11.41)	32.58 (10.38)	0.03755	-3.45 [-6.69;-0.20]
CLDN5	rs13447300	22	17887929	T	C	0	0.0067	0.9933	NA (NA)	35.20 (15.59)	32.25 (10.43)	0.5305	2.95 [-6.26;12.15]
CLDN5	rs885984	22	17891008	A	G	0.0054	0.0886	0.9060	39.75 (6.19)	34.29 (10.74)	32.04 (10.45)	0.03372	2.57 [0.20;4.95]
CLDN5	rs885985	22	17891925	G	A	0.1911	0.4797	0.3293	31.09 (10.12)	32.24 (10.66)	32.84 (10.32)	0.1224	-0.84 [-1.90;0.22]
CLDN5	rs2871029	22	17893930	A	G	0.0187	0.2393	0.7420	31.71 (9.60)	31.63 (10.84)	32.50 (10.37)	0.3439	-0.74 [-2.28;0.79]
CLDN5	rs739371	22	17893998	C	G	0.0548	0.3463	0.5989	33.2 (11.75)	31.47 (10.13)	32.64 (10.52)	0.4523	-0.48 [-1.74;0.77]
CLDN5	rs9604969	22	17894013	G	A	0.0107	0.1486	0.8407	35 (9.53)	32.63 (11.25)	32.23 (10.30)	0.503	0.63 [-1.22;2.49]
CLDN5	rs5748261	22	17898159	G	C	0	0.0333	0.9667	NA (NA)	33.62 (10.39)	32.31 (10.45)	0.5455	1.31 [-2.94;5.56]
CLDN5	rs7290191	22	17902510	G	T	0.0204	0.2343	0.7452	26.13 (10.76)	32.43 (10.18)	32.43 (10.50)	0.2421	-0.92 [-2.46;0.62]

**Supplementary Table 2. Results of association test with MS Age at Onset in the French discovery dataset for the 31 genotyped SNPs**

<i>Gene</i>	<i>SNP</i>	<i>CHR</i>	<i>Position</i>	<i>A1</i>	<i>A2</i>	<i>A1 Freq. in [PP]</i>	<i>A1 Freq. in [RR+SP]</i>	<i>P-Value</i>	<i>OR [CI]</i>
OCLN	rs28616120	5	68821054	G	A	0.2143	0.2278	0.7145	0.92 [0.64;1.33]
OCLN	rs3733742	5	68824017	G	A	0.1701	0.1709	1	0.99 [0.66;1.49]
OCLN	rs34221525	5	68833874	A	G	0.0816	0.1048	0.3750	0.76 [0.44;1.31]
OCLN	rs28633210	5	68838122	C	T	0.2347	0.2667	0.3837	0.84 [0.59;1.2]
OCLN	rs35107257	5	68841372	A	G	0.0306	0.0364	0.8372	0.84 [0.35;1.98]
OCLN	rs28558979	5	68856104	G	T	0.1378	0.1376	1	1 [0.65;1.55]
OCLN	rs35628909	5	68860642	G	A	0.0978	0.1331	0.1951	0.71 [0.42;1.18]
OCLN	rs28703202	5	68862544	A	G	0.0372	0.0549	0.3828	0.67 [0.3;1.47]
OCLN	rs256925	5	68887297	A	G	0.0306	0.0506	0.2818	0.59 [0.25;1.39]
CLDN3	rs12536661	7	72819953	A	G	0.1753	0.2085	0.3388	0.81 [0.54;1.2]
CLDN3	rs6971390	7	72824566	A	C	0.0612	0.1031	0.0704	0.57 [0.31;1.05]
CLDN3	rs6460055	7	72825453	C	T	0.2398	0.2792	0.2647	0.81 [0.57;1.16]
CLDN3	rs4484568	7	72825587	A	G	0.4592	0.4743	0.7017	0.94 [0.7;1.27]
CLDN5	rs2283655	22	17880657	G	C	0.5051	0.4682	0.3571	1.16 [0.86;1.57]
CLDN5	rs13447272	22	17881688	O	G	0	0	1	- [-;-]
CLDN5	rs13447274	22	17881767	A	G	0.0521	0.0490	0.8577	1.07 [0.54;2.12]
CLDN5	rs1548359	22	17883274	G	C	0.1633	0.1941	0.3297	0.81 [0.54;1.21]
CLDN5	rs13447279	22	17883587	A	G	0.0155	0.0321	0.2611	0.47 [0.15;1.55]
CLDN5	rs2073732	22	17884470	T	C	0.2398	0.3048	0.0650	0.72 [0.51;1.02]
CLDN5	rs13447284	22	17885531	T	C	0.0612	0.0738	0.6563	0.82 [0.44;1.52]
CLDN5	rs9606044	22	17886560	A	T	0.0722	0.0636	0.6388	1.15 [0.64;2.06]
CLDN5	rs9606045	22	17886880	A	G	0.0510	0.0512	1	1 [0.5;1.97]
CLDN5	rs13447298	22	17887447	C	A	0.0206	0.0280	0.8110	0.73 [0.26;2.08]
CLDN5	rs13447300	22	17887929	T	C	0.0102	0.0023	0.1322	4.42 [0.73;26.64]
CLDN5	rs885984	22	17891008	A	G	0.0510	0.0499	0.8620	1.02 [0.52;2.03]
CLDN5	rs885985	22	17891925	G	A	0.3969	0.4371	0.3122	0.85 [0.62;1.15]
CLDN5	rs2871029	22	17893930	A	G	0.1531	0.1359	0.5059	1.15 [0.76;1.75]
CLDN5	rs739371	22	17893998	C	G	0.2500	0.2252	0.4650	1.15 [0.81;1.63]
CLDN5	rs9604969	22	17894013	G	A	0.0816	0.0855	1	0.95 [0.55;1.64]
CLDN5	rs5748261	22	17898159	G	C	0.0316	0.0145	0.1194	2.21 [0.87;5.64]
CLDN5	rs7290191	22	17902510	G	T	0.1250	0.1384	0.7350	0.89 [0.56;1.40]

**Supplementary Table 3. Results of association test with MS form in the French discovery dataset for the 31 genotyped SNPs**

## 2.4. Hypothèse 4 : Variants rares et peu fréquents

Jusqu'à présent, les technologies utilisées dans les GWAS ne permettent d'analyser que les SNPs dits fréquents, c'est à dire ayant une MAF (Minor Allele Frequency, i.e. Fréquence d'Allèle Mineur) supérieure à 5% dans la population générale. Cependant, il est probable que des SNPs ayant une MAF inférieure à 5% soient aussi impliqués dans la prédisposition à la SEP (cf. hypothèse d'hétérogénéité développée dans le 1.2.2.) (Manolio, Collins et al. 2009).

Grâce au projet 1000 Genomes (Genomes Project, Abecasis et al. 2010, Genomes Project, Abecasis et al. 2012), de nombreux SNPs ayant une MAF faible, c'est à dire comprise entre 0,5% et 5% dans la population générale, et des variants rares, c'est à dire ayant une MAF inférieure à 0,5% dans la population générale, ont été identifiés.

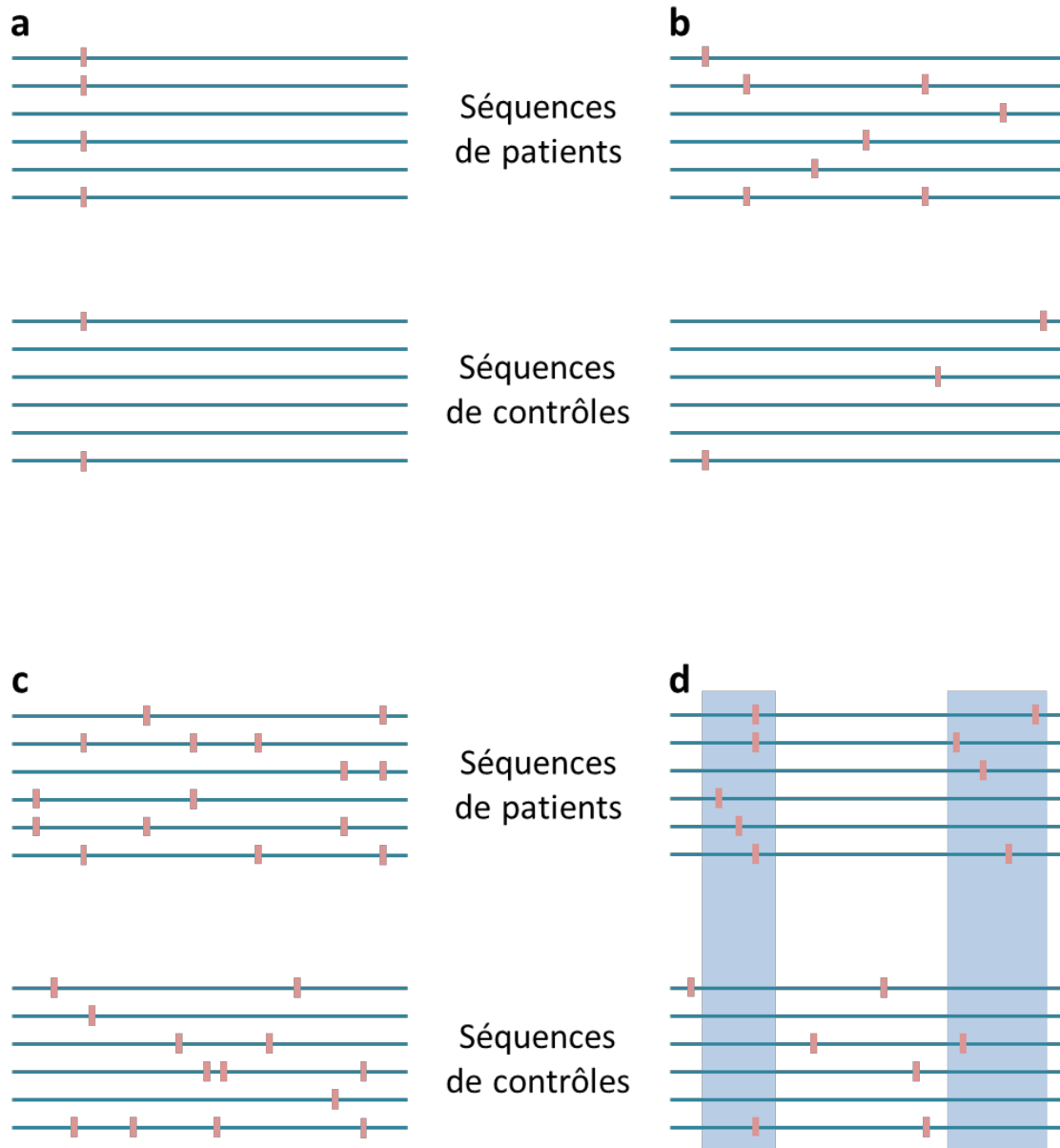
### 2.4.1. Arguments

Le rôle que pourrait avoir les variants rares dans la prédisposition aux maladies complexes est souligné par différents arguments (Gibson 2011) :

- Le plus fort d'entre eux repose sur la théorie de l'évolution. Etant donné que les maladies sont délétères pour la survie des individus, les variants associés devraient disparaître par la sélection naturelle. Ils devraient par conséquent ne pas être fréquents en population, autrement dit, ils devraient être rares. Leur existence serait due à une balance entre les mutations créant de nouveaux variants de prédisposition et la sélection les empêchant de dériver à une plus forte fréquence en population.

- Les données de génétique des populations montrent que les variants délétères sont rares. Il est en effet connu que la distribution des fréquences des variants est biaisée vers un excès de variants rares : plus d'un tiers des polymorphismes ont une fréquence inférieure à 5% en population. Les découvertes faites suite aux données de séquences de whole-exome, montrant que les substitutions non-synonymes sont majoritairement présentes à faible fréquence, reflètent certainement le rôle de sauvegarde de la sélection. Les substitutions semblant délétères, la sélection tend à les garder à de faibles fréquences.

- Le dernier argument repose sur le fait que la plupart des maladies familiales sont dues à des allèles rares à forts effets. Cela s'applique en majorité aux maladies mendéliennes causées par de rares mutations à forte pénétrance mais il existe également des maladies plus fréquentes causées par des variants rares, telles que le cancer du sein avec les mutations rares des gènes *BRCA1* et *BRCA2*.



**Figure 18. Scénarii selon lesquels les variants génétiques permettent de distinguer les cas des contrôles.**

La ligne bleue représente des régions génomiques ; les rectangles rouges représentent les variants.

- (a) A un seul locus, un variant fréquent est plus présent chez les patients que chez les contrôles.
- (b) Plusieurs variants rares contribuent au phénotype de telle sorte que la fréquence collective des variants rares est plus importante chez les patients.
- (c) Plusieurs variants rares contribuent au phénotype mais agissent de manière coordonnée de telle sorte que les patients sont plus à même d’avoir des séquences similaires comparées aux contrôles.
- (d) Plusieurs variants rares contribuent au phénotype mais les variations sont positionnées dans des régions génomiques spécifiques.

De (Bansal, Libiger et al. 2010).

### 2.4.2. Approches expérimentales pour l'étude des variants rares

Deux stratégies sont possibles pour étudier ces variants (Bansal, Libiger et al. 2010, Metzker 2010):

- La première, sans à priori, consiste à étudier les variants rares ou à faible MAF présents dans les exons grâce à de nouvelles puces à ADN: les ExomeChip. Ces études requièrent des échantillons de taille importante. Les estimations du nombre d'échantillons (avec autant de cas et de contrôles) nécessaires pour atteindre une puissance de 80% afin de détecter un OR de 2 à une p-value significative *genome-wide* de  $5 \times 10^{-8}$  augmentent de 12 000 à 117 000 échantillons, lorsque la MAF diminue respectivement de 0,01 à 0,001 (Asimit and Zeggini 2010).

- La deuxième, avec à priori, consiste à utiliser le séquençage pour identifier des variants rares se trouvant dans ou à proximité des régions identifiées comme associées à la maladie par les GWAS.

### 2.4.3. Hypothèses sur le rôle des variants rares dans la prédisposition génétique

Comment les variants rares pourraient-ils jouer un rôle dans la prédisposition aux maladies complexes ? La **Figure 18** reprend les hypothèses les plus probables (Bansal, Libiger et al. 2010) :

- un variant rare pourrait influencer la maladie indépendamment des autres variants
- l'interaction entre des variants rares et des variants communs ou peu fréquents serait à l'origine de la maladie
- quelques variants rares influenceraient un phénotype déterminé par leur localisation dans un élément fonctionnel d'une région d'intérêt

L'hypothèse la plus étudiée est l'Hétérogénéité Allélique Extrême (HEA ou Extreme Allelic Heterogeneity, EAH), dans laquelle un individu ou des petits sous-groupes d'individus présentant un phénotype particulier de la maladie possède un ou quelques sous-ensemble(s) d'un groupe plus large de variants rares qui tous indépendamment les uns des autres ne perturbent qu'un seul gène d'une manière similaire (McClellan and King 2010).



#### 2.4.4. Limites

Avec l'étude des variants rares se pose aussi le problème des méthodes statistiques à utiliser.

La plus simple d'entre elles consiste à tester de manière indépendante les variants rares pour leur association avec la maladie. Cependant, comme mentionné précédemment, cette méthode nécessite des échantillons de grande taille (de 12.000 à 117.000 patients) pour atteindre des puissances statistiques suffisantes.

Une solution permettant de passer outre ce problème de puissance serait d'étudier l'association de sous-groupes de variants. La stratification des variants en sous-groupes est alors basée sur les fonctions des gènes et /ou sur la prédiction du risque plus ou moins délétère que pourrait conférer chaque variant. Ainsi, dans sa forme la plus simple, cette stratégie reposerait sur le fait de compter le nombre d'individus portant un variant rare à n'importe quelle position dans la région génomique d'intérêt ou dans le groupe de régions d'intérêt, puis de calculer la fréquence de ces individus chez les patients et les contrôles et enfin de tester la différence de fréquences entre ces deux groupes (Bansal, Libiger et al. 2010).

**Au cours de mon doctorat, je me suis intéressé à l'hypothèse des variants rares par le biais d'une étude de l'IMSGC, le projet ExomeChip. Ce projet vise à génotyper 243.094 variants dans une cohorte de 16.762 patients et de 10.397 contrôles. Les variants analysés étant rares, l'analyse individuelle des SNPs n'aurait que peu de puissance statistique. C'est pourquoi la stratégie d'analyse sera probablement d'étudier la charge génétique en variants rares au sein d'un gène ou d'un pathway. L'orientation stratégique de l'analyse des résultats est en cours de discussion au sein du consortium.**

**Dans le cadre de ce projet, mon rôle a été de génotyper 400 patients et 200 contrôles français en utilisant la puce ExomeChip créée par Illumina et d'effectuer les premiers contrôles qualités. J'ai aussi été le représentant de l'équipe au cours des différents échanges qu'il y eu jusqu'à maintenant.**

**Ce projet est en cours et n'a pour le moment pas atteint la phase finale d'analyse. Il ne constituera donc pas une partie à part entière de ce mémoire.**





## **2.5. Hypothèse 5 : L'interaction de variants d'un même réseau biologique confère un risque de développer la maladie**

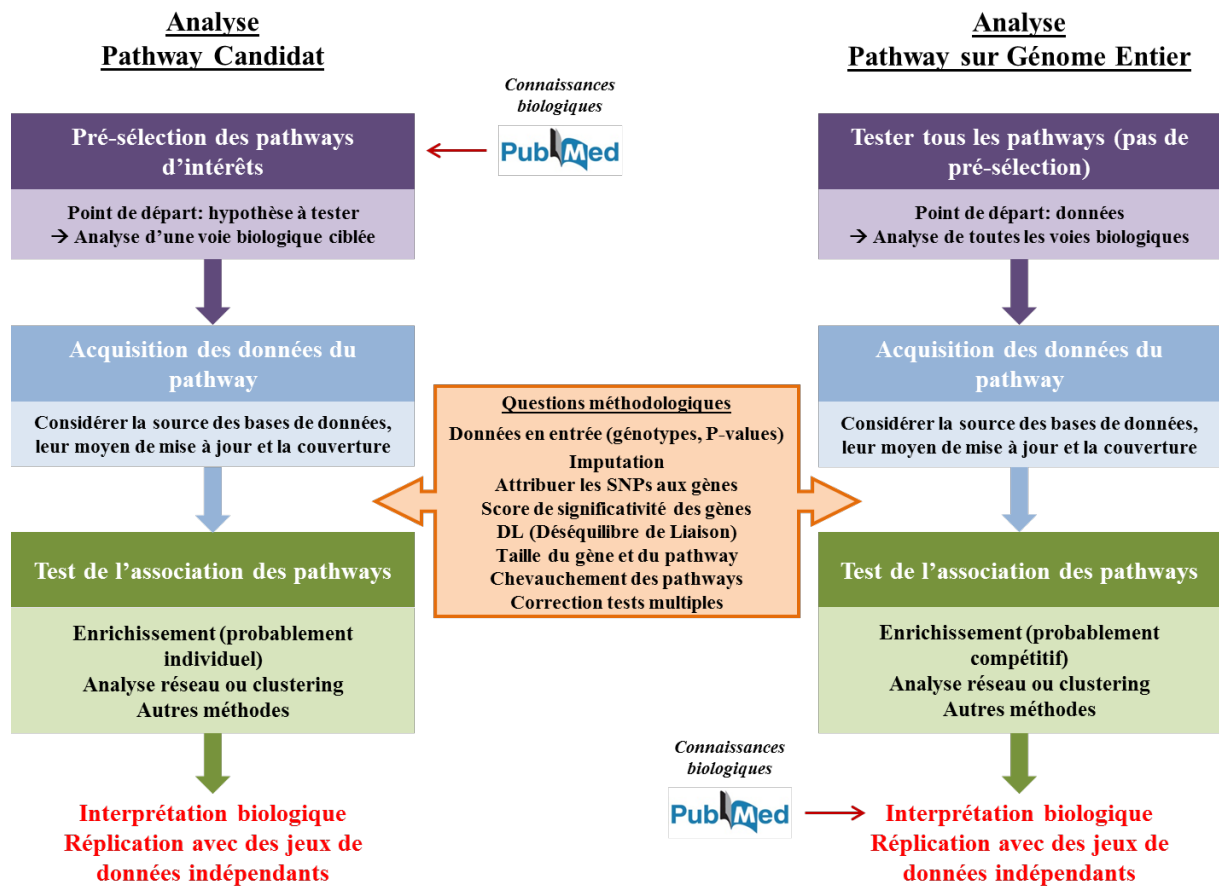
Les études de criblage anonyme du génome ont permis l'identification de régions de prédisposition de différentes maladies complexes comme la SEP. Les analyses statistiques utilisées dans ces projets pan-génomiques ont consisté à tester l'association individuelle de chaque variant avec la maladie (Wang, Li et al. 2010). Cependant, cette stratégie a une puissance limitée pour identifier le rôle des gènes ayant un effet individuel très faible ou nul. Néanmoins, il est probable que parmi ces variants à faible effet individuel, il en existe certains qui, en interaction, peuvent conférer un risque de prédisposition aux maladies. De plus, même si un certain nombre de gènes ont été identifiés par les GWAS dans différentes maladies, très peu d'entre eux peuvent être reliés à des mécanismes biologiques expliquant la physiopathologie des maladies. Enfin, même s'il est primordial d'identifier les gènes de prédisposition aux maladies, il est également nécessaire d'étudier et d'identifier les *pathways* ou les réseaux biologiques pouvant être impliqués dans leur développement (Ramanan, Shen et al. 2012).

Il a déjà été montré dans certaines pathologies, comme la maladie de Crohn ou le diabète de type 2, que des gènes impliqués dans la même fonction biologique influencent collectivement la prédisposition à la maladie, même si les variants de ces gènes considérés de manière individuelle ne confèrent aucun risque génétique (Wang, Li et al. 2010, Zhong, Yang et al. 2010).

Ainsi, les analyses des réseaux biologiques (ici analyses *pathways*) sont aujourd'hui le lien indispensable à identifier entre les données génétiques de prédisposition et les phénotypes cliniques des patients. Même si les analyses *pathways* sont aujourd'hui très populaires, les stratégies d'analyses ont progressé ces dernières années sans véritables lignes directrices (Ramanan, Shen et al. 2012).

### 2.5.1. Définition d'une analyse pathway

Dans une analyse *pathway*, un groupe de gènes impliqués dans une même voie biologique, le *pathway*, est testé pour l'enrichissement en signaux d'association avec la maladie. Cette analyse permet de répondre à la question suivante : ***Le groupe de gènes montre-t-il significativement plus d'association que ce à quoi on s'attendrait sous l'hypothèse nulle ?*** (Holmans 2010)



**Figure 19. Guide pour les analyses pathways.**

La figure ci-dessus reprend les grandes étapes et les questions méthodologiques qui se posent lors des études Pathways Candidats et les études Pathway sur Génome Entier.

Adapté de (Ramanan, Shen et al. 2012).

Deux types d'analyse *pathway* peuvent être effectuées suivant l'hypothèse nulle qui est testée (Holmans 2010, Wang, Li et al. 2010) :

- les **tests compétitifs** ou *competitive tests* comparent l'association entre un groupe de gènes et la maladie par rapport à l'association entre tous les autres groupes de gènes testés et la maladie
- les **tests individuels** ou *self-contained tests* recherchent une association significative entre le groupe de gènes et la maladie, sans faire de comparaison avec les autres groupes de gènes testés.

### 2.5.2. Les stratégies d'études *pathway*

Comme pour les analyses d'identification de gènes de prédisposition, deux stratégies peuvent être utilisées lors d'analyse *pathways* (Ramanan, Shen et al. 2012) (**Figure 19**):

- les analyses ***pathway(s) candidat(s)*** (ou **CPA pour Candidate Pathway Analysis**), ou encore analyse *pathway* avec à priori. Les voies biologiques sont présélectionnées suivant les connaissances dont on dispose sur la maladie étudiée et sur les mécanismes physiopathologiques mis en jeu. Dans ce type d'analyse ne sont utilisées que les données génétiques générées par les GWAS et concernant les régions où se trouvent les gènes du ou des *pathway(s) candidat(s)*.
- les analyses ***pathways sur génome entier*** (ou **GWPA, Genome-Wide Pathway Analysis**) ou sans à priori. Toutes les voies biologiques préalablement déterminées et disponibles dans les bases de données sont analysées, impliquant l'utilisation de la majorité des données générées par les GWAS.

Il existe de nombreux avantages à l'approche CPA. Etant donné que l'analyse ne porte que sur un ou quelques *pathways*, les analyses sont plus rapides et moins gourmandes en ressources informatiques. De plus, se focaliser sur un nombre restreint de *pathways* permet de définir avec précisions les gènes impliqués dans les voies biologiques ciblées par recherche bibliographique sur chacun des gènes impliqués. Cette démarche est aussi possible lorsque l'on teste tous les *pathways* mais demande alors beaucoup plus de temps.

Cependant, le fait de ne cibler que des *pathways* candidats restreint le champ de l'analyse qu'à des voies biologiques dont le rôle dans la maladie est évident.

Par conséquent, la CPA est plutôt recommandée lorsqu'il existe de forts arguments laissant supposer le rôle de ce *pathway* candidat dans la maladie.



Les études GWPA, quant-à-elles, utilisent la majorité des données génétiques disponibles. Ainsi elles possèdent l'avantage de pouvoir détecter des relations inattendues, d'une part entre des *pathways* et une maladie, alors que ces voies biologiques ne semblaient pas à priori avoir de rôle dans la pathologie, et d'autre part entre des maladies qui vont partager des *pathways* de prédisposition communs.

Cependant, les GWPA possèdent quelques inconvénients. Elles impliquent d'importantes corrections pour tests multiples, ce qui nécessite de très larges cohortes. De plus, les résultats des GWPA doivent être systématiquement bien analysés pour d'une part, prendre en compte le fait qu'il y ait souvent beaucoup de gènes qui jouent un rôle dans différents *pathways* et d'autre part, évaluer la fonction biologique du ou des *pathways* mis en évidence et ainsi voir s'il(s) peu(ven)t avoir un rôle dans la maladie.

### 2.5.3. Méthodes statistiques des analyses pathways

#### *2.5.3.1. Analyse de sur-représentation (Overrepresentation analysis)*

La base d'une analyse de sur-représentation est constituée d'une liste de gènes significativement associés à la maladie, obtenue à partir des résultats d'un GWAS. La proportion de gènes du *pathway* apparaissant dans cette liste est comparée à la proportion des gènes de la liste n'appartenant pas au *pathway* étudié (Holmans 2010).

L'analyse statistique permet de tester s'il existe une différence entre ces proportions. Si la proportion des gènes dans le *pathway* apparaissant dans la liste est significativement plus forte que la proportion de gènes qui ne sont pas dans le *pathway*, le *pathway* est dit sur-représenté dans les résultats significatifs du GWAS. Il s'agit de tests compétitifs.

#### *2.5.3.2. Gene-set enrichment analysis*

Un des problèmes majeurs de la méthode de sur-représentation est qu'elle requiert de spécifier un seuil de significativité pour définir la liste des gènes ou SNPs significatifs.

L'analyse Gene-set enrichment s'affranchit de ce problème en classant les gènes par significativité, puis teste les différences qu'il peut exister entre les classements de gènes d'un *pathway* comparé aux autres gènes. Il s'agit encore ici de tests compétitifs (Holmans 2010).



Comparé à la méthode de sur-représentation, les méthodes de gene-set enrichment présentent deux avantages :

- elles permettent de s'affranchir d'un seuil de significativité arbitraire pour déterminer la liste des gènes significatifs

- le classement prenant en compte l'association des gènes avec la maladie, elles donnent plus de poids aux gènes significativement associés à la maladie. Cependant, il se peut également qu'un unique gène fortement associé à la maladie donne un fort score d'enrichissement à chaque *pathway* auquel il appartient.

#### 2.5.3.3. *Set-based methods*

Les méthodes set-based intègrent les signaux d'association de tous les gènes ou SNPs d'un *pathway* en un seul score statistique et testent si ce score est plus important que celui auquel on s'attendrait sous l'hypothèse nulle, c'est à dire qu'il n'y a pas d'association entre les gènes du *pathway* et le phénotype (Holmans 2010).

#### 2.5.4. Considérations post-analyses

Une bonne interprétation des résultats est impérative et doit prendre en compte les biais qui ont pu être introduits dans l'étude (Wang, Li et al. 2010, Ramanan, Shen et al. 2012). Les deux biais les plus importants sont dus :

- à la taille du *pathway* (en nombre de gènes). En effet, un *pathway* avec un faible nombre de gènes peut générer des résultats faux-positifs dus à la forte association d'un seul gène ou SNP à la maladie. Inversement, un *pathway* composé d'un grand nombre de gènes est plus susceptible de montrer une association seulement par chance.

- à la possibilité que certains polymorphismes soient assignés à plusieurs gènes, du fait de leur proximité génomique.

Même si à l'heure actuelle aucune règle n'a été définie, il est primordial d'appliquer des corrections pour tests multiples aux analyses *pathways* afin de contrôler les faux-positifs.

**Au cours de mon doctorat, je me suis particulièrement intéressé à l'interaction des gènes du réseau des molécules d'adhésion dans la prédisposition à la SEP. Ce projet a abouti à un article présenté ci-après, en cours de préparation (*Publication Damotte et al., en préparation, présentée page 169*).**





**Article (1<sup>er</sup> auteur) :**

A candidate pathway analysis highlights the role of the cell adhesion molecules in Multiple Sclerosis susceptibility

**Vincent Damotte**, Léna Guillot-Noel, Nikolaos Patsopoulos, Lohith Madireddy, Mohamed El Behi, IMSGC, WTCCC2, Philip De Jager, Sergio Baranzini, Isabelle Cournu-Rebeix\*, Bertrand Fontaine\*

\* equal contribution

(En préparation)



**Introduction :** Dans le contexte des analyses *pathways*, nous nous sommes intéressés au réseau des molécules d'adhésions, d'intérêt dans la physiopathologie de la SEP. Les molécules appartenant à ce réseau sont essentielles pour l'adhésion des lymphocytes T à la surface des cellules endothéliales de la barrière hémato-encéphalique et à leur transmigration dans le SNC. Elles sont donc essentielles à la colonisation du SNC par les lymphocytes T préactivés. Parmi ces molécules appartenant à ce réseau, seul le gène *VCAMI*, codant pour un récepteur aux intégrines se trouvant à la surface des cellules endothéliales, est localisé dans une région génomique associée à la prédisposition à la SEP (IMSGC, WTCCC2 et al. 2011).

Au vu de l'importance du passage de la BHE par les cellules immunitaires dans la physiopathologie de la SEP, nous avons émis l'hypothèse que d'autres gènes interagissant au sein de ce réseau pourraient contribuer à la prédisposition à la maladie. Nous avons donc étudié le réseau des molécules d'adhésion par une Analyse *Pathway* Candidat.

**Matériel et Méthodes :** Nous avons tout d'abord établi une liste exhaustive des molécules appartenant à ce réseau. Nous avons utilisé la base de données KEGG pour rechercher les molécules d'adhésion et les autres molécules interagissant directement avec elles. Nous avons aussi utilisé la base de données SabioSciences pour rechercher les 10 facteurs de transcription semblant le plus impliqués dans la régulation de l'expression des molécules d'adhésion.

Pour chaque gène, nous avons calculé une p-value représentative du gène dans chacun des 7 datasets (issus des GWAS successifs) disponibles grâce au logiciel VEGAS qui permet de combiner les p-values de tous les SNPs d'un gène. Nous avons combiné 6 d'entre eux pour avoir une unique p-value représentative de chaque gène dans 6 datasets, le dataset de l'IMSGC de 2011 étant utilisé séparément.

Nous avons ensuite cherché si certains sous-réseaux étaient enrichis en gènes ayant de faibles p-values dans l'un ou l'autre dataset en utilisant le logiciel CytoScape et le plugin jActives module.

**Résultats :** La recherche des molécules appartenant au réseau nous a permis d'établir une liste de 76 gènes (dont 10 facteurs de transcription).

L'analyse *pathway* nous a permis de mettre en évidence 5 sous-réseaux enrichis en gènes ayant de faibles p-values.



**Conclusions :** Parmi les 5 sous-réseaux mis en évidence par l'étude, le *pathway* composé d' ITGAL, d' ICAM1 et d'ICAM3 est particulièrement d'intérêt. En effet, ICAM3 est connu comme activant le ligand lymphocytaire ITGAL, permettant ainsi sa fixation à son récepteur endothélial ICAM1.

Bloquer l'interaction entre ces 3 molécules pourrait être une piste pour le développement d'un futur traitement. Un traitement similaire, le Natalizumab, bloquant l'interaction entre ITGA4, un autre ligand lymphocytaire et son récepteur endothélial VCAM1 constitue par ailleurs un des traitements les plus efficaces dans la SEP.

**Ma contribution :** Au cours de mon doctorat, j'ai mené ce projet de sa conception à l'analyse des résultats. Seule la recherche des sous-réseaux par le logiciel CytoScape a été fait en collaboration avec Sergio Baranzini, de UCSF. J'ai écrit la première version de l'article et intégré les différentes corrections de relecture.



## **A gene pathway analysis highlights the role of cellular adhesion molecules in Multiple Sclerosis susceptibility**

Vincent Damotte<sup>1</sup>, Léna Guillot-Noel<sup>1</sup>, Nikolaos A. Patsopoulos<sup>2,3,4,5</sup>, Lohith Madireddy<sup>6</sup>, Mohamed El Behi<sup>1,7</sup>, International Multiple Sclerosis Genetics Consortium<sup>¶</sup>, Wellcome Trust Case Control Consortium 2<sup>¶</sup>, Philip L. De Jager<sup>2,4,5</sup>, Sergio E. Baranzini<sup>6</sup>, Isabelle Courneau-Rebeix<sup>1,8\*</sup>, Bertrand Fontaine<sup>1,8\*</sup>

\* equal contribution

<sup>1</sup>INSERM UMR S 975 CRICM UPMC 75013 PARIS France

<sup>2</sup>Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham & Women's Hospital, Boston, 02115 MA, USA

<sup>3</sup>Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>4</sup>Harvard Medical School, Boston, MA 02115, USA

<sup>5</sup>Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>6</sup>Department of Neurology, University of California at San Francisco, Sandler Neurosciences Center, 675 Nelson Rising Lane, San Francisco, CA 94158, USA

<sup>7</sup>Institut du Cerveau et de la Moelle épinière, ICM, 75013 Paris, France

<sup>8</sup>Département de Neurologie Pitié-Salpêtrière AP-HP 75013 PARIS France

<sup>¶</sup> Membership of the International Multiple Sclerosis Genetics Consortium (IMSGC) and the Wellcome Trust Case Control Consortium 2 (WTCCC2) is provided in the Acknowledgments



**Abstract:**

Typically Genome-Wide Association Studies (GWASs) perform per-SNP association tests in order to identify variants involved in disease or trait susceptibility. However, such an approach is not powerful enough to unravel genes that are not individually contributing to the disease/trait but that may play a role in interaction with other genes as a group. Pathway analysis is an alternative way to highlight such group of genes. Using SNPs association p-values from 8 Multiple Sclerosis (MS) GWAS datasets, we performed a candidate pathway analysis for MS susceptibility considering genes interacting in cell adhesion molecules (CAMs) network using Cytoscape software. This network is a strong candidate since it is involved in the crossing of the blood brain barrier by the T cells, an early event in MS pathophysiology, and used as an efficient therapeutic target to treat the disease. We drew up a list of 76 genes belonging to the CAMs network. We highlighted 64 sub-networks that were enriched with CAMs genes with low p-values. After filtering by a percentage of CAMs genes up to 50% and rejecting enriched signals mainly driven by transcription factors, we highlighted 5 sub-networks associated with MS susceptibility. One of them, constituted of ITGAL, ICAM1 and ICAM3 genes could be of interest to develop novel therapeutic targets.

Keywords: Multiple Sclerosis, Susceptibility, Pathway Analysis, Cell Adhesion Molecules

**Author summary (wide audience):**

Multiple Sclerosis is a common inflammatory and demyelinating disease affecting the Central Nervous System (CNS) i.e. the brain and the spinal cord. It is caused by both environmental and genetics factors. Several studies have been published and highlighted several genetic variants conferring individually a risk to the disease. In this article, I report a candidate pathway analysis to highlight some genes that are associated with Multiple Sclerosis (MS) susceptibility as a group. We analyzed the Cell Adhesion Molecules (CAMs) network, since it is involved in the transmigration of inflammatory cells into the CNS. We highlighted 5 groups of CAMs genes that contribute as groups to susceptibility to the disease. One of them consisted in 3 genes coding for molecules that could be used as potential therapeutic targets in the future.

## Introduction

Multiple Sclerosis (MS) is a common inflammatory and demyelinating disease of the central nervous system [1]. Epidemiological studies have proved the multifactorial causes of the disease, resulting from the interaction between genetic factors and currently unknown environmental factors [2].

To date, MHC genetic variants, as well as 62 non-MHC variants, have been associated with MS susceptibility by GWAS [3-6]. However, it has been estimated that several other genes contribute to disease susceptibility and are yet to be identified [7].

Several hypotheses have been proposed to explain the missing heritability [8 ,9]. One of them points at interactions between genes involved in the same biological pathway [10 ,11]. According to this hypothesis, some of the genes implicated in MS susceptibility cannot, on their own, reach GWAS significance but grouped in a pathway may collectively contribute to MS genetic component.

The initial event leading to the development of an MS lesion is blood-brain-barrier (BBB) disruption and the crossing of the latter by peripherally activated T cells [12]. This step, which is a key mechanism in our current understanding of MS pathophysiology, requires the interaction of integrins, cell surface molecules expressed by T cells, to the adhesion molecules expressed by the BBB endothelium, ultimately allowing the lymphocytes transmigration into the brain [13-16].

The importance of the BBB crossing in MS pathophysiology is stressed by the use of treatments active on adhesion molecules. For example, Natalizumab, a monoclonal antibody blocking the interaction between VLA-4 (very late antigen-4) molecule (composed of ITGB1, integrin beta 1, and ITGA4, integrin alpha 4, integrins subunits) and its receptor VCAM1 (vascular cell adhesion molecule-1), is among the most efficient therapies available for MS patients to date [17 ,18]. The role of adhesion molecules in MS has also been reinforced by the demonstration that the expression of ALCAM (activated leucocyte cell adhesion molecule) located at the surface of the BBB endothelium, was increased during the inflammatory process in MS patients. In the same study, it has been shown that in vitro expression of ALCAM, ICAM-1 (intercellular cell adhesion molecule-1) and VCAM-1 was dependent of activation of the human BBB endothelial cells by pro-inflammatory cytokines (TNF, Tumor Necrosis Factor et IFN- $\gamma$ , Interferon-gamma) [19]

Moreover, genetic studies pointed out the importance of adhesion molecules in the susceptibility to the disease [20 ,21]. In 2003, we reported that a rare haplotype of the *ICAM-1* gene (intercellular adhesion molecule 1, expressed at the surface of endothelial cells) was

under-transmitted to patients, suggesting a protective effect of this haplotype [20]. In 2009, a pathway meta-analysis combining two GWAS uncovered genetic variants that interacted in the CAM (Cell Adhesion Molecules) pathway potentially contributing to MS development [21]. This new strategy led to the identification of new genes that do not significantly contribute to MS susceptibility individually but play an important role in interaction with other genes of the biological pathway. An earlier protein interaction network-based pathway analysis (PINBPA) considered all the SNPs with a p-value below of 0.05, without any prior analysis on genes function. That study showed significant interactions among adhesion molecule-coding genes including *ICAM-1*, *ITGB2* (*integrin beta 2*), *ITGAM* (*integrin alpha M*), *ITGA6* (*integrin alpha 6*), *CD58* (*CD58 molecule*), *CD2* (*CD2 molecule*) and *CD4* (*CD4 molecule*). This observation, coming from a meta-analysis of two independent GWAS, suggests that the biological network of adhesion molecules is involved in MS susceptibility.

In order to decipher the role of the CAMs pathway in MS susceptibility, we conducted a candidate pathway analysis (CPA) focusing on adhesion molecules. We found 5 sub-networks of the CAMs pathway enriched in low p-values for genes interacting synergistically to confer MS susceptibility.

## Results

### CAMs pathway selected genes

Using KEGG database and Sabiosciences website, we identified 76 genes of interest involved in the adhesion molecule pathway. More than 85% of them (66 out of 76) were involved in 3 sub-pathways: Adhesion, Adherens Junctions and Tight Junctions, and 10 encode for transcription factors highly involved in adhesion molecule genes regulation (see Supplementary Table 1).

70 of the 76 genes were represented by at least one SNP in one or more of the 7 datasets included in D1, whereas 68 genes of the 76 were represented by at least one SNP in the WTCCC2 dataset (D2 dataset). The number of available SNPs per gene and per dataset is given in Supplementary Table 2. D1 and D2 gene-wise p-values, computed using VEGAS software (see Materials and Methods section) are given in Table 1. As expected, *CD58*, *NFKB1* and *STAT3* genes previously identified as associated with MS in GWAS show low gene-wise p-values,  $P_{D1} = 1.89 \times 10^{-6}$ ,  $P_{D2} = 0$  (meaning that  $P < 89 \times 10^{-6}$ , see Materials and Methods section);  $P_{D1} = 0.0125$ ,  $P_{D2} = 0.0011$ ;  $P_{D1} = 2.94 \times 10^{-5}$ ,  $P_{D2} = 1.5 \times 10^{-5}$  respectively.

Interestingly, *VCAM1*, previously identified as associated to the disease, with rs11581062 being the most statistical significant SNP, did not reach a significant p-value ( $P_{D1}$

=0.7224;  $P_{D2}$  =0.993) neither in the discovery nor in the replication datasets. The published associated SNP (rs11581062) is 202kb far from VCAM1 gene. We conclude that even if VCAM1 is a good candidate gene for MS physiopathology, not enough evidence support its involvement within the rs11581062 region. Furthermore, none CIS-e-QTL regulating VCAM1 expression was described within the 1 Mb region around the gene (Pritchard lab resources).

### **Sub-networks associated with MS susceptibility**

As described by IMSGC et al. [22], we used a curated human protein interaction network (PIN) dataset, consisted in a network of more than 400,000 interactions among ~25,000 proteins. Using Cytoscape software, we attributed gene-wise p-values to each of the CAMs genes (listed in Supplementary Table 1). As mentioned in IMSGC et al. (ref), Cytoscape plugin jActives modules was used to calculate a global score (Z-score) for all the possible sub-networks that could be form from the PIN dataset, using D1 and D2 p-values. Sub-networks with Z-scores greater than 3.0 are generally considered significant, i.e. these sub-networks are enriched in CAMs genes showing D1 and/or D2 gene-wise significant p-values. The cytoscape software was applied on the entire PIN dataset (~25,000 proteins , ~400,000 interactions) containing CAMs genes (see supplementary Table 1) and non-CAMs ones. 64 sub-networks were highlighted regarding the enrichment of CAMs genes with significant p-Values. Focusing on the process of BBB transmigration by T-cells, we only considered as relevant for our study 6 sub-networks with at least 50% of CAMs genes. Finally, we removed subnets that were only enriched in CAMs transcription factors in order to eliminate non-specific association signal only driven by transcription factors (see Table 2). We identified 5 sub-networks constituted by genes known to interact together within the CAMs pathway and meeting all expressed conditions (See Figure 1).

### **Discussion**

GWAS and replication studies have successfully identified approximately 62 non-MHC MS susceptibility genes [3-5]. However, an important part of genetic heritability remains to be discovered [3]. One of the hypotheses which can be put forward postulates that genes without individual effect could influence the susceptibility to the disease through genetic interactions [10]. Here, we report the first candidate pathway analysis on the Cell Adhesion Molecules (CAMs) network in MS. Our results highlight 5 sub-networks enriched

in CAMs genes with significant p-values, reflecting a potential genetic contribution of these molecules in MS susceptibility.

In 2009, Baranzini et al. identified 7 genes of the CAMs pathway as associated with MS susceptibility considering their interaction [21]. Focusing on the CAMs pathway and using two powerful datasets (5,545+9,772 patients and 12,153+17,376 controls), our study refines the role of these genes as MS genetic factors.

Two of the 5 identified sub-networks (sub-network 2 and 27) appear to be the highest contributors to MS physiopathology. The sub-network 2 is constituted by ICAM1, ICAM3 and ITGAL genes that are involved in the adhesion process of the T cells on the endothelial cells of the Blood-Brain Barrier (BBB). The adhesion process between endothelial and T cells is one of the earliest event in MS development leading to the T cells transmigration into the central nervous system across the BBB. Of note, ITGAL, ICAM1 and ICAM3 molecules constitute ligand-receptor complexes, ITGAL interacting both with ICAM1 and ICAM3. Our hypothesis on the role of these genes in MS susceptibility is that some intra- and inter-gene combinations of polymorphisms could influence the adhesion process of T cells on the BBB leading to a positive or negative modulation of the inflammatory cells flow into the CNS.

The sub-network 27 is constituted by CD82 (KAI1), ITGA4, ITGB1, ITGB2, and HLA-DMA genes. The ITGB1 and ITGA4 genes code for the two subunits of the VLA4 molecule. CD82 is notably known for its inhibitory effect on ITGB1 activation [23]. The effective role of CD82 in inflammatory response has not been yet investigated. In the MS context, it would be relevant to analyze the development and the severity of EAE (Experimental Autoimmune Encephalomyelitis; MS animal model) in CD82 deficient mice. Since CD82 initiates the differentiation of oligodendrocytes precursors into mature myelinating cells [24], quantification of remyelination in this model would be of interest.

Interestingly we identified two sub-networks containing the ITGA4 molecule which is the target of Natalizumab. The interaction between VLA-4 and his receptor VCAM1 is known to be crucial for the transmigration of T cells across the BBB and potentially the molecular basis for this treatment's effectiveness.

In conclusion, our results highlighting the role of CAMs genes interactions in MS susceptibility could be of high interest to identify new targets for efficient treatments for MS. A monoclonal antibody directed against ITGA4 that prevent its interaction with VCAM1 already exists [17]. We propose that targeting ITGAL or ICAM1 or ICAM3 in order to block their interactions could also be an efficient way to prevent the crossing of the BBB by T cells.

## **Materials and Methods**

### **Selection of genes coding for CAMs pathway molecules**

The selection of the genes belonging to the network has been performed in 3 phases: Using KEGG (Kyoto Encyclopedia for Genes and Genomes) database (<http://www.genome.jp/kegg/>) [25 ,26], we searched for adhesion molecules involved in tight and adherens junctions between BBB endothelial cells, and for those involved in T cells transmigration. Still Using KEGG database, we then identified molecules interacting directly with adhesion molecules. In a last step, we selected the 10 transcription factors that are most involved in adhesion molecule-coding genes expression and regulation using the SabioSciences website (<http://www.sabiosciences.com>) through GeneCards database (<http://www.genecards.org/>) [27]. Using this 3-step strategy, we compiled a list of 76 genes of interest.

### **Datasets**

Eight datasets were used in this study. Seven datasets (IMSGC UK, IMSGC US, ANZGene, GeneMSA DU, GeneMSA SW, GeneMSA US and BWH/MIGEN) came from a previous meta-analysis as described in Patsopoulos et al. [4]. The 8<sup>th</sup> dataset consists in the MS GWAS published by the IMSGC and WTCCC2 in 2011 [3]. From these 8 datasets, we used calculated p-values of selected genes of (+/- 1kb from the 5' and 3' UTR).

### **Statistical analysis**

#### **Gene-wise p-values computation**

Since the IMSGC-WTCCC2 dataset is to date the most powerful GWAS published in MS, we decided to use this dataset separately to avoid signals losses. For the 8 datasets we used VEGAS software [28] to compute an individual gene-wise p-value corresponding to the association of each gene with MS in each dataset. VEGAS assigns SNPs to each of 17,787 autosomal genes according to positions on the UCSC Genome Browser (hg18 assembly). For the capture of regulatory regions and SNPs in LD, gene boundaries are defined as +/- 50kb of each gene. VEGAS takes into account LD patterns between markers within a gene by using Monte-Carlo simulations from the multivariate normal distribution on the basis of the LD structure of a set of reference individuals (the HapMap2 CEU [Utah residents with ancestry from northern and western Europe from the CEPH collection] population). In VEGAS, the number of simulations per gene is determined adaptively. In the first stage, 10<sup>3</sup> simulations are performed. If the resulting empirical p value is less than 0.1, 10<sup>4</sup> simulations are then

performed. If the empirical p value from  $10^4$  simulations is less than 0.001, the program will perform  $10^6$  simulations. At each stage, the simulations are mutually exclusive. For computational reasons, if the empirical p value is 0, then no more simulations will be performed. An empirical p value of 0 from  $10^6$  simulations can be interpreted as  $p < 10^{-6}$ , which exceeds a Bonferroni-corrected threshold of  $p < 2.8 \times 10^{-6}$ . Then, we applied Fisher's method to meta-analyze the gene-wise p-values across the seven datasets described above (named D1) for each of the 76 selected genes. The gene-wise p-values from the IMSGC-WTCCC2 dataset were used separately, as a second independent dataset (named D2).

### **Protein interaction network-based pathway analysis (PINBPA)**

As described by IMSGC et al. in 2013 [22], we integrated data from a curated human protein interaction network (PIN) dataset in Cytoscape software [29]. We assigned the D1 and D2 gene-wise p-values obtained previously to each of the 76 selected genes. Cytoscape plugin jActives modules was used to calculate a global score for each sub-network enriched in CAMs genes showing D1 and/or D2 gene-wise significant p-values. For each selected gene, we searched for sub-networks enriched in CAMs genes showing D1 and/or D2 gene-wise significant p-values.

### **Acknowledgments.**

This study was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Fondation pour la Recherche sur la Sclérose En Plaques (ARSEP), the Association Française contre les Myopathies (AFM), GIS-IBISA and ICM Carnot Institute. The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06. We thank ICM, CIC Pitié-Salpêtrière, Généthon and REFGENSEP's and IMSGC's members for their help and support as well as Jorge Oksenberg and Pierre-Antoine Gourraud. VD received a travel grant from the Fondation ARSEP and ICM Carnot Institute. Philip L. De Jager is a Harry Weaver Neuroscience Scholar of the National MS Society. SEB is a Harry Weaver Neuroscience fellow from the US National MS Society. This investigation was supported (in part) by a Postdoctoral Fellowship from the National Multiple Sclerosis Society to Nikolaos A. Patsopoulos (FG 1938-A-1).



The members of the **International Multiple Sclerosis Genetics Consortium (IMSGC)** and **Wellcome Trust Case Control Consortium 2 (WTCCC2)** are:

Stephen Sawcer<sup>1</sup>, Garrett Hellenthal<sup>2</sup>, Matti Pirinen<sup>2</sup>, Chris C.A. Spencer<sup>2</sup>, Nikolaos A. Patsopoulos<sup>3,5</sup>, Loukas Moutsianas<sup>6</sup>, Alexander Dilthey<sup>6</sup>, Zhan Su<sup>2</sup>, Colin Freeman<sup>2</sup>, Sarah E. Hunt<sup>7</sup>, Sarah Edkins<sup>7</sup>, Emma Gray<sup>7</sup>, David R. Booth<sup>8</sup>, Simon C. Potter<sup>7</sup>, An Goris<sup>9</sup>, Gavin Band<sup>2</sup>, Annette Bang Oturai<sup>10</sup>, Amy Strange<sup>2</sup>, Janna Saarela<sup>11</sup>, Céline Bellenguez<sup>2</sup>, Bertrand Fontaine<sup>12</sup>, Matthew Gillman<sup>7</sup>, Bernhard Hemmer<sup>13</sup>, Rhian Gwilliam<sup>7</sup>, Frauke Zipp<sup>14,15</sup>, Alagurevathi Jayakumar<sup>7</sup>, Roland Martin<sup>16</sup>, Stephen Leslie<sup>17</sup>, Stanley Hawkins<sup>18</sup>, Eleni Giannoulatou<sup>2</sup>, Sandra D'alfonso<sup>19</sup>, Hannah Blackburn<sup>7</sup>, Filippo Martinelli Boneschi<sup>20</sup>, Jennifer Liddle<sup>7</sup>, Hanne F. Harbo<sup>21,22</sup>, Marc L. Perez<sup>7</sup>, Anne Spurkland<sup>23</sup>, Matthew J Waller<sup>7</sup>, Marcin P. Mycko<sup>24</sup>, Michelle Ricketts<sup>7</sup>, Manuel Comabella<sup>25</sup>, Naomi Hammond<sup>7</sup>, Ingrid Kockum<sup>26</sup>, Owen T. McCann<sup>7</sup>, Maria Ban<sup>1</sup>, Pamela Whittaker<sup>7</sup>, Anu Kemppinen<sup>1</sup>, Paul Weston<sup>7</sup>, Clive Hawkins<sup>27</sup>, Sara Widaa<sup>7</sup>, John Zajicek<sup>28</sup>, Serge Dronov<sup>7</sup>, Neil Robertson<sup>29</sup>, Suzannah J. Bumpstead<sup>7</sup>, Lisa F. Barcellos<sup>30,31</sup>, Rathi Ravindrarajah<sup>7</sup>, Roby Abraham<sup>27</sup>, Lars Alfredsson<sup>32</sup>, Kristin Ardlie<sup>4</sup>, Cristin Aubin<sup>4</sup>, Amie Baker<sup>1</sup>, Katharine Baker<sup>29</sup>, Sergio E. Baranzini<sup>33</sup>, Laura Bergamaschi<sup>19</sup>, Roberto Bergamaschi<sup>34</sup>, Allan Bernstein<sup>31</sup>, Achim Berthele<sup>13</sup>, Mike Boggild<sup>35</sup>, Jonathan P. Bradfield<sup>36</sup>, David Brassat<sup>37</sup>, Simon A. Broadley<sup>38</sup>, Dorothea Buck<sup>13</sup>, Helmut Butzkueven<sup>39,42</sup>, Ruggero Capra<sup>43</sup>, William M. Carroll<sup>44</sup>, Paola Cavalla<sup>45</sup>, Elisabeth G. Celius<sup>21</sup>, Sabine Cepok<sup>13</sup>, Rosetta Chiavacci<sup>36</sup>, Françoise Clerget-Darpoux<sup>46</sup>, Katleen Clysters<sup>9</sup>, Giancarlo Comi<sup>20</sup>, Mark Cossburn<sup>29</sup>, Isabelle Cournu-Rebeix<sup>12</sup>, Mathew B. Cox<sup>47</sup>, Wendy Cozen<sup>48</sup>, Bruce A.C. Cree<sup>33</sup>, Anne H. Cross<sup>49</sup>, Daniele Cusi<sup>50</sup>, Mark J. Daly<sup>4,51,52</sup>, Emma Davis<sup>53</sup>, Paul I.W. de Bakker<sup>3,4,54,55</sup>, Marc Debouverie<sup>56</sup>, Marie Beatrice D'hooghe<sup>57</sup>, Katherine Dixon<sup>53</sup>, Rita Dobosi<sup>9</sup>, Bénédicte Dubois<sup>9</sup>, David Ellinghaus<sup>58</sup>, Irina Elovaara<sup>59,60</sup>, Federica Esposito<sup>20</sup>, Claire Fontenille<sup>12</sup>, Simon Foote<sup>61</sup>, Andre Franke<sup>58</sup>, Daniela Galimberti<sup>62</sup>, Angelo Ghezzi<sup>63</sup>, Joseph Glessner<sup>36</sup>, Refujia Gomez<sup>33</sup>, Olivier Gout<sup>64</sup>, Colin Graham<sup>65</sup>, Struan F.A. Grant<sup>36,66,67</sup>, Franca Rosa Guerini<sup>68</sup>, Hakon Hakonarson<sup>36,66,67</sup>, Per Hall<sup>69</sup>, Anders Hamsten<sup>70</sup>, Hans-Peter Hartung<sup>71</sup>, Rob N. Heard<sup>8</sup>, Simon Heath<sup>72</sup>, Jeremy Hobart<sup>28</sup>, Muna Hoshi<sup>13</sup>, Carmen Infante-Duarte<sup>73</sup>, Gillian Ingram<sup>29</sup>, Wendy Ingram<sup>28</sup>, Talat Islam<sup>48</sup>, Maja Jagodic<sup>26</sup>, Michael Kabesch<sup>74</sup>, Allan G. Kermodé<sup>44</sup>, Trevor J. Kilpatrick<sup>39,40,75</sup>, Cecilia Kim<sup>36</sup>, Norman Klopp<sup>76</sup>, Keijo Koivisto<sup>77</sup>, Malin Larsson<sup>70</sup>, Mark Lathrop<sup>72</sup>, Jeannette S. Lechner-Scott<sup>47,78</sup>, Maurizio A. Leone<sup>79</sup>, Virpi Leppä<sup>11,80</sup>, Ulrika Liljedahl<sup>81</sup>, Izaura Lima Bomfim<sup>26</sup>, Robin R. Lincoln<sup>33</sup>, Jenny Link<sup>26</sup>, Jianjun Liu<sup>82</sup>, Aslaug R. Lorentzen<sup>22,83</sup>, Sara Lupoli<sup>50,84</sup>, Fabio Macciardi<sup>50,85</sup>, Thomas Mack<sup>48</sup>, Mark Marriott<sup>39,40</sup>, Vittorio Martinelli<sup>20</sup>, Deborah Mason<sup>86</sup>, Jacob L. McCauley<sup>87</sup>, Frank Mentch<sup>36</sup>, Inger-Lise Mero<sup>21,83</sup>, Tania Mihalova<sup>27</sup>, Xavier Montalban<sup>25</sup>, John Mottershead<sup>88,89</sup>, Kjell-Morten Myhr<sup>90,91</sup>, Paola Naldi<sup>79</sup>, William Ollier<sup>53</sup>, Alison Page<sup>92</sup>, Aarno Palotie<sup>7,11,93,94</sup>, Jean Pelletier<sup>95</sup>, Laura Piccio<sup>49</sup>, Trevor Pickersgill<sup>29</sup>, Fredrik Piehl<sup>26</sup>, Susan Pobywajlo<sup>5</sup>, Hong L. Quach<sup>30</sup>, Patricia P. Ramsay<sup>30</sup>, Mauri Reunanen<sup>96</sup>, Richard Reynolds<sup>97</sup>, John D. Rioux<sup>98</sup>, Mariaemma Rodegher<sup>20</sup>, Sabine Roesner<sup>16</sup>, Justin P. Rubio<sup>39</sup>, Ina-Maria Rückert<sup>76</sup>, Erika Salvi<sup>50,100</sup>, Adam Santaniello<sup>33</sup>, Catherine A. Schaefer<sup>31</sup>, Stefan Schreiber<sup>58,101</sup>, Christian Schulze<sup>102</sup>, Rodney J. Scott<sup>47</sup>, Finn Sellebjerg<sup>10</sup>, Krzysztof W. Selmaj<sup>24</sup>, David Sexton<sup>103</sup>, Ling Shen<sup>31</sup>, Brigid Simms-Acuna<sup>31</sup>, Sheila Skidmore<sup>1</sup>, Patrick M.A. Sleiman<sup>36,66</sup>, Cathrine Smestad<sup>21</sup>, Per Soelberg Sørensen<sup>10</sup>, Helle Bach Søndergaard<sup>10</sup>, Jim Stankovich<sup>61</sup>, Richard C. Strange<sup>27</sup>, Anna-Maija Sulonen<sup>11,80</sup>, Emilie Sundqvist<sup>26</sup>, Ann-Christine Syvänen<sup>81</sup>, Francesca Taddeo<sup>100</sup>, Bruce Taylor<sup>61</sup>, Jenefer M. Blackwell<sup>104,105</sup>, Pentti Tienari<sup>106</sup>, Elvira Bramon<sup>107</sup>, Ayman Tourbah<sup>108</sup>, Matthew A. Brown<sup>109</sup>, Ewa Tronczynska<sup>24</sup>, Juan P. Casas<sup>110</sup>, Niall Tubridy<sup>40,111</sup>, Aiden Corvin<sup>112</sup>, Jane Vickery<sup>28</sup>, Janusz Jankowski<sup>113</sup>, Pablo Villoslada<sup>114</sup>, Hugh S. Markus<sup>115</sup>, Kai Wang<sup>36,66</sup>, Christopher G. Mathew<sup>116</sup>, James Wason<sup>117</sup>, Colin N.A. Palmer<sup>118</sup>, H-Erich Wichmann<sup>76,119,120</sup>, Robert Plomin<sup>121</sup>, Ernest Willoughby<sup>122</sup>, Anna Rautanen<sup>2</sup>, Juliane Winkelmann<sup>13,123,124</sup>, Michael Wittig<sup>58,125</sup>, Richard C. Trembath<sup>116</sup>,

Jacqueline Yaouanq<sup>126</sup>, Ananth C. Viswanathan<sup>127</sup>, Haitao Zhang<sup>36,66</sup>, Nicholas W. Wood<sup>128</sup>, Rebecca Zuvich<sup>103</sup>, Panos Deloukas<sup>7</sup>, Cordelia Langford<sup>7</sup>, Audrey Duncanson<sup>129</sup>, Jorge R. Oksenberg<sup>33</sup>, Margaret A. Pericak-Vance<sup>87</sup>, Jonathan L. Haines<sup>103</sup>, Tomas Olsson<sup>26</sup>, Jan Hillert<sup>26</sup>, Adrian J. Ivinson<sup>51,130</sup>, Philip L. De Jager<sup>4,5,51</sup>, Leena Peltonen<sup>7,11,80,93,94</sup>, Graeme J. Stewart<sup>8</sup>, David A. Hafler<sup>4,131</sup>, Stephen L. Hauser<sup>33</sup>, Gil McVean<sup>2</sup>, Peter Donnelly<sup>2,6</sup> and Alastair Compston<sup>1</sup>

<sup>1</sup>University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, Cambridge, UK

<sup>2</sup>Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, UK

<sup>3</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>4</sup>Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>5</sup>Center for Neurologic Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, MA, USA

<sup>6</sup>Dept Statistics, University of Oxford, Oxford, UK

<sup>7</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

<sup>8</sup>Westmead Millennium Institute, University of Sydney, Australia

<sup>9</sup>Laboratory for Neuroimmunology, Section of Experimental Neurology, Katholieke Universiteit Leuven, Leuven, Belgium

<sup>10</sup>Danish Multiple Sclerosis Center, Department of Neurology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

<sup>11</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

<sup>12</sup>INSERM UMR S 975 CRICM, UPMC, Département de Neurologie Pitié-Salpêtrière, AP-HP, Paris, France

<sup>13</sup>Department of Neurology, Klinikum Rechts der Isar der Technischen Universität, Munich, Germany

<sup>14</sup>Department of Neurology, University Medicine Mainz, Johannes Gutenberg University Mainz, Mainz, Germany

<sup>15</sup>Max Delbrueck Center for Molecular Medicine, Berlin, Germany

<sup>16</sup>Institute for Neuroimmunology and Clinical MS Research (inims), Centre for Molecular Neurobiology, Hamburg, Germany

<sup>17</sup>Department of Clinical Pharmacology, University of Oxford, Old Road Campus Research Building, Oxford, UK

<sup>18</sup>Queen's University Belfast, University Road, Belfast, Northern Ireland, UK

<sup>19</sup>Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy

<sup>20</sup>Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy

<sup>21</sup>Department of Neurology, Oslo University Hospital, Oslo, Norway

<sup>22</sup>Department of Neurology, University of Oslo, Oslo, Norway

<sup>23</sup>Institute of Basal Medical Sciences, University of Oslo, Oslo, Norway

<sup>24</sup>Department of Neurology, Laboratory of Neuroimmunology, Medical University of Lodz, Lodz, Poland

<sup>25</sup>Clinical Neuroimmunology Unit, Multiple Sclerosis Center of Catalonia (CEM-Cat), Vall d'Hebron University Hospital, Barcelona, Spain

<sup>26</sup>Department of Clinical Neurosciences, Centre for Molecular Medicine CMM, Karolinska Institutet, Karolinska Hospital, Stockholm, Sweden

- <sup>27</sup>Keele University Medical School, Stoke-on-Trent, UK
- <sup>28</sup>Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth, Clinical Neurology Research Group, Tamar Science Park, Plymouth, UK
- <sup>29</sup>Department of Neurology, University Hospital of Wales, Heath Park, Cardiff, UK
- <sup>30</sup>Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, USA
- <sup>31</sup>Kaiser Permanente Northern California Division of Research, CA, USA
- <sup>32</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
- <sup>33</sup>Department of Neurology, University of California San Francisco, San Francisco, CA, USA
- <sup>34</sup>Neurological Institute C. Mondino, IRCCS, Pavia, Italy
- <sup>35</sup>The Walton Centre for Neurology and Neurosurgery, Liverpool, UK
- <sup>36</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
- <sup>37</sup>INSERM U 563 et Pôle Neurosciences, Hôpital Purpan, Toulouse, France
- <sup>38</sup>School of Medicine, Griffith University, Australia
- <sup>39</sup>Florey Neuroscience Institutes, University of Melbourne, Victoria, Australia
- <sup>40</sup>Royal Melbourne Hospital, Parkville, Victoria, Australia,
- <sup>41</sup>Box Hill Hospital, Box Hill, Australia
- <sup>42</sup>Department of Medicine, RMH Cluster, University of Melbourne, Victoria, Australia
- <sup>43</sup>Multiple Sclerosis Centre, Department of Neurology, Ospedali Civili di Brescia, Brescia, Italy
- <sup>44</sup>Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Perth, Australia
- <sup>45</sup>Department of Neurosciences, University of Turin, A.O.U. San Giovanni Battista, Turin, Italy
- <sup>46</sup>INSERM U535, Univ Paris-Sud, Villejuif, France
- <sup>47</sup>University of Newcastle, University Drive, Callaghan NSW, Australia
- <sup>48</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
- <sup>49</sup>Department of Neurology, Washington University, St Louis MO, USA
- <sup>50</sup>University of Milan, Department of Medicine, Surgery and Dentistry, AO San Paolo, University of Milan, c/o Filarete Foundation - Milano, Italy
- <sup>51</sup>Harvard Medical School, Boston, MA, USA
- <sup>52</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA
- <sup>53</sup>The UK DNA Banking Network, Centre for Integrated Genomic Medical Research, University of Manchester, UK
- <sup>54</sup>Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
- <sup>55</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
- <sup>56</sup>Service de Neurologie, Hôpital Central, Nancy, France
- <sup>57</sup>National Multiple Sclerosis Center, Melsbroek, Belgium
- <sup>58</sup>Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany
- <sup>59</sup>Department of Neurology, Tampere University Hospital, Tampere, Finland
- <sup>60</sup>University of Tampere, Medical School, Tampere, Finland
- <sup>61</sup>Menzies Research Institute, Hobart, Tasmania, Australia
- <sup>62</sup>Department of Neurological Sciences, Centro Dino Ferrari, University of Milan, Fondazione Cà Granda, Ospedale Maggiore Policlinico, Milan, Italy
- <sup>63</sup>Centro Studi Sclerosi Multipla, Ospedale di Gallarate, Gallarate (VA), Italy
- <sup>64</sup>Service de Neurologie, Fondation Ophtalmologique Adolphe de Rothschild, Paris, France

- <sup>65</sup>Belfast Health and Social Care Trust, City Hospital, Belfast, Northern Ireland, UK
- <sup>66</sup>Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
- <sup>67</sup>Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
- <sup>68</sup>Laboratory of Molecular Medicine and Biotechnology, Don C. Gnocchi Foundation IRCCS, S. Maria Nascente, Milan, Italy
- <sup>69</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
- <sup>70</sup>Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital Solna, Stockholm, Sweden
- <sup>71</sup>Department of Neurology, Heinrich-Heine-University, Düsseldorf, Germany
- <sup>72</sup>Centre National de Genotypage, Evry Cedex, France
- <sup>73</sup>Experimental and Clinical Research Center, Charité – Universitätsmedizin Berlin and Max Delbrueck Center for Molecular Medicine, Berlin, Germany
- <sup>74</sup>Clinic for Paediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Germany
- <sup>75</sup>Centre for Neuroscience, University of Melbourne, Victoria, Australia
- <sup>76</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Munich, Germany
- <sup>77</sup>Seinäjoki Central Hospital, Seinäjoki, Finland
- <sup>78</sup>Hunter Medical Research Institute, John Hunter Hospital, Lookout Road, New Lambton NSW, Australia
- <sup>79</sup>SCDU Neurology, Maggiore della Carità Hospital, Novara, Italy
- <sup>80</sup>Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland
- <sup>81</sup>Molecular Medicine, Department of Medical Sciences, Uppsala University, Entrance 70, 3rd Floor, Res Dept 2, Univeristy Hospital, Uppsala, Sweden
- <sup>82</sup>Human Genetics and Cancer Biology, Genome Institute of Singapore, Singapore
- <sup>83</sup>Institute of Immunology, Oslo University Hospital, Oslo, Norway
- <sup>84</sup>Institute of Experimental Neurology (INSPE), San Raffaele Scientific Institute, Milan, Italy
- <sup>85</sup>Dept of Psychiatry and Human Behavior, University of California, Irvine (UCI), Irvine CA, USA
- <sup>86</sup>Christchurch School of Medicine, University of Otago, Christchurch, New Zealand
- <sup>87</sup>John P. Hussman Institute for Human Genomics and The Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, USA
- <sup>88</sup>Greater Manchester Centre for Clinical Neurosciences, Hope Hospital, Salford, UK
- <sup>89</sup>The Department of Neurology, Dunedin Public Hospital, Otago, NZ
- <sup>90</sup>The Multiple Sclerosis National Competence Centre, Department of Neurology, Haukeland University Hospital, Bergen, Norway
- <sup>91</sup>Department of Clinical Medicine, University of Bergen, Bergen, Norway
- <sup>92</sup>Plymouth Hospitals NHS Trust, Department of Neurology, Derriford Hospital, Plymouth, UK
- <sup>93</sup>Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland
- <sup>94</sup>Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
- <sup>95</sup>Pôle Neurosciences Cliniques, Service de Neurologie, Hôpital de la Timone, Marseille, France
- <sup>96</sup>Department Neurology, Oulu University Hospital, Oulu, Finland

<sup>97</sup>UK MS Tissue Bank, Wolfson Neuroscience Laboratories, Imperial College London, Hammersmith Hospital, London, UK

<sup>98</sup>Université de Montréal & Montreal Heart Institute, Research Center, Montreal, Quebec, Canada

<sup>100</sup>KOS Genetic Srl, Milan - Italy

<sup>101</sup>Department of General Internal Medicine, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany

<sup>102</sup>Systems Biology and Protein-Protein Interaction, Center for Molecular Neurobiology, Hamburg, Germany

<sup>103</sup>Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, USA

<sup>104</sup>Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Australia

<sup>105</sup>Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, UK

<sup>106</sup>Department of Neurology, Helsinki University Central Hospital and Molecular Neurology Programme, Biomedicum, University of Helsinki, Helsinki, Finland

<sup>107</sup>Division of Psychological Medicine and Psychiatry, Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London, UK

<sup>108</sup>Service de Neurologie et Faculté de Médecine de Reims, Université de Reims Champagne-Ardenne, Reims, France

<sup>109</sup>University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Australia

<sup>110</sup>Dept Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

<sup>111</sup>St. Vincent's University Hospital, Dublin, Ireland

<sup>112</sup>Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland

<sup>113</sup>Centre for Gastroenterology, Bart's and the London School of Medicine and Dentistry, London, UK

<sup>114</sup>Department of Neurosciences, Institute of Biomedical Research August Pi Sunyer (IDIBAPS), Hospital Clinic of Barcelona, Barcelona, Spain

<sup>115</sup>Clinical Neurosciences, St George's University of London, London, UK

<sup>116</sup>Dept Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK

<sup>117</sup>Medical Research Council Biostatistics Unit, Robinson Way, Cambridge, UK

<sup>118</sup>Biomedical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK

<sup>119</sup>Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany

<sup>120</sup>Klinikum Grosshadern, Munich, Germany

<sup>121</sup>King's College London, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London, UK

<sup>122</sup>Department of Neurology, Auckland City Hospital, Grafton Road, Auckland, New Zealand

<sup>123</sup>Institut für Humangenetik, Technische Universität München, Germany

<sup>124</sup>Institut für Humangenetik, Helmholtz Zentrum München, Germany

<sup>125</sup>Popgen Biobank, Christian-Albrechts University Kiel, Kiel, Germany

<sup>126</sup>Pôle Recherche et Santé Publique, CHU Pontchaillou, Rennes, France

<sup>127</sup>NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK

<sup>128</sup>Dept Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK

<sup>129</sup>Molecular and Physiological Sciences, The Wellcome Trust, London, UK

<sup>130</sup>Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA, USA

<sup>131</sup>Department of Neurology & Immunology, Yale University Medical School, New Haven, CT, USA

## References

1. Nylander A, Hafler DA. Multiple sclerosis. *The Journal of clinical investigation* 2012;**122**(4):1180-8.
2. Compston A, Coles A. Multiple sclerosis. *Lancet* 2008;**372**(9648):1502-17.
3. IMSGC, WTCCC2, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, D'Alfonso S, Blackburn H, Boneschi FM, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kempainen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BA, Cross AH, Cusi D, Daly MJ, Davis E, de Bakker PI, Debouverie M, D'Hooghe M B, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SF, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W, Islam T, Jagodic M, Kabesch M, Kermodé AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppä V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, Macciardi F, Mack T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Ruckert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellebjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PM, Smestad C, Sorensen PS, Sondergaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvanen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CN, Wichmann HE, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivinson AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P, Compston A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011;**476**(7359):214-9.
4. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, Sandbrink R, Pohl C, Edan G, Kappos L, Miller D, Montalban J, Polman CH, Freedman MS, Hartung HP, Arnason BG, Comi G, Cook S, Filippi M, Goodin DS, Jeffery D, O'Connor P, Ebers GC, Langdon D, Reder AT, Traboulsee A, Zipp F, Schimrigk S, Hillert J, Bahlo M, Booth DR, Broadley S, Brown MA, Browning BL, Browning SR, Butzkueven H, Carroll WM, Chapman C, Foote SJ, Griffiths L,

- Kermode AG, Kilpatrick TJ, Lechner-Scott J, Marriott M, Mason D, Moscato P, Heard RN, Pender MP, Perreau VM, Perera D, Rubio JP, Scott RJ, Slee M, Stankovich J, Stewart GJ, Taylor BV, Tubridy N, Willoughby E, Wiley J, Matthews P, Boneschi FM, Compston A, Haines J, Hauser SL, McCauley J, Ivinson A, Oksenberg JR, Pericak-Vance M, Sawcer SJ, De Jager PL, Hafler DA, de Bakker PI. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011;**70**(6):897-912.
5. Lill CM, Schjeide BM, Graetz C, Liu T, Damotte V, Akkad DA, Blaschke P, Gerdes LA, Kroner A, Luessi F, Cournu-Rebeix I, Hoffjan S, Winkelmann A, Touze E, Pico F, Corcia P, Otaegui D, Antiguiedad A, Alcina A, Comabella M, Montalban X, Olascoaga J, Matesanz F, Dorner T, Li SC, Steinhagen-Thiessen E, Lindenberger U, Chan A, Rieckmann P, Hartung HP, Aktas O, Lohse P, Buttmann M, Kumpfel T, Kubisch C, Zettl UK, Epplen JT, Fontaine B, Zipp F, Vandebroek K, Bertram L. Genome-wide significant association of ANKRD55 rs6859219 and multiple sclerosis risk. *Journal of medical genetics* 2013.
  6. Lill CM, Schjeide BM, Graetz C, Ban M, Alcina A, Ortiz MA, Perez J, Damotte V, Booth D, Lopez de Lapuente A, Broer L, Schilling M, Akkad DA, Aktas O, Alloza I, Antiguiedad A, Arroyo R, Blaschke P, Buttmann M, Chan A, Compston A, Cournu-Rebeix I, Dorner T, Epplen JT, Fernandez O, Gerdes LA, Guillot-Noel L, Hartung HP, Hoffjan S, Izquierdo G, Kempainen A, Kroner A, Kubisch C, Kumpfel T, Li SC, Lindenberger U, Lohse P, Lubetzki C, Luessi F, Malhotra S, Mescheriakova J, Montalban X, Papeix C, Paredes LF, Rieckmann P, Steinhagen-Thiessen E, Winkelmann A, Zettl UK, Hintzen R, Vandebroek K, Stewart G, Fontaine B, Comabella M, Urcelay E, Matesanz F, Sawcer S, Bertram L, Zipp F. MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. *Brain* 2013;**136**(Pt 6):1778-82.
  7. Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA, Haines JL. Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. *Am J Hum Genet* 2010;**86**(4):621-5.
  8. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009;**461**(7265):747-53.
  9. Marian AJ. Elements of 'missing heritability'. *Current opinion in cardiology* 2012;**27**(3):197-201.
  10. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010;**11**(12):843-54.
  11. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends in genetics : TIG* 2012;**28**(7):323-32.
  12. Holman DW, Klein RS, Ransohoff RM. The blood-brain barrier, chemokines and multiple sclerosis. *Biochim Biophys Acta* 2011;**1812**(2):220-30.
  13. Engelhardt B, Wolburg H. Mini-review: Transendothelial migration of leukocytes: through the front door or around the side of the house? *Eur J Immunol* 2004;**34**(11):2955-63.
  14. Engelhardt B, Ransohoff RM. The ins and outs of T-lymphocyte trafficking to the CNS: anatomical sites and molecular mechanisms. *Trends Immunol* 2005;**26**(9):485-95.
  15. Engelhardt B. Immune cell entry into the central nervous system: involvement of adhesion molecules and chemokines. *J Neurol Sci* 2008;**274**(1-2):23-6.
  16. Man S, Ubogu EE, Ransohoff RM. Inflammatory cell migration into the central nervous system: a few new twists on an old tale. *Brain Pathol* 2007;**17**(2):243-50.
  17. Yednock TA, Cannon C, Fritz LC, Sanchez-Madrid F, Steinman L, Karin N. Prevention of experimental autoimmune encephalomyelitis by antibodies against alpha 4 beta 1 integrin. *Nature* 1992;**356**(6364):63-6.
  18. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, Phillips JT, Lublin FD, Giovannoni G, Wajgt A, Toal M, Lynn F, Panzara MA, Sandrock AW. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2006;**354**(9):899-910.

19. Cayrol R, Wosik K, Berard JL, Dodelet-Devillers A, Ifergan I, Kebir H, Haqqani AS, Kreymborg K, Krug S, Moumdjian R, Bouthillier A, Becher B, Arbour N, David S, Stanimirovic D, Prat A. Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. *Nat Immunol* 2008;**9**(2):137-45.
20. Cournu-Rebeix I, Genin E, Lesca G, Azoulay-Cayla A, Tubridy N, Noe E, Clanet M, Edan G, Clerget-Darpoux F, Semana G, Fontaine B. Intercellular adhesion molecule-1: a protective haplotype against multiple sclerosis. *Genes Immun* 2003;**4**(7):518-23.
21. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, Wu W, Uitdehaag BM, Kappos L, Polman CH, Matthews PM, Hauser SL, Gibson RA, Oksenberg JR, Barnes MR. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009;**18**(11):2078-90.
22. IMSGC. Network-Based Multiple Sclerosis Pathway Analysis with GWAS Data from 15,000 Cases and 30,000 Controls. *Am J Hum Genet* 2013.
23. Lee HA, Park I, Byun HJ, Jeoung D, Kim YM, Lee H. Metastasis suppressor KAI1/CD82 attenuates the matrix adhesion of human prostate cancer cells by suppressing fibronectin expression and beta1 integrin activation. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2011;**27**(5):575-86.
24. Mela A, Goldman JE. CD82 blocks cMet activation and overcomes hepatocyte growth factor effects on oligodendrocyte precursor differentiation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2013;**33**(18):7952-60.
25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000;**28**(1):27-30.
26. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* 2012;**40**(Database issue):D109-14.
27. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. *Trends in genetics : TIG* 1997;**13**(4):163.
28. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010;**87**(1):139-45.
29. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. *Nature protocols* 2007;**2**(10):2366-82.



## Legend

### Figure 1. Subnets enriched in CAMs genes with significant D1 and/or D2\_pvalue

Each node represents a gene product and each edge, a physical interaction reported in at least two independent publications. Each subnet is enriched with CAMs genes having low p-values. Nodes are colored according to D1\_p-values (red: highly significant, orange: moderate, yellow: D1\_p-value = 0.05, white: not significant). Analysis was performed using Cytoscape and *jActive module plugin*.

### Table 1. D1\_ and D2\_ Gene-wise p-values.

D1\_Gene-wise p-values have been computed using VEGAS software and fisher product. D2\_Gene-wise p-values were calculated using VEGAS software only.

### Table 2. Selection criteria of the subnets

In red, the 5 selected subnets that met the criteria.

In white, excluded subnets with less than 50% of CAMs genes in the sub-network

In blue, excluded subnets with more than 50% of CAMs transcription factors among the CAMs genes in the sub-network

# Genes: Number of genes in the sub-network

# CAMs Genes: Number of genes of the network that belong to the CAMs genes list

% CAMs Genes: Ratio (# Genes / # CAMs Genes)

# CAMs TF: Number of CAMs Transcription factors in the sub-network

CAMs TF / CAMs Genes: Ratio (# CAMs TF / # CAMs Genes)

### Supplementary Table 1. Name, localization and function of selected CAMs pathway genes.

Chr: chromosome

Start: Starting Chromosomal Position in NCBI B36

End: Ending Chromosomal Position in NCBI B36

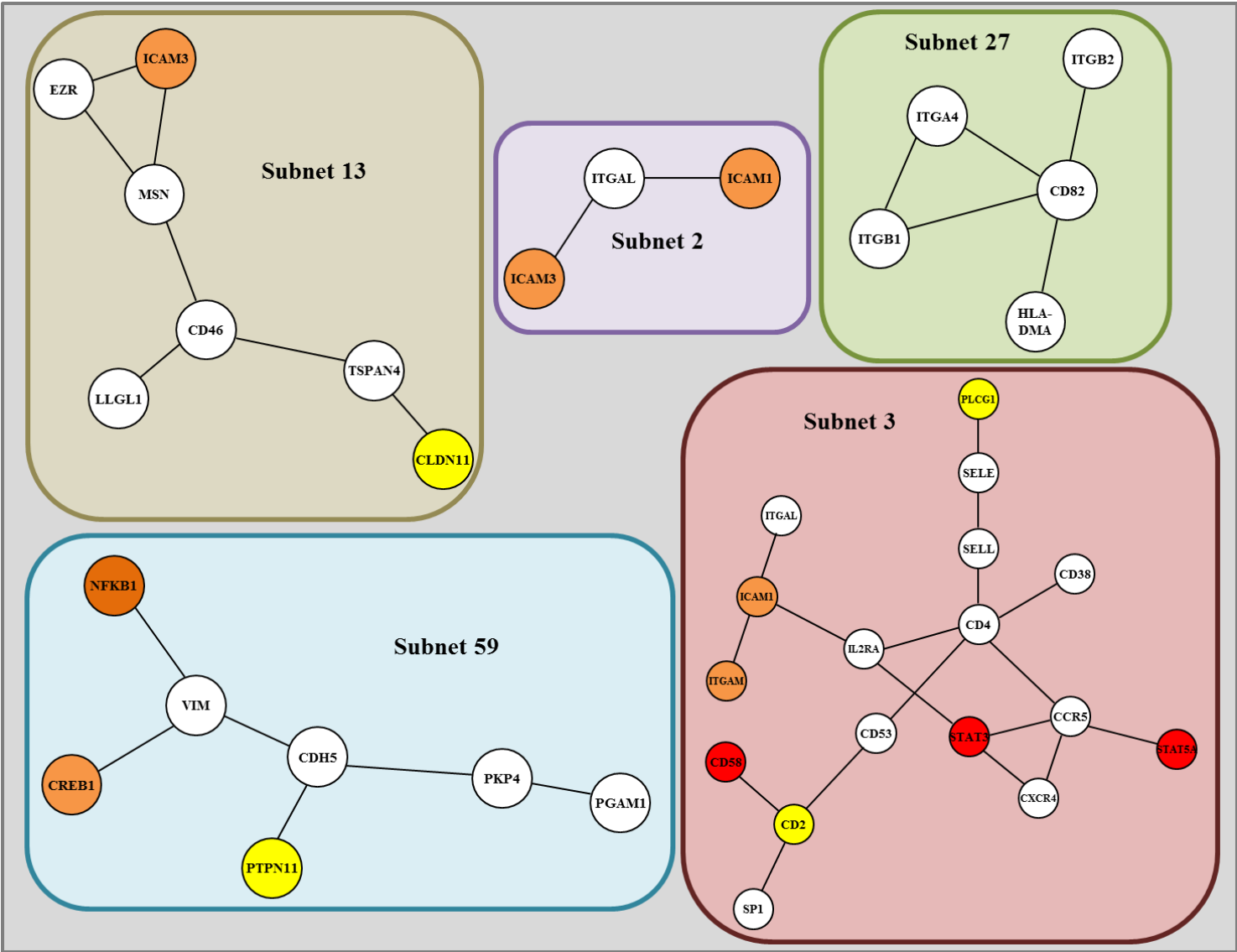
TJ: Tight Junctions

AJ: Adherens Junctions

TF: Transcription Factors

### Supplementary Table 2. Number of genotyped SNPs in each original dataset

Figure 1.



**Table 1**

Gene	D1_Gene-Wise_Pvalue	D2_Gene-Wise_Pvalue
ARHGAP1	0.9839	0.655
CD2	0.0958	0.0498
CD34	0.1544	0.217
CD58	1.895E-06	0
CDH1	0.3354	0.67
CDH5	0.5487	0.0747
CLDN11	0.414	0.0226
CLDN12	0.6658	0.43
CREB1	0.4054	0.0097
CTNNA1	0.8212	0.322
CTNNA2	0.9272	0.51
CTNNA3	0.9391	0.742
CTNNB1	0.2761	0.53
CTNND1	0.739	0.37
CXCL12	0.1875	0.351
ESAM	0.5151	0.76
EZR	0.3449	0.129
F11R	0.7609	0.55
GLG1	0.864	0.335
GNAI1	0.3653	0.754
ICAM1	0.2427	0.0018
ICAM2	0.1718	0.438
ICAM3	0.1969	0.0009
INADL	0.8822	0.983
ITGA4	0.1698	0.704
ITGA9	0.9868	0.472
ITGAL	0.0832	0.259
ITGAM	0.03	0.001
ITGB1	0.2867	0.484
ITGB2	0.7712	0.158
ITGB7	0.7891	0.11
JAM2	0.3714	0.0956
JAM3	0.3393	0.678
JUN	0.9629	0.428
MADCAM1	0.6581	0.393
MAG11	0.8826	0.211
MLLT4	0.8411	0.577
MPDZ	0.8629	0.308
NFKB1	0.0125	0.0011
NR3C1	0.503	0.118
PARD3	0.4014	0.183
PECAM1	0.3346	0.045
PLCG1	0.0434	0.0936
PLCG2	0.4254	0.0792
PPARG	0.8729	0.262
PTPN11	0.452	0.0247
PVRL1	0.5655	0.517
PVRL2	0.7898	0.253
PVRL3	0.5861	NA
RAP1A	0.4118	0.354
RAP1B	0.9663	0.757
RASSF5	0.9414	0.0698
RHOH	0.3895	0.459
SELE	0.1953	0.188
SELL	0.1925	0.194
SELP	0.4889	0.575
SELPLG	0.4166	0.608
SP1	0.1029	0.49
STAT1	0.7344	0.0374
STAT3	2.942E-05	1.5E-05
STAT5A	0.0001	4.2E-05
THY1	0.8038	NA
TJP1	0.3435	0.544
TJP2	0.1755	0.197
TJP3	0.8044	0.484
TP53	0.4371	0.407
VAV1	0.3259	0.497
VAV2	0.7844	0.33
VAV3	0.2729	0.768
VCAM1	0.7224	0.993

**Table 2.**

Sub-net_ID	# Genes	# CAMs Genes	% CAMs Genes	# CAMs TF	CAMs TF/ CAMs Genes
Sub-net 1	3	2	67	2	100
Sub-net 2	3	3	100	0	0
Sub-net 3	15	11	73	3	27
Sub-net 4	5	2	40	2	100
Sub-net 5	5	2	40	2	100
Sub-net 6	5	2	40	1	50
Sub-net 7	5	1	20	1	100
Sub-net 8	5	1	20	1	100
Sub-net 9	5	1	20	1	100
Sub-net 10	5	2	40	0	0
Sub-net 11	5	1	20	1	100
Sub-net 12	5	1	20	1	100
Sub-net 13	7	4	57	0	0
Sub-net 14	5	1	20	1	100
Sub-net 15	5	2	40	2	100
Sub-net 16	5	1	20	1	100
Sub-net 17	11	5	45	2	40
Sub-net 18	5	1	20	0	0
Sub-net 19	5	1	20	1	100
Sub-net 20	5	2	40	2	100
Sub-net 21	5	1	20	1	100
Sub-net 22	10	3	30	3	100
Sub-net 23	5	1	20	1	100
Sub-net 24	5	1	20	1	100
Sub-net 25	5	2	40	0	0
Sub-net 26	5	1	20	1	100
Sub-net 27	5	3	60	0	0
Sub-net 28	5	1	20	1	100
Sub-net 29	5	2	40	1	50
Sub-net 30	6	1	17	1	100
Sub-net 31	5	1	20	1	100
Sub-net 32	5	2	40	1	50
Sub-net 33	5	1	20	1	100
Sub-net 34	5	1	20	1	100
Sub-net 35	6	2	33	1	50
Sub-net 36	5	1	20	1	100
Sub-net 37	5	2	40	1	50
Sub-net 38	5	1	20	1	100
Sub-net 39	5	2	40	0	0
Sub-net 40	5	1	20	1	100
Sub-net 41	5	1	20	1	100
Sub-net 42	11	4	36	1	25
Sub-net 43	10	3	30	2	67
Sub-net 44	5	2	40	0	0
Sub-net 45	5	2	40	2	100
Sub-net 46	6	1	17	1	100
Sub-net 47	15	5	33	3	60
Sub-net 48	5	1	20	1	100
Sub-net 49	17	4	24	2	50
Sub-net 50	5	1	20	1	100
Sub-net 51	8	2	25	1	50
Sub-net 52	6	1	17	1	100
Sub-net 53	6	1	17	1	100
Sub-net 54	5	2	40	1	50
Sub-net 55	7	1	14	1	100
Sub-net 56	5	2	40	1	50
Sub-net 57	5	2	40	2	100
Sub-net 58	5	1	20	1	100
Sub-net 59	7	4	57	2	50
Sub-net 60	10	3	30	2	67
Sub-net 61	10	3	30	1	33
Sub-net 62	15	5	33	5	100
Sub-net 63	5	2	40	0	0
Sub-net 64	11	3	27	1	33

Supplementary Table 1

Gene	Chr	Start	End	Function
JUN/AP-1	1	59019053	59022373	TF
INADL/PATJ	1	61980737	62402177	TJ
VCAM1	1	100957885	100977187	Adhesion
VAV3	1	107915305	108309068	Adhesion
RAP1a	1	111963928	112057623	Adhesion
CD58	1	116858680	116915126	Adhesion
CD2	1	117098625	117113373	Adhesion
JAM1	1	159232608	159275358	TJ/Adhesion
SELP	1	167824714	167866031	Adhesion
SELL	1	167926434	167947461	Adhesion
SELE	1	167958406	167969803	Adhesion
RAPL/RASSF5	1	204747502	204829238	Adhesion
CD34	1	206126507	206151306	Adhesion
CTNNA2	2	79593634	80729415	AJ
CXCR4	2	136588390	136592195	Adhesion
ITGA4	2	182029864	182110711	Adhesion
STAT1	2	191542122	191587181	TF
CREB1	2	208102931	208171806	TF
PPAR-gamma	3	12304349	12450854	TF
ITGA9	3	37468817	37836283	Adhesion
CTNNA1	3	41216016	41256938	AJ/Adhesion
MAGI1	3	65314947	65999549	TJ
PVRL3	3	112273555	112335752	AJ
CLDN11	3	171619359	171634577	TJ
RHOH	4	39874997	39922248	Adhesion
NFKB1	4	103641518	103757506	TF
OCLN	5	68823875	68885886	TJ
CTNNA1	5	138117006	138298621	AJ
NR3C1	5	142637689	142795270	TF
VIL2/EZR	6	159106766	159159247	Adhesion
MLLT4/AF6	6	167970520	168115551	TJ
CLDN3	7	72821264	72822512	TJ
GNAI1/Gi	7	79602076	79686657	Adhesion
CLDN 12	7	89870732	89883202	TJ
MUPP1/MPDZ	9	13096578	13240357	TJ
ZO-2/TJP2	9	70978909	71059938	TJ
VAV2	9	135616838	135847226	Adhesion
ITGB1	10	33229327	33287204	Adhesion
PAR3/PARD3	10	34440104	35143929	TJ
CXCL12/SDF-1	10	44185613	44200548	Adhesion
CTNNA3	10	67349937	69125933	AJ
RhoGAP/ARHGAP1	11	46655208	46678696	Adhesion
CTNND1/p120CTN	11	57285845	57343226	AJ
Thy1	11	118794098	118799064	Adhesion
PVRL1	11	119014018	119104645	AJ
ESAM	11	124128242	124137385	TJ
JAM3	11	133444030	133526859	TJ/Adhesion
ITGB7	12	51871375	51887267	Adhesion
Sp1	12	52060246	52096491	TF
RAP1b	12	67290919	67340642	Adhesion
SELPLG	12	107539811	107551799	Adhesion
SHP-2/PTPN11	12	111340919	111432099	Adhesion
ZO-1/TJP1	15	27779651	27901998	TJ/Adhesion
ITGAL	16	30391572	30442006	Adhesion
ITGAM	16	31178789	31251712	Adhesion
CDH5	16	64958064	64996189	AJ
CDH1	16	67328696	67426945	AJ
GLG1	16	73043357	73198518	Adhesion
PLCG2	16	80370431	80549399	Adhesion
TP53	17	7512445	7531642	TF
STAT5A	17	37693091	37717484	TF
STAT3	17	37718869	37794039	TF
ICAM2	17	59433688	59437839	Adhesion
PECAM1	17	59753596	59817743	AJ/Adhesion
MAcCAM1	19	447490	456342	Adhesion
ZO-3/TJP3	19	3679374	3701681	TJ
VAV1	19	6723722	6808371	Adhesion
ICAM1	19	10242779	10258291	Adhesion
ICAM3	19	10305454	10311300	Adhesion
PVRL2	19	50041233	50084325	AJ
PLCG1	20	39199575	39237771	Adhesion
JAM2	21	25933460	26009106	TJ/Adhesion
ITGB2	21	45130299	45165303	Adhesion
CLDN5	22	17890550	17892860	TJ
CD99	X	2619228	2669348	AJ/Adhesion
MSN	X	64804236	64878517	Adhesion
CD99	Y	2619228	2669348	AJ/Adhesion

Supplementary Table 2.

Gene	ANZGene	BWH/MIGEN	GeneMSA DU	GeneMSA SW	GeneMSA US	IMSGC US	IMSGC UK	WTCCC2
ARHGAP1	1	1	0	0	0	0	0	1
CD2	4	4	2	2	2	1	0	2
CD34	6	15	8	8	8	11	11	8
CD58	3	3	3	3	3	2	2	3
CDH1	8	29	17	17	17	19	19	16
CDH5	6	10	12	12	12	5	5	10
CLDN11	2	4	4	4	4	3	3	4
CLDN12	1	3	3	3	3	2	2	2
CREB1	2	5	5	5	5	2	2	4
CTNNA1	6	20	11	11	11	19	19	10
CTNNA2	216	429	403	403	404	233	233	369
CTNNA3	241	539	434	435	436	319	314	404
CTNNB1	2	6	4	4	4	3	3	4
CTNND1	4	3	5	5	5	2	2	5
CXCL12	6	5	8	8	8	3	3	6
ESAM	2	3	4	4	4	2	2	4
EZR	6	13	10	10	10	13	13	10
F11R	8	6	7	7	7	6	6	8
GLG1	5	25	7	7	7	21	21	7
GNAI1	10	23	19	19	19	10	10	18
ICAM1	6	0	8	8	8	0	0	8
ICAM2	0	0	1	1	1	0	0	1
ICAM3	8	3	9	9	9	2	2	10
INADL	59	105	87	88	88	58	58	82
ITGA4	14	23	23	23	23	15	15	23
ITGA9	56	137	95	95	95	83	83	94
ITGAL	8	3	11	11	11	1	1	10
ITGAM	7	7	7	7	7	3	3	6
ITGB1	9	8	12	12	12	8	8	11
ITGB2	14	14	18	18	18	9	9	18
ITGB7	1	2	1	1	1	2	2	1
JAM2	10	21	14	14	14	16	16	13
JAM3	9	24	16	16	16	14	14	14
JUN	1	0	1	1	1	0	0	1
MADCAM1	3	0	3	3	3	0	0	2
MAG1	117	255	216	216	216	151	150	204
MLLT4	9	17	12	12	12	12	12	12
MPDZ	15	24	25	25	25	14	14	24
NFKB1	7	22	15	15	15	19	21	14
NR3C1	12	14	15	15	15	7	7	14
PARD3	64	165	109	109	109	99	99	107
PECAM1	5	14	9	9	9	9	9	9
PLCG1	2	3	4	4	4	2	2	3
PLCG2	35	106	95	95	95	55	55	94
PPARG	13	35	23	23	23	24	24	22
PTPN11	4	5	5	5	5	2	2	5
PVRL1	5	17	18	18	18	11	11	16
PVRL2	6	7	12	12	12	5	5	12
PVRL3	0	4	0	0	0	3	3	0
RAP1A	13	21	17	17	17	16	16	16
RAP1B	3	3	3	3	3	1	1	3
RASSF5	9	18	11	11	11	15	15	11
RHOH	5	14	9	9	9	4	4	9
SELE	11	36	13	13	13	28	28	13
SELL	11	36	13	13	13	28	28	13
SELP	20	26	26	26	26	18	18	22
SELPLG	2	4	2	2	2	4	4	2
SP1	2	2	3	3	3	2	2	3
STAT1	6	17	13	13	13	10	10	12
STAT3	4	16	6	6	6	13	12	6
STAT5A	3	11	5	5	5	8	8	5
THY1	0	2	0	0	0	1	1	0
TJP1	8	16	12	12	12	8	8	11
TJP2	17	39	19	19	19	29	29	17
TJP3	9	4	8	8	8	0	0	7
TP53	3	6	3	3	3	3	3	4
VAV1	14	19	29	29	29	6	6	29
VAV2	50	64	76	76	76	31	31	72
VAV3	56	134	99	101	99	94	94	84
VCAM1	3	8	8	8	8	5	5	7



## 2.6. Hypothèse 6 : La maladie s'explique à l'échelle individuelle

Pour la plupart des variants de risque identifiés à ce jour dans la SEP, près de la moitié des patients ne possède pas l'allèle à risque d'un polymorphisme donné. Cette constatation aboutit à la conclusion que chaque patient atteint de SEP porte une combinaison particulière de variants à risque mais ne porte pas la totalité des variants identifiés. Cette proportion d'allèle à risque absente chez certains sous-groupes de patients ou « héritabilité manquante individuelle » pourrait être un leurre et ne pas exister. En effet, à ce jour, aucune étude d'interaction génique sans à priori « *pathway* » n'a pu être menée faute de méthode statistique puissante appropriée.

Il est aujourd'hui impossible d'étudier toutes les interactions aléatoires possibles entre les 110 variants, soit près de  $3 \times 10^{52}$  tests si on considère les combinaisons de génotypes et  $1,7 \times 10^{66}$  tests si on considère les combinaisons d'allèles. D'autre part, cette étude à l'échelle de la population ne donnerait probablement aucun résultat, étant donné que peu de patients et de contrôles présenteraient une combinaison commune de variants. Les études statistiques en résultant ne présenteraient que peu de puissance.

Des approches individuelles à but prédictif ou diagnostique, intégrant les données génétiques des variants identifiés ont été menées au cours des dernières années sans réel succès (De Jager, Chibnik et al. 2009, Gourraud, McElroy et al. 2011, Wang, Pappas et al. 2011).

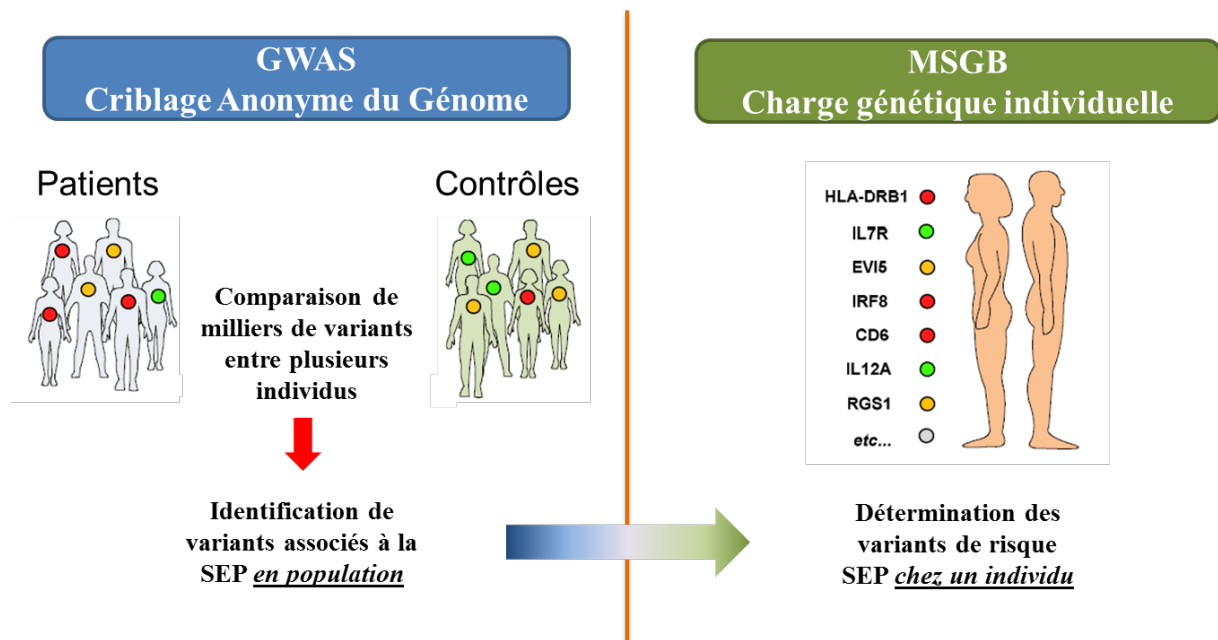
Au cours de mon doctorat, j'ai réalisé une étude d'agrégation à l'échelle familiale, étape intermédiaire entre la population et l'individu, présentée ci-après.

### 2.6.1. Les études individuelles de prédiction

Les mesures les plus simples reflétant la précision de la prédiction sont la *sensibilité* et la *spécificité* du test appliqué, défini respectivement comme la proportion des patients classés comme étant à haut risque et la proportion des individus sains classés comme étant à faible risque. La **courbe ROC** (Receiver Operating Characteristic) est souvent utilisée pour représenter cela. Elle représente le taux de vrais positifs en fonction du taux de faux positifs. L'aire en dessous de la courbe (aussi appelé AUC, Area Under the ROC Curve) a la propriété d'être égale à la probabilité qu'un patient sélectionné au hasard ait un score plus important qu'un individu sain sélectionné au hasard (Jostins and Barrett 2011).

Dans la SEP, la première étude de prédiction a été publiée par De Jager et al. en 2009 (De Jager, Chibnik et al. 2009). Par l'utilisation d'un modèle log-additif et en prenant en



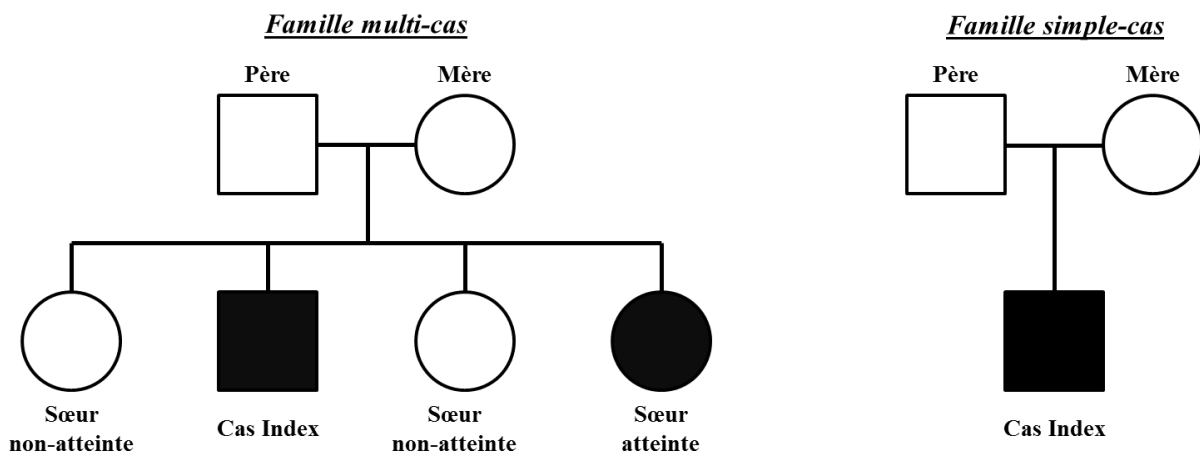


**Figure 20. Concept de charge génétique individuelle.**

Le score MSGB (Multiple Sclerosis Genetics Burden) est basé sur un algorithme qui intègre, suivant un modèle additif, chaque variant de risque porté par un individu donné ainsi que l'effet qui leur est associé. Ainsi le score MSGB représente le risque génétique SEP pour chaque individu.

GWAS : Genome-Wide Association Studies

Adapté de (Gourraud, Harbo et al. 2012)



**Figure 21. Nomenclature famille multi-cas et famille simple-cas.**

Une famille multi-cas possède au moins deux apparentés au premier degré atteints.

Une famille simple-cas ne possède qu'un seul individu atteint.

compte les 16 marqueurs connus en 2009, les auteurs ont calculé pour chaque individu un unique score, le wGRS pour weighted Genetic Risk Score. Ce score prend en compte le nombre d'allèles à risque portés par un individu et le risque associé à chacun de ces allèles ou Odds Ratio (OR). Cette étude a conclu que la comparaison des wGRS entre une population de patients et de contrôles n'était que très modestement prédictive.

Une autre étude publiée en 2011 par Wang et al. s'est aussi intéressée à la prédiction de la SEP (Wang, Pappas et al. 2011). En partant des données de la méta-analyse v1.0, Wang et al. ont identifié un ensemble de 350 variants génétiques pouvant être utilisés pour la prédiction de la SEP. Les résultats de cette étude ont montré que l'utilisation de ces 350 variants augmentait la sensibilité et la spécificité de la prédiction, sans néanmoins atteindre un seuil de prédiction suffisant à l'échelle individuelle.

### *2.6.2. Les études familiales d'agrégation*

En 2011, Gourraud et al. (Gourraud, McElroy et al. 2011) ont proposé une approche identique mais philosophiquement différente de l'utilisation du wGRS De Jager et al. (De Jager, Chibnik et al. 2009), non pas dans un but prédictif mais plus dans l'optique de déterminer un profil génétique de risque individuel. Ce score, appelé le MSGB (Multiple Sclerosis Genetic Burden), reflète ainsi l'agrégation des variants à risque à l'échelle individuelle (**Figure 20**). De manière intéressante, en prenant en compte les 16 variants de risques connus en 2011, les auteurs ont montré que les individus des familles multi-cas, dans lesquelles au minimum deux apparentés du premier degré sont atteints, avaient un plus fort MSGB, c'est à dire une plus forte agrégation de variants de risque, que les individus des familles simple-cas, où un seul individu est atteint (voir **Figure 21** pour des exemples d'arbres de familles simple-cas et multi-cas). De plus, le groupe composé des frères et/ou sœurs atteints d'un cas index<sup>1</sup> présente une charge génétique similaire à celle du groupe des cas index et d'autre part, le groupe composé des frères et/ou sœurs non atteints du cas index présente une charge génétique plus faible que celle du groupe des cas index. Cependant, cette différence n'est pas utilisable pour faire de la prédiction au sein de la famille d'un individu atteint. De plus, aucune différence de MSGB n'a été mise en évidence entre les différentes formes d'évolution de SEP, en comparant les formes RR+SP vs les formes PP.

Cette analyse MSGB a permis pour la première fois de mettre en évidence une différence de charge génétique, c'est à dire d'agrégation de variants à risque en fonction de

---

<sup>1</sup> Cas index : Patient choisi au hasard dans la fratrie ou patient par qui le recrutement de la famille s'est opéré



l'histoire familiale de la SEP, sans cependant conclure à une possible utilisation du MSGB pour la prédiction du développement de la maladie chez les apparentés d'un cas index.

### *2.6.3. Limites aux tests d'agrégation*

Très récemment, Wray et al. (Wray, Yang et al. 2013), ont mis en avant les limites des tests de prédiction qui agrègent les risques associés à chaque SNPs :

- la part environnementale de la maladie n'étant pas prise en compte, une prédiction uniquement basée sur des données génétiques ne pourrait pas caractériser la variation phénotypique dans sa totalité. Cette prédiction ne serait que probabiliste et ne pourrait être utile qu'à l'échelle d'un groupe plutôt qu'à l'échelle individuelle

- comme mentionné avant dans ce manuscrit, les SNPs mis en évidence ne sont probablement pas les variants causaux, et ne pourraient donc pas expliquer à eux seuls, toute la variation génétique de la maladie

- la variation génétique serait d'autant plus exacte que les effets attribués à chaque SNP seraient estimés de manière précise, grâce à l'augmentation de la taille des populations étudiées. En effet, les études étant faites sur des échantillons de tailles finies, les effets associés ne sont que des estimations avec des limites inférieures et supérieures.

Dans cette même revue, les auteurs ont souligné plusieurs pièges à éviter dans la mise en place de telles études :

- les échantillons de découverte et de validation du test doivent être indépendant afin de conclure à son efficacité. De plus, les valeurs de prédiction étant grandement influencées par le nombre de SNPs incluent dans l'étude, les prédictions ne seront définitives qu'après l'identification de l'ensemble des variants génétique de prédisposition à une maladie

- le taux de relations familiales entre la population de découverte et la population de réplication doit être identique au taux de relations familiales entre la population de réplication et la population cible.

- la prédiction sera moins précise si les échantillons de découverte et de validation ne possèdent pas un pattern de stratification similaire à celle de la population cible de la prédiction.

Selon Wray et al, les scores de prédiction permettront d'identifier des groupes d'individus à risque, mais n'aboutirons probablement pas à une prédiction individuelle.



**Au cours de mon doctorat, lors d'un stage à UCSF, je me suis intéressé à répliquer les résultats obtenus par Gourraud et al. en 2011, en prenant en compte les 64 variants de prédisposition connus en 2012 (*Publication Isobe\*, Damotte\*, Lo Re\* et al. Genes Immun. 2013, présentée page 207*)**



**Article (co-1<sup>er</sup> auteur):**

Genetic burden in multiple sclerosis families

Noriko Isobe\*, **Vincent Damotte**\*, Enza Lo Re\*, Maria Ban, Derek Pappas, Lena Guillot-Noel, Isabelle Rebeix, Alastair Compston, Thomas Mack, Wendy Cozen, Bertrand Fontaine, Stephen L. Hauser, Jorge R. Oksenberg, Stephen Sawcer, Pierre-Antoine Gourraud

\* equal contribution

*Genes Immun. 2013 Aug 1. doi: 10.1038/gene.2013.37*





**Introduction** : Depuis l'étude MSGB originelle de 2011 (Gourraud, McElroy et al. 2011), de nouveaux variants ont été identifiés par le GWAS de l'IMSGC (IMSGC, WTCCC2 et al. 2011) et par une nouvelle méta-analyse (Patsopoulos, Esposito et al. 2011). Nous avons donc entrepris en collaboration avec l'équipe de San Francisco ayant publié la première étude MSGB et une équipe de Cambridge, de confirmer les résultats obtenus en 2011 en prenant en compte les nouveaux variants et en étendant les analyses à des familles Américaines, Françaises et Anglaises.

**Matériel et Méthodes** : 809 familles multi-cas, 2013 familles simple-cas provenant de 3 cohortes (française, américaine et anglaise) ainsi que 708 contrôles et 380 cas sporadiques américains ont été génotypés pour 64 marqueurs SEP. Un score de charge unique, le MSGB, a été calculé pour chaque individu suivant un modèle log-additif prenant en compte le nombre d'allèles à risques portés par un individu pour chaque variant de risque et l'Odds Ratio qui lui est associé.

**Résultats** : Nous avons confirmé une différence de charge génétique entre les individus des familles multi-cas et simple-cas, l'absence de prédiction précise du statut dans les fratries mais aussi dans la population générale et enfin l'absence de différence de charge génétique entre les différentes formes de SEP. La prise en compte d'un plus grand nombre de variants de prédisposition génétique à la maladie (64 vs 16) dans cette étude montre d'autre part une meilleure discrimination des patients et des contrôles en terme de charge génétique.

**Conclusions** : Le fait que les allèles de risque soient fréquents en population générale entraîne sans doute la modeste prédiction observée au sein d'une fratrie ou dans la population générale, car ils ne sont pas spécifiques ni aux patients SEP, ni à leurs apparentés. Par ailleurs, l'augmentation du nombre de variants de prédisposition pris en compte (64 vs 16) n'a pas permis d'augmenter significativement l'efficacité de la prédiction.

Un seuil acceptable d'efficacité de la prédiction pourra peut-être être atteint si des variants rares à fort effet individuel, associés à la SEP, sont identifiés et pris en compte dans le calcul du MSGB.

**Ma contribution** : Ma participation à ce projet s'est faite en deux étapes :

- En France, j'ai génotypé les échantillons français sur la plateforme OpenArray® (Applied Biosystems) et effectuer les contrôles qualités sur ces échantillons



- À San Francisco, j'ai effectué les analyses statistiques sur les familles françaises, américaines et anglaises (Figures 1, 4A, 4B et Figure supplémentaire 1)

J'ai participé à l'écriture de l'article et à sa relecture.



ORIGINAL ARTICLE

# Genetic burden in multiple sclerosis families

N Isobe<sup>1,6</sup>, V Damotte<sup>1,2,6</sup>, V Lo Re<sup>3,6</sup>, M Ban<sup>3</sup>, D Pappas<sup>1</sup>, L Guillot-Noel<sup>2</sup>, I Rebeix<sup>2,4</sup>, A Compston<sup>3</sup>, T Mack<sup>5</sup>, W Cozen<sup>5</sup>, B Fontaine<sup>2,4</sup>, SL Hauser<sup>1</sup>, JR Oksenberg<sup>1</sup>, S Sawcer<sup>3</sup> and P-A Gourraud<sup>1</sup>

A previous study using cumulative genetic risk estimations in multiple sclerosis (MS) successfully tracked the aggregation of susceptibility variants in multi-case and single-case families. It used a limited description of susceptibility loci available at the time (17 loci). Even though the full roster of MS risk genes remains unavailable, we estimated the genetic burden in MS families and assess its disease predictive power using up to 64 single-nucleotide polymorphism (SNP) markers according to the most recent literature. A total of 708 controls, 3251 MS patients and their relatives, as well as 117 twin pairs were genotyped. We validated the increased aggregation of genetic burden in multi-case compared with single-case families ( $P = 4.14e - 03$ ) and confirm that these data offer little opportunity to accurately predict MS, even within sibships (area under receiver operating characteristic (AUROC) = 0.59 (0.55, 0.53)). Our results also suggest that the primary progressive and relapsing-type forms of MS share a common genetic architecture ( $P = 0.368$ ; difference being limited to that corresponding to  $\pm 2$  typical MS-associated SNPs). We have confirmed the properties of individual genetic risk score in MS. Comparing with previous reference point for MS genetics (17 SNPs), we underlined the corrective consequences of the integration of the new findings from GWAS and meta-analysis.

*Genes and Immunity* advance online publication, 1 August 2013; doi:10.1038/gene.2013.37

**Keywords:** multiple sclerosis; family study; genetic risk

Extensive epidemiological and laboratory data confirm that genetic variation is an important contributor to risk in multiple sclerosis (MS), a chronic and severe neuro-inflammatory disease that is seen most commonly in Whites of northern European descent.<sup>1</sup> As with many other multifactorial diseases, genome-wide association studies (GWAS) have been highly successful in identifying genomic regions associated with susceptibility. To date, multiple MS GWAS have been completed, and between them have identified nearly 60 susceptibility genomic regions. Given the limited power of this approach to identify rare variants, it is unsurprising that the majority of these associations relates to variants that are common in the general population.<sup>2–13</sup> The significant enrichment for genes with annotated immunological functions is supported by multiple statistical measures<sup>4,14</sup> with only a minority of the candidate genes implicated by these associations having bona fide neuronal functions independent of inflammation. The strongest signal consistently maps to the *HLA-DRB1* gene in the class II region of the major histocompatibility complex (MHC, 6p21.3). Extensive overlap with susceptibility markers for other autoimmune diseases has been well established,<sup>15,16</sup> and follow-up experiments provided important mechanistic insights linking associated variants in *IL7R*, *IL2RA*, *CDS8*, *TNFRSF1A* and *TYK2*<sup>(refs 17–21)</sup> to their functional consequences. These results significantly broaden our understanding of etiology but the full potential of these data to model an individual's lifetime disease risk remains unknown.

A previous attempt to use cumulative genetic risk estimations in MS suggested a higher aggregation of susceptibility variants in multi-case compared with single-case MS families, but resulted in

incomplete affectation status classification accuracy.<sup>22–24</sup> A number of factors could have contributed to the low sensitivity of the genomic burden metrics, including the limited description of susceptibility loci available at the time (17 loci). Even though the full roster of MS risk genes remains unavailable, we build on the most recent data sets<sup>4,25</sup> to estimate the genetic burden in MS families and assess its disease predictive power. We validate the increased aggregation of genetic burden in multi-case compared with single-case families and confirm that these data offer little opportunity to accurately predict MS, even within sibships. We also provide additional genetic evidence for the absence of any meaningful difference in the genetic burden between the primary progressive and relapsing-type forms of MS.

## RESULTS

### Distribution of MS genetic burden (MSGB) scores in multi-case and single-case MS families

The genetic risk captured by the MSGB score was higher in the probands and parents from the multi-case families compared with those with unaffected first-degree relatives ( $P$ -values for probands =  $4.14e - 03$ ,  $P$  for mothers =  $1.16e - 02$  and  $P$  for fathers =  $3.13e - 02$ ) (Figure 1 and Table 1). These results are consistent with those we reported previously using just 17 markers<sup>23</sup> and in line with the increased genetic information conferred by the inclusion of additional MS-associated single-nucleotide polymorphisms (SNPs), which affords more statistical significance when compared with controls. Using the maximum number of MS risk variants genotyped in the University of

<sup>1</sup>Department of Neurology, University of California at San Francisco, San Francisco, CA, USA; <sup>2</sup>UPMC-INSERM-CNRS-ICM, UMR 975-7225, Institut du Cerveau et de la Moelle épinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France; <sup>3</sup>Department of Clinical Neurosciences, Addenbrooke's, Hospital, University of Cambridge, Cambridge, UK; <sup>4</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Département de Neurologie, Hôpital Pitié-Salpêtrière, Paris, France and <sup>5</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. Correspondence: Dr P-A Gourraud, Department of Neurology, University of California at San Francisco, 675 Nelson Rising lane, Room NS235, San Francisco, CA 94158, USA.

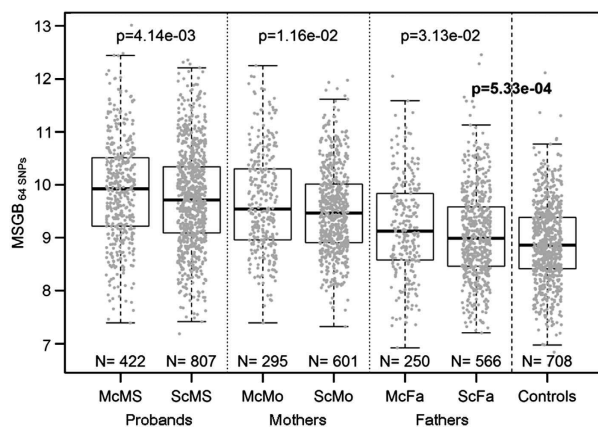
E-mail: Pierreantoine.Gourraud@ucsf.edu

<sup>6</sup>These authors contributed equally to this work.

Received 21 February 2013; revised 15 May 2013; accepted 18 June 2013

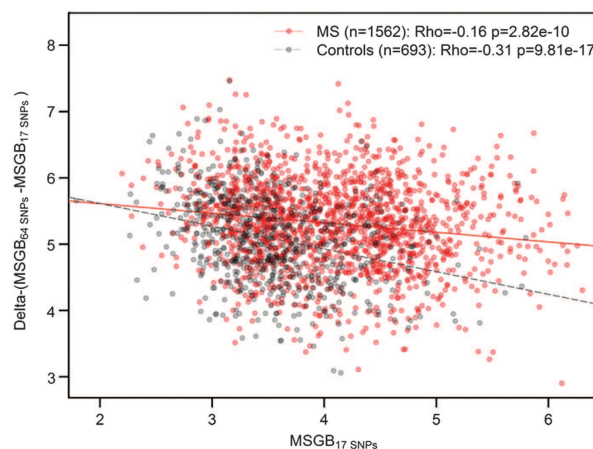
California San Francisco (UCSF) samples, we did not observe any indication of balancing aggregation (Supplementary Figure 1), meaning that neither the presence of the homozygous major risk factor (*HLA-DRB1\*15:01*) nor female status results in any trend toward lower accumulation of the other non-MHC genetic risk variants. However, power computations suggest that we would need ~2.5 times as many samples to reach 80% power for detecting an one-sided statistically significant difference of half the one observed between the group of *HLA-DRB1\*15:01* homozygous female and the group of *HLA-DRB1\*15:01*-negative males. So this negative observation is not unexpected.

To further assess the effect of the additional information content on the individual risk, we plotted  $MSGB_{17SNPs}$  (previous reference point for MS genetics) vs the new part of  $MSGB_{64SNPs}$  (most up-to-date MS Genetics  $MSGB_{64SNPs}$  minus previously used



**Figure 1.** MSGB score differentiates multi-case from single-case MS in UCSF families. The distribution of MSGB is presented using box plots. MSGB is computed using components derived from gender, MHC and non-MHC SNPs. Gray dots represent the MSGB of an individual subject. Groups separated by dotted lines (probands, mothers of probands, fathers of probands and unrelated controls) are divided into multi-case and single-case samples. Spouses of patients were considered genetically unrelated controls. Sample sizes are indicated at the bottom of each box plot. *P*-values in each of the three left panels indicate the significance of Wilcoxon's tests of the null hypothesis that MSGB of members of multi-case families are greater than those of members of single-case MS families (*P*-values without affected mothers and fathers: McMo vs ScMo =  $6.63e-02$  and McFa vs ScFa =  $3.47e-02$ ). The *P*-value in the right panel corresponds to the test that the MSGB of fathers of single-case MS patients is greater than unrelated controls (Wilcoxon's test). Spouses as healthy unrelated individuals are taken as controls. Fa, father; Mc, multi-case; Mo, mother; Sc, single-case.

$MSGB_{17SNPs}$ ) using UCSF samples typed for SNPs required for preparation of both MSGB scores (Figure 2). Surprisingly, this analysis yielded a statistically significant negative correlation in both, cases and controls ( $\rho = -0.16$  in cases  $P = 2.82e-10$  and  $\rho = -0.31$  in controls  $P = 9.81e-17$ ). It shows that our knowledge of MS genetic not only grows in number of regions identified but also refines the association signal in previously identified regions. Several factors are likely to contribute to this correlation. First, the lack of confirmation for *CD226* and *GPC5* in the recent GWAS<sup>4,25</sup> suggest that the previous analyses may have been confounded by the inclusion of unassociated variants. In line with the winners curse, the effect sizes were reduced for 13 SNPs but increased for only 2 (*HLA-DRB1* and *EVI5*). In two regions (*IL2RA* and the MHC), we included some degree of genetic heterogeneity with multiple SNPs capturing second signals. Furthermore, updating the MSGB model with the latest results obtained with independent samples from these family collections not only resulted in the addition of lower effect-size common genetic variants but also tended to replace the low-frequency variants with more common ones at the previously identified genomic

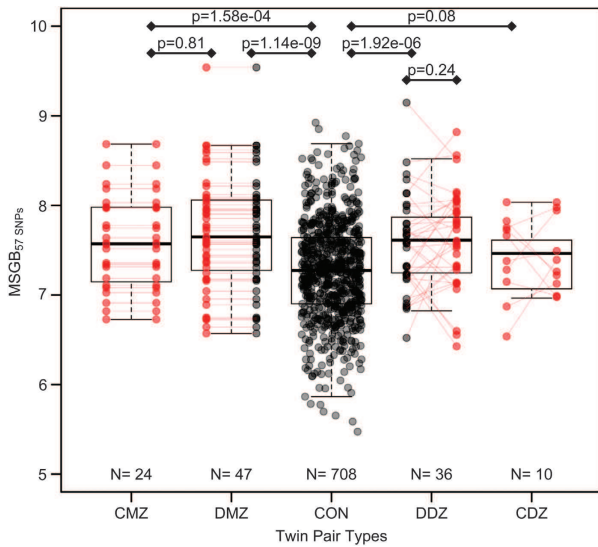


**Figure 2.** Characteristics of the new part of MSGB scores compared with the previous MSGB scores with 17 SNPs. y axis represents the new and eventually corrective part of MSGB scores calculated by subtracting  $MSGB_{17SNPs}$  from  $MSGB_{64SNPs}$ . Similar correlation are obtained by plotting  $MSGB_{17SNPs}$  vs  $MSGB_{64SNPs}$  (data not shown). UCSF samples with available MSGB scores for both  $MSGB_{17SNPs}$  and  $MSGB_{64SNPs}$  are included. For MS patients, probands of the family data set and cases in the case-control data set are enrolled. Negative correlation between  $MSGB_{17SNPs}$  and new part of MSGB ( $MSGB_{64SNPs} - MSGB_{17SNPs}$ ) were seen for both MS patients and controls but controls had significantly stronger negative correlation than patients.

**Table 1.** MSGB differentiates multi-case families from single-case families

MSGB components median, (p25–p75)	Multi-case probands n = 422	Single-case probands n = 807	Multi-case parents n = 545	Single-case parents n = 1167	Controls n = 708	Cuzick z P
Gender, HLA, non-MHC SNPs	9.93 (9.22–10.51)	9.71 (9.09–10.34)	9.35 (8.77–10.10)	9.23 (8.70–9.83)	8.86 (8.42–9.38)	$z = -23.07$ , $P < 1e-04$
HLA, non-MHC SNPs	9.53 (8.92–10.09)	9.33 (8.76–9.94)	9.09 (8.55–9.83)	8.99 (8.45–9.56)	8.61 (8.14–9.10)	$z = -17.73$ , $P < 1e-04$
Gender, non-MHC SNPs	8.84 (8.45–9.29)	8.89 (8.45–9.28)	8.64 (8.24–9.07)	8.64 (8.21–9.06)	8.48 (8.04–8.86)	$z = -17.73$ , $P < 1e-04$
Non-MHC SNPs	8.50 (8.09–8.90)	8.54 (8.10–8.90)	8.42 (7.98–8.79)	8.36 (8.00–8.79)	8.22 (7.80–8.62)	$z = -8.93$ , $P < 1e-04$

Abbreviations: HLA, human leukocyte antigen; MSGB, Multiple Sclerosis Genetic Burden; SNP, single-nucleotide polymorphism.

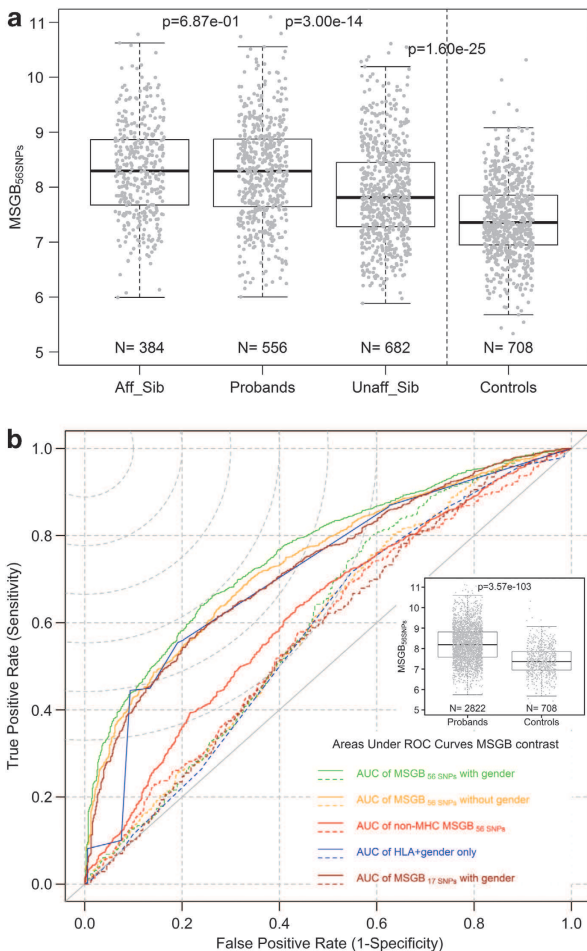


**Figure 3.** Comparison of MSGB scores in monozygotic and dizygotic twin pairs. The distribution of MSGB is represented using box plots. The calculated MSGB does not account for gender or the presence of *HLA-DRB1\*15:01*. Black dots represent the MSGB of unaffected twins or controls; red dots represent the MSGB of affected twins. Individual twin pairs are connected with a red line. *P*-values between adjacent groups or indicated groups (black horizontal line) represent the significance from a Wilcoxon's test. CDZ, concordant dizygotic; CMZ, concordant monozygotic; DDZ, discordant dizygotic; DMZ, discordant monozygotic.

regions because the SNP with the lowest *P*-value was chosen. For example, in the *TYK2* locus, the MS risk SNPs rs34536443 (NP\_003322.3;p.Pro1104Ala – MAF = 0.0285) has been replaced by rs8112449 (NT\_011295.11:g.1782866G>A – MAF = 0.305 in dbSNP). Several changes resulting from the redefinition of SNPs considered to represent the most up-to-date definition of the MSGB probably explain the negative correlation in both cases and controls. Supplementary Figure 2 presents the MSGB score distribution in both cases and controls when increasing the number of SNPs. While the difference between cases and controls increases, there is also a noteworthy increase in the variance of the scores in both cases and controls (0.53 vs 0.74,  $P = 2.09e - 04$  in cases, 0.37 vs 0.55,  $P = 2.19e - 05$  in controls; Supplementary Table 3).  $MSGB_{64SNPs}$  shows larger variance than  $MSGB_{17SNPs}$  in both cases and controls and the skewness of the distribution is reduced in  $MSGB_{64SNPs}$  (from 0.40 to 0.10 in cases, and from 0.73 to 0.25 in controls) showing that the combination of 62 non-MHC SNPs with high frequencies and modest effect sizes in the scores smoothen the distorted distribution observed in  $MSGB_{17SNPs}$ .

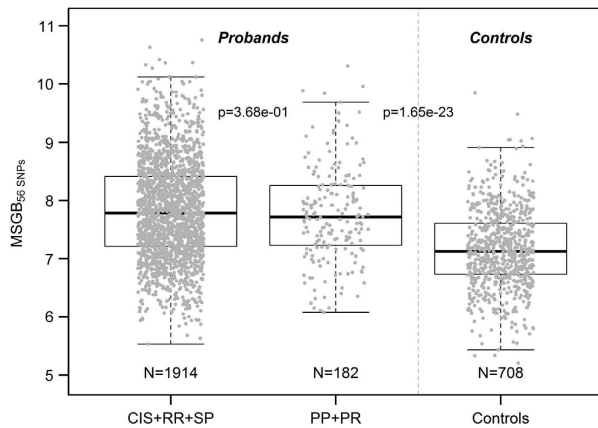
Distribution of MSGB scores in twin and sib pairs

We explored the properties of the MSGB statistics (including HLA and gender) in three additional familial data sets, including 117 twin pairs (Supplementary Table 4, Figure 3). In this analysis, data from 57 markers were available (Supplementary Table 1). As expected, MS twins have higher MSGB compared with controls ( $P = 1.14e - 09$  vs MZ twins pairs and  $P = 1.92e - 06$  for DZ pairs).



**Figure 4.** (a) Distribution of MSGB in siblings of MS families. Computations have been done with the 56 SNPs typed in common with all data sets. Distribution of MSGB in siblings of UCSF and French MS multi-case families using box plots. MSGB is computed using components derived from gender, MHC and non-MHC SNPs. Gray dots correspond to the MSGBs of individual subjects. The three left box plots correspond to subjects' status in sibship (Aff\_Sib = affected sibs; Unaff\_Sib = unaffected sibs). Sample sizes are indicated at the bottom of each box plot. The first *P*-value corresponds to the test that MSGBs of affected sibs are different from MSGBs of probands (Wilcoxon's test). The second *P*-value corresponds to the test that the MSGBs of multi-case probands is greater than MSGBs of unaffected sibs of probands (Wilcoxon's test). The *P*-value overlaying the dotted line indicates the significance of Wilcoxon tests of the null hypothesis that MSGBs of unaffected siblings of probands are greater than the MSGBs of the controls. (b) ROC curves for MS prediction comparing achievement of various MSGBs sib-proband contrasts in sibships compared with the direct use of MSGB scores as predictors in general population. Computations have been done with up to 56 SNPs typed in common with all data sets. ROCs corresponding to: the prediction of MS status of the sibs of the probands from UCSF and French multi-case families based on the contrast between sib's and proband's MSGB (dotted lines); only UCSF multi-case families are used for the brown line; the prediction of MS status of the general population based on the contrast between unrelated UCSF controls and UCSF, French and Cambridge multi-case and single-case families probands (full lines); only UCSF samples are used for the brown line. In green, MSGB contrasts are computed using the gender, MHC and non-MHC SNPs components. In orange, MSGB contrasts are computed using the MHC and non-MHC SNPs components. In red, MSGB contrasts are computed using the only the non-MHC SNPs components. In blue, MSGB contrasts are computed using the gender and the MHC components. In brown, MSGB contrasts are computed using the MSGB values of the previously published study (Gourraud *et al.*<sup>23</sup>) only for UCSF samples of families and case-control data set. The inset corresponds to the distribution of MSGB in probands from UCSF, French and Cambridge multi-case and single-case families and in unrelated UCSF controls, using box plots. Sample sizes are indicated at the bottom of each box plot. The *P*-value corresponds to the test that the MSGBs of probands are greater than MSGBs of controls (Wilcoxon's test). AUC, area under the curve; HLA, human leukocyte antigen.





**Figure 5.** Absence of different MSGB scores calculated with 56 SNPs between patients with relapsing-type MS vs patients with primary progressive MS. The distribution of  $MSGB_{56SNPs}$  scores are shown on y axis. The  $P$ -value in probands indicates the significance of Wilcoxon tests of the null hypothesis that  $MSGB_{56SNPs}$  of relapsing-type MS (CIS + RR + SP) patients are not different from those of primary progressive (PP + PR) patients. The  $P$ -value in the right part of the figure corresponds to the Wilcoxon test for the null hypothesis that the  $MSGB_{56SNPs}$  of PP + PR MS patients are not different from those of unrelated controls. CIS, clinically isolated syndrome; PP, primary progressive; PR, progressive relapsing; RR, relapsing remitting; SP, secondary progressive.

Given the limited power of this data set, it is unsurprising that we saw no statistically significant difference in MSGB score when comparing concordant with discordant monozygotic twins ( $P=0.81$ , estimated power = 30%), nor between discordant dizygotic twins ( $P=0.24$ , estimated power = 61%).

In siblings of the same sibship, having a greater or equal  $MSGB_{56SNPs}$  than the proband is significantly associated with MS with an odds ratio (OR) = 2.10 (1.4, 3.1) (conditional logistic regression  $P=1e-04$ , 348 informative pedigrees, 804 individuals) (Figure 4a). However, this statistically significant association had very little predictive value in the sibship. Figure 4b displays the receiver operating characteristic (ROC) curves corresponding to the  $MSGB_2$  sib-proband contrast and confirms that contrasting MSGB in MS sibships is statistically significant but not suitable for prediction: The AUROC (area under ROC) areas is 0.58 (0.55, 0.62), whereas the contrasting MSGB in sibling results only in AUROC areas between 0.5 and 0.6. (AUROC 0.59 (0.55, 0.53) for  $MSGB_{56SNPs}$  including HLA and gender, and 0.58 (0.54, 0.62) for non-MHC SNPs only). In siblings, the full MSGB risk score does not appear to do better than *HLA-DRB1\*15:01*-tagging SNP and gender alone ( $P=0.97$  Hanley test (not shown)). In the general population, on the other hand, the full  $MSGB_{56SNPs}$  does better than HLA and gender only (AUROC = 0.75 (0.73, 0.77) vs 0.72 (0.69, 0.73)  $P=1e-4$ ). However, the scores fail to reach the high specificity values that would be required for the prediction of a relatively rare condition such as MS in the general population.

Interestingly, using this updated version of genetic risk score, we confirmed the lack of differences in MSGB between progressive (PP (primary progressive) + PR (progressive relapsing),  $n=182$ ) and relapsing-type (CIS (clinically isolated syndrome) + RR (relapsing remitting) + SP (secondary progressive),  $n=1914$ ) MS (Figure 5) ( $P=0.368$ ). However, classical statistical methods are primarily developed to establish the statistical significance of a difference, the non-significant  $P$ -value of  $P=0.368$  do not provide correct statistical evidence to claim for similarity of MSGB scores. We therefore used equivalence testing and found that primary progressive MS and relapsing-type MS have similar scores. This difference being limited to that corresponding to  $\pm 2$  typical MS-associated SNPs (equivalence test  $P=5.1e-03$ ). It means that if

different, the MSGB scores in PP MS and RR MS differs of less than the contribution of two risk alleles. Taken together, these results confirm that primary progressive and relapsing-type MS share the same underlying genetic architecture.

## DISCUSSION

MS is a prototypic multifactorial disease in which susceptibility is determined by polygenic inheritance. Over the last 5 years, large and multi-center DNA collections have been successfully established and the application of the hypothesis-free GWAS approach has resulted in remarkable progress in identifying the non-*HLA* genetic components of MS genetic risk. The most recent multi-center collaborative GWAS involving 9772 cases of European descent collected by 23 research groups working in 15 different countries replicated nearly all of the previously GWAS-suggested associations together with the identification of 29 novel susceptibility loci.<sup>4</sup> Understanding how these associated alleles exert their effects on risk constitutes a priority in MS genetics. We used the most updated genetic data to build a genetic profile associated with the cumulative genetic risk measured by the probability of an individual having MS.

We have confirmed previously obtained results (17 MS-associated SNPs) with the updated version of the MSGB scores using up to 64 MS markers. Our observations are as follows: (1) individuals from multi-case families have a greater MS genetic load than members of single-case families; (2) association of MSGB with disease co-occurrence in sibships is highly significant and yet, of limited power in terms of disease-prediction; (3) we confirmed the similarity of genetic load in primary progressive MS and relapsing-type MS. While the significance of any comparison between MS family members and healthy controls increases with the addition of the recently identified SNPs in the MSGB score, the current results further highlight the limited discriminative power of common alleles when patients and their relatives are being compared. Whereas a greater MSGB in siblings of MS patients was associated with an increased risk of MS, the ROC curves of MSGB differences between probands and sibs show that case-control status prediction using 56 rather than 17 markers did not really increase the predictive value of genetic risk score. The predictive power of genetic burden scores in the population is challenged by the false positive signal generated by common alleles, which by definition, are neither specific of MS patients nor they are for MS families. Interestingly, as suggested by the striking similarity of the distribution of MSGB score when comparing the relapse type of MS (CIS, RR, SP) to the progressive forms of MS (PR, PP), the normality test significance in patients with  $MSGB_{64SNPs}$  do not suggest that genetic risk score would support the existence of etiologically heterogeneity in MS. Similarly, the analysis of the residuals in the negative correlation reveals no evidence of new subgroups patients that could account specifically for the newest genetic risk discovered. In addition, no multi-modal distribution of the score supports a high predictive power of genetic risk score for a small proportion of strongly genetically predisposed patients.

Modeling of the potential multiple non-independent contributions of variants within several regions will certainly require more elaborated genetic risk score computations than the log-additive model presented in this study. This particularly applies to current knowledge in the MHC region, underlying the need of full HLA characterization. Our current findings are based on a relatively simplistic modeling of the MHC contribution to MS risk. Much information may be lost by using just two SNPs partially tagging for *HLA-DRB1\*15:01* and *HLA-B\*44*. It underlines that the log-additive model inherited from GWAS has limited power to describe complex genotype risk hierarchies, including dominant and recessive models of association. Finally, allelic heterogeneity is an established feature of polygenic inheritance<sup>26</sup> and it seems likely that many of the MS-associated loci will ultimately prove to contain additional risk allele associations.

The information of the 56-marker model is barely any greater than that explained by the 17 SNP model. Unremarkably, most of the classification accuracy comes from the variables with larger effects included in the model (gender and *HLA-DRB1\*15:01*). These variables with large effects account for most of the information, and the number of associated SNPs that would need to be added to significantly increase the information captured and thereby improve the classification accuracy is substantial. Unfortunately, while adding variants to the model increases the mean difference in the MSGB score between cases and controls, it also increases the variance in this score, that is, the extent to which the distribution of the scores may overlap. Given the absence of significant disease linkage signals outside of the MHC, it is unlikely that large-effect rare variants ( $OR > 2$ ) remains undetected after high-density GWAS.

With several large-scale replication and genome-wide scans currently going on, one must expect the list of MS-associated genomic regions to continue its exponential growth. Our results on the correction and enhancement of the MSGB<sub>17SNPs</sub> (2010–2011) to account for novel discoveries underline the risk for excess fitting of the data. In this study, we used historic familial data sets that were used in the first generation of GWAS, which could suggest that previously reported MSGB associations were slightly inflated. The accumulation of new variants will slowly increase the power of genetic risk score for the genetic study beyond susceptibility and will most likely require large multi-center efforts.

However, the need for summarizing the wealth of information represented by hundreds of disease-associated regions reinforce the need to further develop genetic risk score models with parameters emerging from the most replicated results of the literature. These will be of great importance when studying quantitative phenotypic traits such as age of onset, progression or response to treatment.

## MATERIALS AND METHODS

### Subjects

Subject recruitment occurred at four centers, two in the United States (University of California San Francisco and University of South Carolina (USC)), one in France (French Network for MS Genetics) and one in the United Kingdom (University of Cambridge) (Table 2). The study was approved by the corresponding Institutional Review Board or ethics committee, and written informed consent was obtained from all study participants.

DNA samples from 3251 MS patients and some of their relatives were available in each data sets. A group of 708 controls DNA (including 253 familial controls and spouses of probands) from UCSF were also included using stringent inclusion and exclusion criteria as previously described.<sup>23</sup> We first used the full UCSF cohort familial data set, consisting of 422 multi-case families in which at least one first-degree relative of the affected proband also had clinically definite MS, and 807 'single-case' families in which the affected individual reported no known history of MS in any family member. The study also included 551 samples from a French familial

**Table 2.** Demographic features of the enrolled individuals

	Multi-case families <sup>a</sup>			Single-case families			UCSF case/control	UCSF additional PP MS <sup>b</sup>
	UCSF	French	Cambridge	UCSF	French	Cambridge		
<b>Probands</b>								
Number	422	134	253	807	417	789	380	49
M/F/unknown (n)	103/319/0	47/87/0	60/193/0	192/615/0	134/283/0	202/587/0	123/257/0	24/25/0
Age of onset (yo)	30.15 ± 8.72	23.62 ± 9.09	26.67 ± 7.80	30.79 ± 9.07	26.13 ± 7.09	27.21 ± 6.87	35.01 ± 9.73	37.18 ± 10.99
Disease duration (years)	11.98 ± 8.98	14.51 ± 9.41	10.81 ± 8.11	10.23 ± 8.35	9.02 ± 7.15	11.16 ± 7.22	9.32 ± 9.31	NA
Disease course (CIS + RR + SP/PP + PR/unknown) (n)	387/28/7	68/2/64	231/18/4	738/56/13	365/31/21	725/60/4	356/16/8	0/49/0
MSSS (mean ± s.d.)	4.84 ± 2.51	3.29 ± 3.03	5.82 ± 2.74	4.59 ± 2.69	4.87 ± 2.78	5.98 ± 2.54	4.13 ± 2.18	6.80 ± 1.38
Sample call rates (mean ± s.d.) (%)	97.26 ± 5.48	99.30 ± 1.33	99.30 ± 1.75	98.42 ± 2.79	99.32 ± 1.20	99.29 ± 1.65	99.23 ± 1.56	92.38 ± 4.99
MSGB (n = 56) (mean ± s.d.)	8.54 ± 0.88	8.37 ± 0.82	8.01 ± 0.95	8.4 ± 0.88	8.39 ± 0.84	7.85 ± 0.82	8.24 ± 0.84	8.25 ± 0.77
<b>Relatives (n)</b>								
Mother (affected/ unaffected/unknown)	22/272/1	11/94/0						
Father (affected/ unaffected/unknown)	11/239/0	3/69/0						
Siblings (affected/ unaffected/unknown)	258/570/16	126/112/0	0/0/0	0/368/0	0/0/0	0/0/0		
<b>Unrelated controls</b>								
Number	253						455	
M/F/unknown (n)	189/64/0						146/309/0	
Sample call rates (mean ± s.d.) (%)	97.53 ± 4.01						99.34 ± 1.43	
MSGB (n = 56) (mean ± s.d.)	7.76 ± 0.66						7.66 ± 0.72	

Abbreviations: CIS, clinically isolated syndrome; F, female; M, Male; MS, multiple sclerosis; MSGB, multiple sclerosis genetic burden; MSSS, multiple sclerosis severity score; NA, not applicable; PP, primary progressive; PR, progressive relapsing; RR, relapsing-remitting; SP, secondary progressive. <sup>a</sup>Definition of multi-case family is as follows: UCSF and French group defined it as families with at least one affected first-degree relative of the affected proband, while Cambridge defined it as families with more than one affected individuals with no degree-related limitation. <sup>b</sup>In the analysis of the comparison of the MSGB scores according to disease course shown in Figure 5, UCSF PPMS patients were included in addition to the patients in the multi-case and single-case families. In total, 1914 patients with relapsing-type MS (CIS + RR + SP) and 182 primary progressive MS patients (PP + PR) from UCSF and French group are enrolled in the analysis.

data set (134 cases with reported first-degree relative with MS and 417 without), and 1042 samples from University of Cambridge (253 with reported family history of MS and 789 without). Families or cases with ambiguous records of co-occurrence were omitted from the study. Diagnostic criteria and ascertainment protocols of patients were similar for all data sets and are summarized elsewhere.<sup>27,28</sup> In addition to sib pairs from the familial data sets, a collection of 117 twin pairs from USC and UCSF were also studied (24 concordant monozygotic twins, 47 discordant monozygotic twins, 10 concordant dizygotic twins, 36 discordant dizygotic twins; Supplementary Table 4).

### SNP genotyping

Sixty four MS-SNPs were genotyped using either individual TaqMan assays or the TaqMan OpenArray genotyping technology (Life Technologies, Carlsbad, CA, USA). MS-SNPs were selected from the latest GWAS<sup>4</sup> and meta-analysis.<sup>25</sup> Samples were loaded into customized TaqMan OpenArray genotyping plates with the OpenArray Autoloader and amplified in a Dual Flat Block GeneAmp PCR System 9700, as recommended by the manufacturer. The OpenArray Genotyping Analysis Software and the Taqman Genotyper Software were used for assigning genotypes.

### Quality control

SNPs had to meet several criteria to be included in the analysis: <10% missing genotypes, Hardy-Weinberg proportion test  $P > 0.001$  in controls, and no significant SNP call rate differences between patients and controls ( $P > 0.001$ ) (Table 2).

### MSGB statistics

The MSGB score was computed based on a weighted scoring algorithm using independent 64 MS-SNPs, typically one in each genomic region of interest. This method extends the log-additive models used in previous analyses<sup>23</sup> with weights given to each SNP based on its effect size as reported by odds ratios in the largest GWAS and meta-analysis<sup>3,4,25</sup> (Supplementary Table 1 for markers and equation). Briefly, the computation of the MSGB cumulative genetic risk scores follows a log-additive model. It corresponds to the trend test typically used in the literature to identify SNP associated with MS susceptibility. In the weighted genetic risk score, the log of odds ratio associated to the presence of a single dose of risk allele is added to the score of each risk allele carried by the subject. For example, on the presence of one allele of HLA-DRB1\*15:01 would add 1.089 ( $\log(\text{OR} = 2.97)$ ) to the MSGB score; or, for a typical MS-associated SNP such as the IL7R SNP rs6897932 in IL7R, one risk allele would add 0.104 ( $\log(1.11)$ ). Where possible, missing values are substituted by proxy SNPs that tag the primarily identified risk SNPs (Supplementary Table 2,  $R^2 > 0.8$  in samples of European ancestry from the 1000 genome project<sup>29</sup>) or if not available, compensated with average frequency of risk allele dose in controls (Supplementary Table 1). When SNPs were not included in one of the centers, the risk SNP was excluded; therefore MSGB score do not account for this risk component in the analysis and the exact number of SNPs used for computation is indicated in subscript (for example,  $\text{MSGB}_{64\text{SNPs}}$ ,  $\text{MSGB}_{57\text{SNPs}}$  and  $\text{MSGB}_{56\text{SNPs}}$ ; details are in Supplementary Table 1). The minimum number of SNPs genotyped in common is 56. Mean and s.d. of the score are given in Table 2. To facilitate comparison with previous model, gender was optionally included in the score and assigned an odds ratio (OR) of 1.6 as a lower bound of sex ratio observed in epidemiological longitudinal studies.<sup>23</sup> To identify the specific effects on familial aggregation of the MHC component of MS risk, we also computed a non-MHC version of the score without contribution from HLA-DRB1\*15:01 and HLA-B\*44 (see primary and associated proxy SNPs in Supplementary Tables 1 and 2). Statistical test used are indicated in the legends of the tables and the figures. ROC curves were computed to assess the predictive power of MSGB scores in identifying affection status in patients and controls, which represent the various false positive and false negative rate of quantitative metric for all possible cutoff thresholds. s.e. were computed using the Hanley method. When power computation are presented, expected difference between the group of patients of the comparison is set to half of the gap observed between all the cases and controls.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This study was supported by the National Institutes of Health (grant RO1NS26799 RO1NS19142), the National Multiple Sclerosis Society (grant RG2901) and the Cambridge NIHR Biomedical Research Centre, the Institut National de la Santé et de la Recherche Médicale (INSERM), the Fondation d'Aide pour la Recherche sur la Sclérose En Plaques (ARSEP), the Association Française contre les Myopathies (AFM) and GIS-IBISA. The research leading to these results has received funding from the program 'Investissements d'avenir' ANR-10-IAIHU-06. We acknowledge use of the cohort of the REFGENSEP and thank ICM, CIC Pitié-Salpêtrière, Généthon and REFGENSEP's members for their help and support. NI is a postdoctoral fellow supported by Japan Society for Promotion of Science. VD received a travel grant from ARSEP. P-AG is a recipient of the Nancy Davis Foundation young investigator award. We acknowledge the assistance of Stacy Caillier, Adam Santaniello, Vinod Bakthavachalam Jinyang Wang and Matthew Chan in sample and data management.

### REFERENCES

- Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat Rev Genet* 2008; **9**: 516–526.
- Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA et al. Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 2010; **86**: 621–625.
- Gourraud PA, Harbo HF, Hauser SL, Baranzini SE. The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev* 2012; **248**: 87–103.
- Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; **476**: 214–219.
- Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007; **39**: 1329–1337.
- Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007; **357**: 851–862.
- Comabella M, Craig DW, Camina-Tato M, Morcillo C, Lopez C, Navarro A et al. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. *PLoS One* 2008; **3**: e3490.
- ANZGENE-consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009; **41**: 824–828.
- Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009; **18**: 767–778.
- De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 2009; **41**: 776–782.
- Jakkula E, Leppä V, Sulonen AM, Varilo T, Kallio S, Kempainen A et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am J Hum Genet* 2010; **86**: 285–291.
- Nischwitz S, Cepok S, Kroner A, Wolf C, Knop M, Müller-Sarnowski F et al. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. *J Neuroimmunol* 2010; **227**: 162–166.
- Sanna S, Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet* 2010; **42**: 495–497.
- Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009; **18**: 2078–2090.
- Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C et al. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011; **7**: e1002254.
- Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci USA* 2009; **106**: 18680–18685.
- Couturier N, Bucciarelli F, Nurtudinov RN, Debouverie M, Lebrun-Frenay C, Defer G et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* 2011; **134**(Pt 3): 693–703.
- De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L et al. The role of the CD58 locus in multiple sclerosis. *Proc Natl Acad Sci USA* 2009; **106**: 5264–5269.
- Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, Butter F et al. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 2012; **488**: 508–511.

- 20 Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A *et al*. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 2007; **39**: 1083–1091.
- 21 Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, Severson C *et al*. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 2009; **5**: e1000322.
- 22 De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC *et al*. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* 2009; **8**: 1111–1119.
- 23 Gourraud PA, McElroy JP, Caillier SJ, Johnson BA, Santaniello A, Hauser SL *et al*. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. *Ann Neurol* 2011; **69**: 65–74.
- 24 Sawcer S, Ban M, Wason J, Dudbridge F. What role for genetics in the prediction of multiple sclerosis? *Ann Neurol* 2010; **67**: 3–10.
- 25 Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J *et al*. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011; **70**: 897–912.
- 26 Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F *et al*. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832–838.
- 27 Barcellos LF, Oksenberg JR, Green AJ, Bucher P, Rimmler JB, Schmidt S *et al*. Genetic basis for clinical expression in multiple sclerosis. *Brain* 2002; **125**(Pt 1): 150–158.
- 28 Goodkin DE, Doolittle TH, Hauser SS, Ransohoff RM, Roses AD, Rudick RA. Diagnostic criteria for multiple sclerosis research involving multiply affected families. *Arch Neurol* 1991; **48**: 805–807.
- 29 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938–2939.

Supplementary Information accompanies this paper on Genes and Immunity website (<http://www.nature.com/gene>)



### **3. Discussion et Perspectives**

La sclérose en plaques est une maladie multifactorielle impliquant des facteurs environnementaux et des facteurs génétiques. Grâce à des projets d'envergure menés au cours de ces dernières années au sein d'un consortium international, dont certains sont présentés dans ce manuscrit, nous connaissons aujourd'hui 110 variants génétiques associés à la prédisposition à la maladie, hors HLA.

Toutes les analyses GWAS et post-GWAS menées jusqu'à présent se focalisent sur :

- la publication de méta-analyses combinant plusieurs GWAS SEP
- l'augmentation de la taille des échantillons pour répliquer les résultats obtenus, pour identifier de nouveaux variants de risques à la SEP, ou déterminer les variants causaux au sein des régions d'intérêts.

Cependant, aujourd'hui, la totalité des variants identifiés ne permettent d'expliquer que 23% de la part génétique de la SEP, conduisant à la recherche de l'héritabilité manquante.

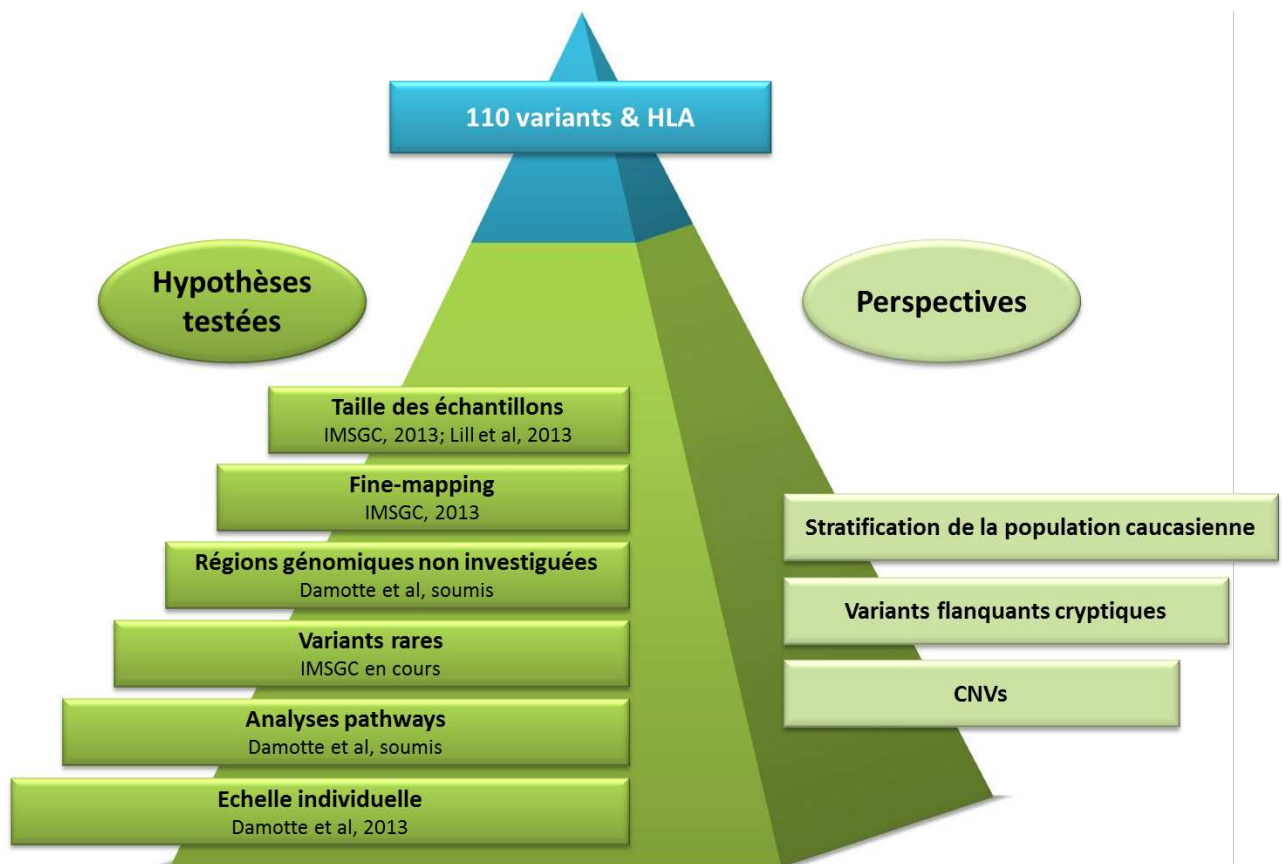
Des travaux de Visscher en 2010 (Visscher, Yang et al. 2010, Yang, Benyamin et al. 2010), indiquent que la part d'héritabilité manquante pourrait être surestimée.

Comme dit précédemment (Partie 2), le calcul classique de l'héritabilité manquante repose sur l'estimation de la corrélation génétique entre germains pour les seuls variants identifiés comme associés par les GWAS. Par une approche statistique différente, basée sur l'estimation des corrélations entre individus non apparentés pour l'ensemble des SNPs analysés (294.831), Yang et al. expliquent 45% de la variance phénotypique de la taille humaine, alors que l'évaluation classique, basée sur l'étude des seuls variants connus comme associés à la taille (environ 50) ne permet d'expliquer que 5% de cette variance.

Les différences de résultats obtenus par ces deux approches s'expliquent par :

- Les variants de prédisposition non-identifiés par GWAS auraient un très faible effet sur le trait étudié. Or les seuils de significativité utilisés pour conclure à une association étant extrêmement bas pour éliminer les faux positifs, ils ne permettent pas d'identifier des variants à très faible effet et entraînent donc de faux-négatifs.

- Les variants causaux ont probablement une MAF inférieure aux variants communs identifiés par les GWAS. Or cette différence de fréquence influence le déséquilibre de liaison entre 2 variants. Dans ce contexte, le seuil de déséquilibre de liaison classique de 0,8 avec les variants associés, augmente la probabilité de faux-négatifs.



**Figure 22. Les possibles explications de l'héritabilité manquante.**

La partie bleue de la pyramide correspond aux 27% de l'héritabilité identifiée. La partie verte correspond aux 73% de l'héritabilité manquante, restant donc à identifier.

Les hypothèses à gauche correspondent à celles qui ont été présentées dans ce manuscrit et que j'ai étudiées pendant mon doctorat.

Les perspectives à gauche correspondent aux trois hypothèses développées dans la discussion pour expliquer la part de l'héritabilité manquante.

Très récemment ont été publiés par l'équipe de Visscher les résultats obtenus pour la SEP (Lee, Harold et al. 2013). En ne considérant que les SNPs identifiés par le GWAS de l'IMSGC (IMSGC, WTCCC2 et al. 2011), environ 6% de la variance phénotypique est expliquée, contre 30% en considérant l'ensemble des SNPs analysés.

Au cours de mon doctorat, je me suis attaché à vérifier plusieurs hypothèses sur la nature de cette héritabilité manquante (**Figure 22**, partie gauche):

- une augmentation de taille d'échantillon a permis de mettre en évidence le rôle du gène *ANKRD55* et de valider celui des 5 gènes *MANBA*, *CXCR5*, *SOX8*, *RPS6KB1* and *ZBTB46* dans la prédisposition à la SEP

- l'utilisation de la puce ImmunoChip a affiné l'étude des régions chromosomiques mises en évidence par le GWAS de l'IMSGC (IMSGC, WTCCC2 et al. 2011) et d'identifier 48 nouveaux variants de prédisposition

- l'étude des régions génomiques où sont localisés les gènes *CLDN3*, *CLDN5* et *OCLN* et qui n'avaient jusqu'alors jamais été investiguées n'a pas permis de valider le rôle de ces gènes dans la SEP.

- l'étude des variants rares est en cours par le biais du projet ExomeChip de l'IMSGC

- l'étude du réseau des molécules d'adhésion a mis en évidence un sous-réseau composé des gènes *ITGAL*, *ICAM1* et *ICAM3* associé à la prédisposition à la SEP

- l'étude de charge génétique MSGB a montré une différence de charge génétique d'une part entre les cas index des familles multi-cas et leur frère ou sœur non-atteint et d'autre part entre une population de patients et de contrôles. Dans les deux cas, les différences ne sont que très peu prédictives, laissant supposer qu'il reste encore une part génétique à identifier.

Il pourrait en fait exister d'autres explications potentielles pour l'héritabilité manquante. Dans le paragraphe qui suit, je détaillerai 3 hypothèses qu'il serait intéressant d'explorer dans le futur (**Figure 22**, partie droite).

### **3.1. Autres pistes pour identifier l'héritabilité manquante**

#### *3.1.1. Hypothèse 1 : Stratification de la population Caucasienne*

Bien que les études antérieures aient été effectuées sur des populations homogènes (de même ethnie) et que les analyses intègrent des corrections pour prendre en compte la





stratification des populations, **il se peut que cette approche par ethnie ait masqué le rôle de certains variants de risque.**

Un variant non identifié à risque dans la population caucasienne en générale pourrait conférer une prédisposition à la maladie dans une ou plusieurs sous-populations en particulier. Inversement, un variant mis en évidence comme étant à risque dans la population caucasienne pourrait ne pas jouer de rôle dans une ou plusieurs sous-populations, pouvant même jouer un rôle protecteur dans celles-ci.

Dans les études antérieures, les résultats d'associations des différentes populations étant combinés, **les signaux d'association conférés par les variants de risque population-spécifique sont effacés par les signaux négatifs obtenus dans les autres sous-populations.**

Des études pangénomiques sur des populations définies (française, allemande, espagnole, norvégienne...) pourraient aboutir à l'identification de nouveaux variants de risque population-spécifique. Cette hypothèse prend aujourd'hui tout son sens avec les premières études d'associations sur les variants rares, qui sont plus à même d'être à risque pour une population spécifique.

**La prédisposition génétique à la SEP pourrait donc reposer sur un ensemble de variants communs à tous les individus caucasiens, en interaction avec des variants de risques spécifiques à des sous-populations.** Cette hypothèse pourrait expliquer les différences de prévalence que l'on observe entre deux pays caucasiens se trouvant à une même latitude.

Cependant, ce type d'approche population-spécifique n'est pour le moment que peu réalisable. Cette stratégie nécessiterait d'importants moyens financiers et des cohortes très importantes pour une sous-population donnée. Sachant que la France compte aujourd'hui environ 80 000 patients atteints de SEP, un GWAS français portant sur 10 000 patients pour la phase de découverte et 10 000 patients pour la phase de répliation conduirait à étudier 25% des patients français.

### 3.1.2. Hypothèse 2 : Variants flanquants cryptiques

Comme je l'ai mentionné précédemment, les études GWAS portent sur des SNPs fréquents dans la population générale ( $MAF > 0,05$ ) et l'analyse de ces projets consiste à comparer les fréquences alléliques de chacun des variants étudiés entre une population de patients et une population contrôle. Cependant, même si ces SNPs sont fréquents, les allèles à risques ne sont pas présents chez tous les patients. Par exemple, sur les 110 variants connus aujourd'hui comme étant associés à la SEP, 22 d'entre eux ont une fréquence d'allèle à risque



inférieure à 25% chez les contrôles. Cela veut dire qu'au plus 44% ( $p^2 + 2pq$  avec  $p=0,25$  et  $q=0,75$ ) des contrôles possèdent l'allèle à risque. Du fait du faible risque associé à ces variants, on peut légitimement affirmer que pour ces 22 variants, **près de la moitié des patients ne possèdent pas l'allèle à risque d'un polymorphisme donné.**

Dans ce contexte, deux hypothèses peuvent être avancées :

- les patients non-porteurs d'allèles de risques de certaines régions portent les allèles de risque des autres régions chromosomiques identifiées

- les patients non-porteurs d'allèles de risques de certaines régions portent des allèles de risque d'un autre variant localisé dans la même région que le principal variant identifié, et n'étant pas en déséquilibre de liaison avec celui-ci.

Dans le cas de la 2<sup>ème</sup> hypothèse, les méthodes d'analyses actuelles des GWAS ne permettent pas d'identifier ce variant à risque colocalisé avec le principal variant identifié.

Comme je l'ai mentionné, les analyses actuelles d'association cherchent à identifier une différence de fréquence allélique entre une population de patients et une autre de contrôles. En considérant ces populations, il y a effectivement une différence de fréquences alléliques pour les variants identifiés. Cependant, **en considérant ces populations dans leur ensemble**, il n'est pas possible de mettre en évidence de différence significative de fréquences alléliques pour les variants colocalisés (ou flanquants), **même si certains d'entre eux sont à risque pour la maladie.** Une solution à ce problème serait, dans une deuxième étape de l'analyse d'association, de réaliser des tests d'associations sur les autres variants appartenant à la région d'intérêt et **en ne prenant en compte que les individus qui ne portent pas l'allèle à risque du variant identifié.** Cette stratégie pourrait permettre l'identification d'un ou d'autres variants associés à la SEP. Cela sous-entend qu'il existerait une différence de fréquence allélique d'un autre ou d'autres variant(s) de la région d'intérêt **entre les patients et les contrôles ne portant pas l'allèle à risque du premier variant identifié.**

Bien entendu, le fait de ne prendre en compte que les individus ne portant pas l'allèle à risque du SNP identifié pourrait entraîner une perte de puissance statistique car les tailles des populations de patients et de contrôles diminueraient. Cependant, si les analyses suivantes ne sont faites que sur les SNPs appartenants aux régions identifiées (par exemple + ou - 100kb autour du premier SNP identifié), le nombre de tests à effectuer diminuerait, n'affectant que très peu la puissance statistique de l'étude.



Ainsi, en n'utilisant que des données déjà générées, peut-être serait-il possible d'identifier de nouveaux variants de prédisposition à la SEP, dans des régions déjà identifiées comme étant associées à la maladie.

### 3.1.3. Hypothèse 3 : Le rôle des Copy-Number Variations

Une autre source de l'héritabilité manquante dans la SEP et dans d'autres maladies complexes pourrait se trouver dans les **Copy-Number Variations** (CNVs) (Maher 2008, Manolio, Collins et al. 2009). Ces CNVs consistent en des duplications ou des délétions de séquences d'ADN pouvant aller de dix à des centaines de paires de bases. Des variations rares dans ces séquences ont été montrées comme étant associées à certains troubles comme la schizophrénie (International Schizophrenia 2008, Stefansson, Rujescu et al. 2008) ou l'autisme (Weiss, Shen et al. 2008, Miller, Shen et al. 2009, Pinto, Pagnamenta et al. 2010, Sato, Lionel et al. 2012). Ces variations apparaissent généralement de novo, c'est à dire que cette variation est présente chez l'individu atteint mais pas chez ses parents.

Une large étude du Wellcome Trust publiée en 2010 avait pour but d'identifier des CNVs associées à 8 maladies : trouble bipolaire, cancer du sein, maladie coronarienne, maladie de Crohn, hypertension, diabètes de type 1 et de type 2 (Wellcome Trust Case Control, Craddock et al. 2010). La SEP n'était pas incluse dans l'étude. L'analyse a porté sur 3432 CNVs fréquents et a mis en évidence 3 loci où des CNVs sont associés avec certaines de ces 8 maladies, dont le locus *HLA*. Cependant, les CNVs identifiés dans cette étude se trouvaient dans des régions déjà mises en évidence par les études GWAS de chaque maladie respective.

Dans la SEP, aucune étude à grande échelle n'a été menée à ce jour pour identifier des CNVs associés à la prédisposition génétique. Il pourrait donc être intéressant de se focaliser par la suite sur ces polymorphismes.

## **3.2. Limites des études génétiques**

Comme je l'ai précisé au cours de ce manuscrit, certaines limites aux études génétiques actuelles doivent être apportés.

Tout d'abord le seuil de significativité *genome-wide* de  $5 \times 10^{-8}$  permettant d'exclure les résultats « faux positifs », **ne permet pas de conclure sur les résultats négatifs**. Des variants non mis en évidence par les GWAS, du fait de leur résultat d'association qui n'a pas atteint la p-value seuil, sont peut-être bien à risque pour la SEP : ils pourraient être des résultats « faux négatifs ».



La majorité des GWAS se sont focalisés sur la population Caucasienne. Il serait intéressant maintenant de faire ces études sur d'autres ethnies. Les résultats pourraient être soit sensiblement équivalents à ceux de la population caucasienne, soit différents. Dans ce cas, les polymorphismes ethnique-spécifiques identifiés pourraient apporter de nouvelles informations pour la compréhension des mécanismes biologiques mis en jeu dans la SEP. Une récente étude de Isobe et al. (Isobe, Gourraud et al. 2013) a montré qu'en plus d'HLA, seuls 8 variants de prédisposition non-HLA étaient communs à la population Caucasienne et à la population Afro-Américaine.

Enfin, les importants déséquilibres de liaison existant au sein des régions identifiées ne permettent pas de déterminer avec certitude le gène de chaque région impliqué dans la maladie, ce qui retarde et complique les études fonctionnelles.

Cependant les deux points précédents sont à mettre en relation avec le déséquilibre de liaison plus fort observé dans la population caucasienne que dans la population africaine par exemple. La population caucasienne s'étant développée à partir d'un petit groupe d'individus, la diversité génétique est moindre chez les caucasiens. L'existence d'une plus grande diversité génétique chez les Africains représente donc un intérêt majeur pour affiner l'identification des variants causaux au sein des régions déjà mises en évidence.

### **3.3. Rôle fonctionnel des variants de risques identifiés**

La compréhension du rôle fonctionnel associé aux variants à risques identifiés constitue le défi majeur des prochaines années. Deux stratégies peuvent être développées : étudier les polymorphismes individuellement, ou étudier des réseaux de gènes associés à la maladie et étant impliqués dans une même voie biologique.

D'autre part, dans la plupart de régions génomiques identifiées à ce jour, il n'y a aucune certitude sur l'identité du ou des variants causaux. Il paraît donc plus pertinent aujourd'hui d'initier des études fonctionnelles chez l'Homme, à partir de cohortes de patients porteurs ou non du variant de risque, que d'envisager l'utilisation de modèles animaux génétiquement modifiés.

#### *3.2.1. L'étude individuelle des polymorphismes*

Cette stratégie consiste à corrélérer, pour chaque polymorphisme, les génotypes des patients à un possible effet fonctionnel. Si un effet est identifié, celui-ci peut être dû directement au variant identifié, ce qui sous-entend que les analyses génétiques ont identifié le





variant causal. Mais l'effet fonctionnel identifié peut aussi ne pas être dû à un effet direct du variant identifié, en fort déséquilibre avec le variant causal.

A ce jour, un certain nombre d'études cherchant à corréler les variants de risque SEP à un possible effet fonctionnel ont été publiées.

Il a été montré un effet du variant de prédisposition rs6897932 sur l'expression du gène IL7R. Ce variant est situé dans l'exon 6 du gène IL7R, qui peut subir un épissage alternatif :

- les transcrits incluant l'exon 6 codent pour une forme membranaire de la sous-unité  $\alpha$  du gène IL7R

- les transcrits sans exon 6 codent pour une forme soluble de la protéine.

Les auteurs ont montré que la présence de l'allèle de risque C doublait le taux d'épissage de l'exon 6 comparé aux transcrits ayant l'allèle T. Ainsi les individus portant l'allèle à risque C produiraient moins de forme membranaire du récepteur à l'IL7, et donc plus de forme soluble (Gregory, Schmidt et al. 2007).

Une étude récente publiée par Couturier et al. (Couturier, Bucciarelli et al. 2011) a investigué le rôle du variant à risque rs34536443 du gène *TYK2* (*Tyrosine kinase 2*) dans la SEP. Après avoir confirmé son association dans la population française, Couturier et al. ont montré que l'allèle protecteur C entraîne une diminution de l'activité de la Tyrosine kinase 2. Cette diminution d'activité est corrélée avec un changement du profil sécrétoire cytokinique, en faveur des cytokines anti-inflammatoires de type  $T_H2$ . Cela suggère que le variant de prédisposition rs34536443 pourrait favoriser la différenciation des lymphocytes T vers un phénotype  $T_H2$ , modifiant la balance entre signaux pro- et anti-inflammatoires. Les auteurs concluent que la modulation de l'activité de la Tyrosine kinase 2 pourrait représenter une nouvelle stratégie thérapeutique pour la sclérose en plaques.

### 3.2.2. L'étude de réseaux de gènes

Comme l'ont montré mes travaux sur le MSGB, intégrer dans un seul score toutes les données génétiques que nous possédons à l'heure actuelle ne permet pas de faire de la prédiction ou du diagnostic.

En revanche, ce score MSGB pourrait permettre de stratifier les patients en différents sous-groupes, ce qui pourrait être d'intérêt pour le développement de thérapies ciblées.



La plupart des variants identifiés sont localisés dans des régions contenant des gènes ayant des fonctions connues, comme des gènes immunitaires impliqués dans les voies de différenciation des lymphocytes T APC/T<sub>H0</sub>, T<sub>H1</sub>, T<sub>H2</sub>, T<sub>H17</sub>, Treg, T<sub>FH</sub>/LB (**Figure 8**, page 54). D'autres gènes localisés dans les régions génomiques de prédisposition sont impliqués dans le processus de remyélinisation spontanée. Les gènes peuvent donc être répartis par réseaux, selon leur fonction biologique.

Partant de ce constat, nous pouvons émettre l'hypothèse que certains patients sont porteurs de plus d'allèles de risque de gènes appartenant à un réseau particulier comparé aux autres groupes et qu'ainsi il soit possible de définir des sous-groupes de patients présentant une charge génétique prépondérante dans un réseau génétique spécifique.

Chaque sous-groupe de patients génétiquement identifié pourrait ainsi présenter des caractéristiques cliniques différentes en ce qui concerne l'âge de début de la maladie, la sévérité ou la forme de la maladie. Il serait aussi intéressant d'évaluer les capacités de remyélinisation de chaque sous-groupe de patients par un suivi longitudinal en IRM.

### **3.3. Corrélations Génotype-Phénotype**

La majorité des projets présentés dans ce mémoire avaient pour but d'identifier les facteurs génétiques de prédisposition à la SEP. D'autres stratégies d'études génétiques sont menées depuis quelques années pour identifier la part génétique de la sévérité de la maladie et de la réponse aux traitements.

#### 3.3.1. Génétique de la sévérité de la Sclérose En Plaques

Plusieurs études ont été menées pour tenter d'identifier des marqueurs génétiques associés à la sévérité de la maladie. Elles se sont soit focalisées sur un ou des gène(s) candidat(s) (Jensen, Stankovich et al. 2010, Gemmati, Zeri et al. 2012, Pauwels, Cosemans et al. 2013), soit ont été menées sur génome entier (Baranzini, Wang et al. 2009, Brynedal, Wojcik et al. 2010, IMSGC 2011).

La plus importante d'entre elles a été menée par l'IMSGC. Au cours de cette étude, 2,5 millions de SNPs ont été analysés chez 1470 patients américains et anglais mais cela n'a pas permis d'identifier de facteurs de sévérité de manière certaine (IMSGC 2011).

Aujourd'hui, seuls deux marqueurs du gène *MGAT5* semblent être potentiellement associés à la sévérité de la maladie. Ils ont été mis en évidence par un GWAS sur la sévérité mené dans une population de 1040 patients provenant de France, d'Italie et de Suède (Brynedal, Wojcik et al. 2010). Cependant les résultats d'association n'ont pas atteint le seuil



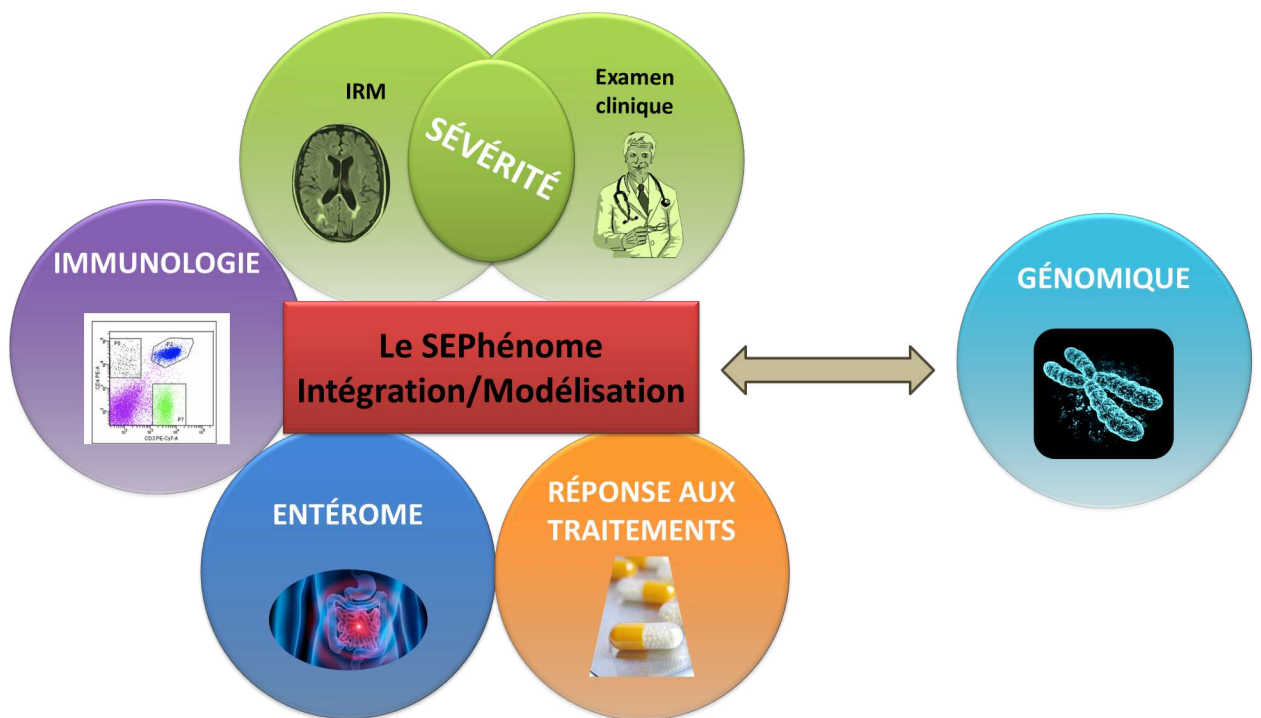
de significativité *genome-wide* de  $5 \times 10^{-8}$ . Des études de réplifications seront donc nécessaires pour confirmer l'association de ces deux marqueurs avec la sévérité de la SEP.

### 3.3.2. Réponses aux traitements

Depuis quelques années, la recherche de marqueurs génétiques prédictifs de la réponse aux traitements est en plein essor. De la même manière que pour les études d'association avec la prédisposition à la maladie, des études pharmacogénétiques de gènes candidats et *genome-wide* sont menées.

Deux GWAS ont été publiés sur la réponse à l'IFN- $\beta$  (Byun, Caillier et al. 2008, Comabella, Craig et al. 2009). Le plus important porte sur une cohorte de 99 répondeurs et 107 non répondeurs provenant d'Espagne et de France. Après l'ajout de 79 individus supplémentaires (correspondant à 44 répondeurs et 35 non répondeurs) pour augmenter la puissance statistique de l'étude, un certain nombre de SNPs, dont 7 intragéniques, ont été mis en évidence comme étant associé à la réponse à l'IFN- $\beta$ . Une étude ultérieure a par la suite validé l'association du gène *GPC5* avec la réponse à l'IFN- $\beta$  (Cenit, Blanco-Kelly et al. 2009).

Cependant, la notion de répondeurs et non-répondeurs à un traitement est difficile à appréhender et standardiser malgré plusieurs publications. Les études de Byun et al. et de Comabella et al. (Byun, Caillier et al. 2008, Comabella, Craig et al. 2009) considèrent que les répondeurs n'ont aucune poussée et pas d'augmentation d'EDSS durant les 2 ans de suivi longitudinal. En revanche, la définition des non-répondeurs est légèrement différente dans ces deux études. Buyn et al. considèrent que les non-répondeurs présentent au moins 2 poussées ou une augmentation d'un point d'EDSS sur cette même période. Comabella et al., quant-à-eux, considèrent comme non-répondeurs les patients présentant au moins une poussée et une augmentation d'un point d'EDSS sur les 2 ans de suivi. Cette difficulté de définition des répondeurs et non répondeurs, combinée à des petites cohortes, limite l'impact des études pharmaco-génomiques menées jusqu'à présent.



**Figure 23. Le SEPhénome : Phénomène exhaustif de la SEP.**

Présentation des différentes données phénotypiques à intégrer et à corrélérer avec des données génomiques

### 3.3.3. Phénomène exhaustif de la SEP : le SEPhénomène

La plupart des études sur la sévérité et la réponse aux traitements, reposent sur des tests d'association entre les variants génétiques et d'une seule caractéristique clinique indépendante.

Le challenge aujourd'hui consiste à identifier des groupes de patients de « SEPhénomènes homogènes » en intégrant (**Figure 23**) :

- des données cliniques (telles que l'âge de début de la maladie, le score de handicap EDSS, la forme de la maladie, la sévérité)
- des données IRM prenant en compte le niveau d'inflammation moyen et la capacité de remyélinisation qu'il est possible aujourd'hui d'évaluer grâce aux technologies d'imagerie émergentes
- des données immunologiques comme la proportion de chaque sous-type cellulaire circulant ( $T_H1$ ,  $T_H2$ ,  $T_H17$ , Treg), la nature et la quantité des cytokines et chimiokines sécrétées par ces sous-types cellulaires
- des données « d'entérome » reflétant la composition de la flore intestinale

Les études pourraient :

- se focaliser sur les variants de prédisposition déjà identifiés et étudier leur association avec un ou des groupes de patients présentant un SEPhénomène similaire
- essayer de corréler les groupes de patients à des profils génétiques particuliers, par exemple en partant des données de méta-analyse.

Ce type d'étude présente des limites qu'il faudra anticiper:

- les groupes de patients étant déterminés sur la base de plusieurs critères indépendants, la cohorte principale sera stratifiée. Si pour  $n$  critères cliniques, on détermine 5 classes quantitatives ou qualitatives, on obtient alors  $5^n$  groupes de patients. L'intégration de 5 critères phénotypiques, présentant chacun 5 classes, entraîne donc potentiellement la création de 3125 groupes de patients. La taille des cohortes nécessaires pour ces analyses devra par conséquent être importante, pour avoir un nombre statistiquement suffisant de patients dans chaque groupe.
- la création de groupes de patients requiert également la nécessité d'avoir des données de qualités et standardisées, que ce soit dans leur acquisition mais aussi leur analyse. Par exemple, aujourd'hui, les données d'imagerie sont acquises par l'intermédiaire de différentes





IRM et analysées avec l'aide de logiciels différents suivant le lieu de l'étude. Avoir des données de qualités et standardisées introduira une certaine robustesse dans la définition des groupes.

- enfin la SEP étant une maladie qui évolue au cours du temps, il est indispensable d'opérer un suivi longitudinal des patients.

En France, la cohorte OFSEP (Observatoire Française de la Sclérose En Plaques), actuellement en cours de recrutement, pourrait fournir un grand nombre de patients SEP génotypés et phénotypés de manière standardisée. Cependant, le coût que représente la création d'une telle base de données constitue un frein important.



## **4. Conclusion**

Les avancées récentes en génétique de la SEP permettent aujourd'hui de disposer de connaissances robustes dans le domaine. L'identification de 110 variants de prédisposition à la maladie ouvre des perspectives de recherche fonctionnelle et de corrélation génotype-phénotype.

La réussite des études génétiques et fonctionnelles futures dépendra fortement du développement de nouvelles stratégies et méthodologies statistiques, de systèmes d'intégration et de modélisation mathématiques, d'outils de standardisation des mesures phénotypiques et enfin de constitution de collections biologiques de grande envergure.



## **Références bibliographiques**

Abbott, N. J., A. A. Patabendige, D. E. Dolman, S. R. Yusof and D. J. Begley (2010). "Structure and function of the blood-brain barrier." Neurobiol Dis **37**(1): 13-25.

Alloza, I., D. Otaegui, A. L. de Lapuente, A. Antiguedad, J. Varade, C. Nunez, R. Arroyo, E. Urcelay, O. Fernandez, L. Leyva, M. Fedetz, G. Izquierdo, M. Lucas, B. Oliver-Martos, A. Alcina, A. Saiz, Y. Blanco, M. Comabella, X. Montalban, J. Olascoaga, F. Matesanz and K. Vandebroek (2012). "ANKRD55 and DHCR7 are novel multiple sclerosis risk loci." Genes Immun **13**(3): 253-257.

ANZgene (2009). "Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20." Nat Genet **41**(7): 824-828.

Ascherio, A. and K. L. Munger (2007). "Environmental risk factors for multiple sclerosis. Part I: the role of infection." Ann Neurol **61**(4): 288-299.

Ascherio, A. and K. L. Munger (2007). "Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors." Ann Neurol **61**(6): 504-513.

Asimit, J. and E. Zeggini (2010). "Rare variant association analysis methods for complex traits." Annu Rev Genet **44**: 293-308.

Ballabh, P., A. Braun and M. Nedergaard (2004). "The blood-brain barrier: an overview: structure, regulation, and clinical implications." Neurobiol Dis **16**(1): 1-13.

Bansal, V., O. Libiger, A. Torkamani and N. J. Schork (2010). "Statistical analysis strategies for association studies involving rare variants." Nat Rev Genet **11**(11): 773-785.

Baranzini, S. E., N. W. Galwey, J. Wang, P. Khankhanian, R. Lindberg, D. Pelletier, W. Wu, B. M. Uitdehaag, L. Kappos, C. H. Polman, P. M. Matthews, S. L. Hauser, R. A. Gibson, J. R. Oksenberg and M. R. Barnes (2009). "Pathway and network-based analysis of genome-wide association studies in multiple sclerosis." Hum Mol Genet **18**(11): 2078-2090.

Baranzini, S. E., J. Mudge, J. C. van Velkinburgh, P. Khankhanian, I. Khrebtukova, N. A. Miller, L. Zhang, A. D. Farmer, C. J. Bell, R. W. Kim, G. D. May, J. E. Woodward, S. J. Caillier, J. P. McElroy, R. Gomez, M. J. Pando, L. E. Clendenen, E. E. Ganusova, F. D. Schilkey, T. Ramaraj, O. A. Khan, J. J. Huntley, S. Luo, P. Y. Kwok, T. D. Wu, G. P. Schroth, J. R. Oksenberg, S. L. Hauser and S. F. Kingsmore (2010). "Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis." Nature **464**(7293): 1351-1356.

Baranzini, S. E., J. Wang, R. A. Gibson, N. Galwey, Y. Naegelin, F. Barkhof, E. W. Radue, R. L. Lindberg, B. M. Uitdehaag, M. R. Johnson, A. Angelakopoulou, L. Hall, J. C. Richardson, R. K. Prinjha, A. Gass, J. J. Geurts, J. Kragt, M. Sombekke, H. Vrenken, P. Qualley, R. R. Lincoln, R. Gomez, S. J. Caillier, M. F. George, H. Mousavi, R. Guerrero, D.

T. Okuda, B. A. Cree, A. J. Green, E. Waubant, D. S. Goodin, D. Pelletier, P. M. Matthews, S. L. Hauser, L. Kappos, C. H. Polman and J. R. Oksenberg (2009). "Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis." Hum Mol Genet **18**(4): 767-778.

Barcellos, L. F., J. R. Oksenberg, A. B. Begovich, E. R. Martin, S. Schmidt, E. Vittinghoff, D. S. Goodin, D. Pelletier, R. R. Lincoln, P. Bucher, A. Swerdlin, M. A. Pericak-Vance, J. L. Haines, S. L. Hauser and G. Multiple Sclerosis Genetics (2003). "HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course." Am J Hum Genet **72**(3): 710-716.

Benedikz, J., H. Magnusson and G. Guthmundsson (1994). "Multiple sclerosis in Iceland, with observations on the alleged epidemic in the Faroe Islands." Ann Neurol **36 Suppl 2**: S175-179.

Bertrams, H. J. and E. K. Kuwert (1976). "Association of histocompatibility haplotype HLA-A3-B7 with multiple sclerosis." J Immunol **117**(5 Pt.2): 1906-1912.

Brynedal, B., J. Wojcik, F. Esposito, V. Debailleul, J. Yaouanq, F. Martinelli-Boneschi, G. Edan, G. Comi, J. Hillert and H. Abderrahim (2010). "MGAT5 alters the severity of multiple sclerosis." J Neuroimmunol **220**(1-2): 120-124.

Bustamante, C. D., E. G. Burchard and F. M. De la Vega (2011). "Genomics for the world." Nature **475**(7355): 163-165.

Byun, E., S. J. Caillier, X. Montalban, P. Villoslada, O. Fernandez, D. Brassat, M. Comabella, J. Wang, L. F. Barcellos, S. E. Baranzini and J. R. Oksenberg (2008). "Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis." Arch Neurol **65**(3): 337-344.

Cenit, M. D., F. Blanco-Kelly, V. de las Heras, M. Bartolome, E. G. de la Concha, E. Urcelay, R. Arroyo and A. Martinez (2009). "Glypican 5 is an interferon-beta response gene: a replication study." Mult Scler **15**(8): 913-917.

Comabella, M., D. W. Craig, C. Morcillo-Suarez, J. Rio, A. Navarro, M. Fernandez, R. Martin and X. Montalban (2009). "Genome-wide scan of 500,000 single-nucleotide polymorphisms among responders and nonresponders to interferon beta therapy in multiple sclerosis." Arch Neurol **66**(8): 972-978.

Compston, A. (2005). McAlpine's multiple sclerosis. Philadelphia, Churchill Livingstone Elsevier.

Compston, A. and A. Coles (2002). "Multiple sclerosis." Lancet **359**(9313): 1221-1231.

Compston, A. and A. Coles (2008). "Multiple sclerosis." Lancet **372**(9648): 1502-1517.

Compston, D. A., J. R. Batchelor and W. I. McDonald (1976). "B-lymphocyte alloantigens associated with multiple sclerosis." Lancet **2**(7998): 1261-1265.

Confavreux, C. and S. Vukusic (2006). "[The natural history of multiple sclerosis]." Rev Prat **56**(12): 1313-1320.

Cortes, A. and M. A. Brown (2011). "Promise and pitfalls of the Immunochip." Arthritis Res Ther **13**(1): 101.

Cotsapas, C., B. F. Voight, E. Rossin, K. Lage, B. M. Neale, C. Wallace, G. R. Abecasis, J. C. Barrett, T. Behrens, J. Cho, P. L. De Jager, J. T. Elder, R. R. Graham, P. Gregersen, L. Klareskog, K. A. Siminovitch, D. A. van Heel, C. Wijmenga, J. Worthington, J. A. Todd, D. A. Hafler, S. S. Rich, M. J. Daly and F. O. N. o. Consortia (2011). "Pervasive sharing of genetic effects in autoimmune disease." PLoS Genet **7**(8): e1002254.

Couturier, N., F. Bucciarelli, R. N. Nurtdinov, M. Debouverie, C. Lebrun-Frenay, G. Defer, T. Moreau, C. Confavreux, S. Vukusic, I. Cournu-Rebeix, R. H. Goertsches, U. K. Zettl, M. Comabella, X. Montalban, P. Rieckmann, F. Weber, B. Muller-Myhsok, G. Edan, B. Fontaine, L. T. Mars, A. Saoudi, J. R. Oksenberg, M. Clanet, R. S. Liblau and D. Brassat (2011). "Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility." Brain **134**(Pt 3): 693-703.

De Jager, P. L., L. B. Chibnik, J. Cui, J. Reischl, S. Lehr, K. C. Simon, C. Aubin, D. Bauer, J. F. Heubach, R. Sandbrink, M. Tyblova, P. Lelkova, E. Havrdova, C. Pohl, D. Horakova, A. Ascherio, D. A. Hafler and E. W. Karlson (2009). "Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score." Lancet Neurol **8**(12): 1111-1119.

De Jager, P. L., X. Jia, J. Wang, P. I. de Bakker, L. Ottoboni, N. T. Aggarwal, L. Piccio, S. Raychaudhuri, D. Tran, C. Aubin, R. Briskin, S. Romano, S. E. Baranzini, J. L. McCauley, M. A. Pericak-Vance, J. L. Haines, R. A. Gibson, Y. Naeglin, B. Uitdehaag, P. M. Matthews, L. Kappos, C. Polman, W. L. McArdle, D. P. Strachan, D. Evans, A. H. Cross, M. J. Daly, A. Compston, S. J. Sawcer, H. L. Weiner, S. L. Hauser, D. A. Hafler and J. R. Oksenberg (2009). "Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci." Nat Genet **41**(7): 776-782.

Dempfle, A., A. Scherag, R. Hein, L. Beckmann, J. Chang-Claude and H. Schafer (2008). "Gene-environment interactions for complex traits: definitions, methodological requirements and challenges." Eur J Hum Genet **16**(10): 1164-1172.

Engelhardt, B. (2008). "Immune cell entry into the central nervous system: involvement of adhesion molecules and chemokines." J Neurol Sci **274**(1-2): 23-26.

Engelhardt, B. and R. M. Ransohoff (2005). "The ins and outs of T-lymphocyte trafficking to the CNS: anatomical sites and molecular mechanisms." Trends Immunol **26**(9): 485-495.

Fernandes de Abreu, D. A., M. C. Babron, I. Rebeix, C. Fontenille, J. Yaouanq, D. Brassat, B. Fontaine, F. Clerget-Darpoux, F. Jehan and F. Feron (2009). "Season of birth and not vitamin



D receptor promoter polymorphisms is a risk factor for multiple sclerosis." Mult Scler **15**(10): 1146-1152.

Fox, E. J. (2004). "Immunopathology of multiple sclerosis." Neurology **63**(12 Suppl 6): S3-7.

Gale, C. R. and C. N. Martyn (1995). "Migrant studies in multiple sclerosis." Prog Neurobiol **47**(4-5): 425-448.

Gemmati, D., G. Zeri, E. Orioli, F. E. De Gaetano, F. Salvi, I. Bartolomei, S. D'Alfonso, C. Dall'osso, M. A. Leone, A. V. Singh, R. Asselta and P. Zamboni (2012). "Polymorphisms in the genes coding for iron binding and transporting proteins are associated with disability, severity, and early progression in multiple sclerosis." BMC Med Genet **13**: 70.

Genomes Project, C., G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Genomes Project, C., G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth and G. A. McVean (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**(7422): 56-65.

Gibson, G. (2011). "Rare and common variants: twenty arguments." Nat Rev Genet **13**(2): 135-145.

Gold, R. and J. S. Wolinsky (2011). "Pathophysiology of multiple sclerosis and the place of teriflunomide." Acta Neurol Scand **124**(2): 75-84.

Gourraud, P. A., H. F. Harbo, S. L. Hauser and S. E. Baranzini (2012). "The genetics of multiple sclerosis: an up-to-date review." Immunol Rev **248**(1): 87-103.

Gourraud, P. A., J. P. McElroy, S. J. Caillier, B. A. Johnson, A. Santaniello, S. L. Hauser and J. R. Oksenberg (2011). "Aggregation of multiple sclerosis genetic risk variants in multiple and single case families." Ann Neurol **69**(1): 65-74.

Gregory, S. G., S. Schmidt, P. Seth, J. R. Oksenberg, J. Hart, A. Prokop, S. J. Caillier, M. Ban, A. Goris, L. F. Barcellos, R. Lincoln, J. L. McCauley, S. J. Sawcer, D. A. Compston, B. Dubois, S. L. Hauser, M. A. Garcia-Blanco, M. A. Pericak-Vance and J. L. Haines (2007). "Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis." Nat Genet **39**(9): 1083-1091.

Handel, A. E., G. Giovannoni, G. C. Ebers and S. V. Ramagopalan (2010). "Environmental factors and their timing in adult-onset multiple sclerosis." Nat Rev Neurol **6**(3): 156-166.

Hartsock, A. and W. J. Nelson (2008). "Adherens and tight junctions: structure, function and connections to the actin cytoskeleton." Biochim Biophys Acta **1778**(3): 660-669.

Hawkins, B. T. and T. P. Davis (2005). "The blood-brain barrier/neurovascular unit in health and disease." *Pharmacol Rev* **57**(2): 173-185.

Hemminki, K., X. Li, J. Sundquist, J. Hillert and K. Sundquist (2009). "Risk for multiple sclerosis in relatives and spouses of patients diagnosed with autoimmune and related conditions." *Neurogenetics* **10**(1): 5-11.

Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." *Nat Rev Genet* **6**(2): 95-108.

Hogart, C. J., T. G. Clark, M. De Iorio, J. C. Whittaker and D. J. Balding (2008). "Genome-wide significance for dense SNP and resequencing data." *Genet Epidemiol* **32**(2): 179-185.

Holick, M. F. (2004). "Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease." *Am J Clin Nutr* **80**(6 Suppl): 1678S-1688S.

Holmans, P. (2010). "Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits." *Adv Genet* **72**: 141-179.

Hunt, K. A., V. Mistry, N. A. Bockett, T. Ahmad, M. Ban, J. N. Barker, J. C. Barrett, H. Blackburn, O. Brand, O. Burren, F. Capon, A. Compston, S. C. Gough, L. Jostins, Y. Kong, J. C. Lee, M. Lek, D. G. MacArthur, J. C. Mansfield, C. G. Mathew, C. A. Mein, M. Mirza, S. Nutland, S. Onengut-Gumuscu, E. Papouli, M. Parkes, S. S. Rich, S. Sawcer, J. Satsangi, M. J. Simmonds, R. C. Trembath, N. M. Walker, E. Wozniak, J. A. Todd, M. A. Simpson, V. Plagnol and D. A. van Heel (2013). "Negligible impact of rare autoimmune-locus coding-region variants on missing heritability." *Nature* **498**(7453): 232-235.

IMSGC (2011). "Genome-wide association study of severity in multiple sclerosis." *Genes Immun* **12**(8): 615-625.

IMSGC, D. A. Hafler, A. Compston, S. Sawcer, E. S. Lander, M. J. Daly, P. L. De Jager, P. I. de Bakker, S. B. Gabriel, D. B. Mirel, A. J. Iverson, M. A. Pericak-Vance, S. G. Gregory, J. D. Rioux, J. L. McCauley, J. L. Haines, L. F. Barcellos, B. Cree, J. R. Oksenberg and S. L. Hauser (2007). "Risk alleles for multiple sclerosis identified by a genomewide study." *N Engl J Med* **357**(9): 851-862.

IMSGC, WTCCC2, S. Sawcer, G. Hellenthal, M. Pirinen, C. C. Spencer, N. A. Patsopoulos, L. Moutsianas, A. Dilthey, Z. Su, C. Freeman, S. E. Hunt, S. Edkins, E. Gray, D. R. Booth, S. C. Potter, A. Goris, G. Band, A. B. Oturai, A. Strange, J. Saarela, C. Bellenguez, B. Fontaine, M. Gillman, B. Hemmer, R. Gwilliam, F. Zipp, A. Jayakumar, R. Martin, S. Leslie, S. Hawkins, E. Giannoulatou, S. D'Alfonso, H. Blackburn, F. M. Boneschi, J. Liddle, H. F. Harbo, M. L. Perez, A. Spurkland, M. J. Waller, M. P. Mycko, M. Ricketts, M. Comabella, N. Hammond, I. Kockum, O. T. McCann, M. Ban, P. Whittaker, A. Kempainen, P. Weston, C. Hawkins, S. Widaa, J. Zajicek, S. Dronov, N. Robertson, S. J. Bumpstead, L. F. Barcellos, R. Ravindrarajah, R. Abraham, L. Alfredsson, K. Ardlie, C. Aubin, A. Baker, K. Baker, S. E. Baranzini, L. Bergamaschi, R. Bergamaschi, A. Bernstein, A. Berthele, M. Boggild, J. P. Bradfield, D. Brassat, S. A. Broadley, D. Buck, H. Butzkueven, R. Capra, W. M. Carroll, P. Cavalla, E. G. Celius, S. Cepok, R. Chiavacci, F. Clerget-Darpoux, K. Clysters, G. Comi, M.

Cosburn, I. Cournu-Rebeix, M. B. Cox, W. Cozen, B. A. Cree, A. H. Cross, D. Cusi, M. J. Daly, E. Davis, P. I. de Bakker, M. Debouverie, B. D'Hooghe M, K. Dixon, R. Dobosi, B. Dubois, D. Ellinghaus, I. Elovaara, F. Esposito, C. Fontenille, S. Foote, A. Franke, D. Galimberti, A. Ghezzi, J. Glessner, R. Gomez, O. Gout, C. Graham, S. F. Grant, F. R. Guerini, H. Hakonarson, P. Hall, A. Hamsten, H. P. Hartung, R. N. Heard, S. Heath, J. Hobart, M. Hoshi, C. Infante-Duarte, G. Ingram, W. Ingram, T. Islam, M. Jagodic, M. Kabesch, A. G. Kermodé, T. J. Kilpatrick, C. Kim, N. Klopp, K. Koivisto, M. Larsson, M. Lathrop, J. S. Lechner-Scott, M. A. Leone, V. Leppa, U. Liljedahl, I. L. Bomfim, R. R. Lincoln, J. Link, J. Liu, A. R. Lorentzen, S. Lupoli, F. Macciardi, T. Mack, M. Marriott, V. Martinelli, D. Mason, J. L. McCauley, F. Mentch, I. L. Mero, T. Mihalova, X. Montalban, J. Mottershead, K. M. Myhr, P. Naldi, W. Ollier, A. Page, A. Palotie, J. Pelletier, L. Piccio, T. Pickersgill, F. Piehl, S. Pobywajlo, H. L. Quach, P. P. Ramsay, M. Reunanen, R. Reynolds, J. D. Rioux, M. Rodegher, S. Roesner, J. P. Rubio, I. M. Ruckert, M. Salvetti, E. Salvi, A. Santaniello, C. A. Schaefer, S. Schreiber, C. Schulze, R. J. Scott, F. Sellebjerg, K. W. Selmaj, D. Sexton, L. Shen, B. Simms-Acuna, S. Skidmore, P. M. Sleiman, C. Smestad, P. S. Sorensen, H. B. Sondergaard, J. Stankovich, R. C. Strange, A. M. Sulonen, E. Sundqvist, A. C. Syvanen, F. Taddeo, B. Taylor, J. M. Blackwell, P. Tienari, E. Bramon, A. Tourbah, M. A. Brown, E. Tronczynska, J. P. Casas, N. Tubridy, A. Corvin, J. Vickery, J. Jankowski, P. Villoslada, H. S. Markus, K. Wang, C. G. Mathew, J. Wason, C. N. Palmer, H. E. Wichmann, R. Plomin, E. Willoughby, A. Rautanen, J. Winkelmann, M. Wittig, R. C. Trembath, J. Yaouanq, A. C. Viswanathan, H. Zhang, N. W. Wood, R. Zuvich, P. Deloukas, C. Langford, A. Duncanson, J. R. Oksenberg, M. A. Pericak-Vance, J. L. Haines, T. Olsson, J. Hillert, A. J. Iverson, P. L. De Jager, L. Peltonen, G. J. Stewart, D. A. Hafler, S. L. Hauser, G. McVean, P. Donnelly and A. Compston (2011). "Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis." *Nature* **476**(7359): 214-219.

International HapMap, C. (2005). "A haplotype map of the human genome." *Nature* **437**(7063): 1299-1320.

International Schizophrenia, C. (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." *Nature* **455**(7210): 237-241.

Isobe, N., P. A. Gourraud, H. F. Harbo, S. J. Caillier, A. Santaniello, P. Khankhanian, M. Maiers, S. Spellman, N. Cereb, S. Yang, M. J. Pando, L. Piccio, A. H. Cross, P. L. De Jager, B. A. Cree, S. L. Hauser and J. R. Oksenberg (2013). "Genetic risk variants in African Americans with multiple sclerosis." *Neurology*.

Jagodic, M., C. Colacios, R. Nohra, A. S. Dejean, A. D. Beyeen, M. Khademi, A. Casemayou, L. Lamouroux, C. Duthoit, O. Papapietro, L. Sjöholm, I. Bernard, D. Lagrange, I. Dahlman, F. Lundmark, A. B. Oturai, H. B. Soendergaard, A. Kemppinen, J. Saarela, P. J. Tienari, H. F. Harbo, A. Spurkland, S. V. Ramagopalan, D. A. Sadovnick, G. C. Ebers, M. Seddighzadeh, L. Klareskog, L. Alfredsson, L. Padyukov, J. Hillert, M. Clanet, G. Edan, B. Fontaine, G. J. Fournie, I. Kockum, A. Saoudi and T. Olsson (2009). "A role for VAV1 in experimental autoimmune encephalomyelitis and multiple sclerosis." *Sci Transl Med* **1**(10): 10ra21.

Jensen, C. J., J. Stankovich, A. Van der Walt, M. Bahlo, B. V. Taylor, I. A. van der Mei, S. J. Foote, T. J. Kilpatrick, L. J. Johnson, E. Wilkins, J. Field, P. Danoy, M. A. Brown, J. P. Rubio and H. Butzkueven (2010). "Multiple sclerosis susceptibility-associated SNPs do not

influence disease severity measures in a cohort of Australian MS patients." PLoS One **5**(4): e10003.

Jersild, C., T. Fog, G. S. Hansen, M. Thomsen, A. Svejgaard and B. Dupont (1973). "Histocompatibility determinants in multiple sclerosis, with special reference to clinical course." Lancet **2**(7840): 1221-1225.

Jersild, C., A. Svejgaard and T. Fog (1972). "HL-A antigens and multiple sclerosis." Lancet **1**(7762): 1240-1241.

Johnson, K. P., B. R. Brooks, J. A. Cohen, C. C. Ford, J. Goldstein, R. P. Lisak, L. W. Myers, H. S. Panitch, J. W. Rose and R. B. Schiffer (1995). "Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group." Neurology **45**(7): 1268-1276.

Jostins, L. and J. C. Barrett (2011). "Genetic risk prediction in complex disease." Hum Mol Genet **20**(R2): R182-188.

Kampman, M. T., T. Wilsgaard and S. I. Mellgren (2007). "Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic Circle." J Neurol **254**(4): 471-477.

Kappos, L., E. W. Radue, P. O'Connor, C. Polman, R. Hohlfeld, P. Calabresi, K. Selmaj, C. Agoropoulou, M. Leyk, L. Zhang-Auberson, P. Burtin and F. S. Group (2010). "A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis." N Engl J Med **362**(5): 387-401.

Kurtuncu, M. and E. Tuzun (2008). "Multiple sclerosis: could it be an epigenetic disease?" Med Hypotheses **71**(6): 945-947.

Kurtzke, J. F. (1983). "Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)." Neurology **33**(11): 1444-1452.

Lavasani, S., B. Dzhambazov, M. Nouri, F. Fak, S. Buske, G. Molin, H. Thorlacius, J. Alenfall, B. Jeppsson and B. Westrom (2010). "A novel probiotic mixture exerts a therapeutic effect on experimental autoimmune encephalomyelitis mediated by IL-10 producing regulatory T cells." PLoS One **5**(2): e9009.

Lee, S. H., D. Harold, D. R. Nyholt, M. E. Goddard, K. T. Zondervan, J. Williams, G. W. Montgomery, N. R. Wray and P. M. Visscher (2013). "Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis." Hum Mol Genet **22**(4): 832-841.

Lindsey, J. W. (2005). "Familial recurrence rates and genetic models of multiple sclerosis." Am J Med Genet A **135**(1): 53-58.

Lublin, F. D. and S. C. Reingold (1996). "Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis." Neurology **46**(4): 907-911.

Lundmark, F., K. Duvefelt, E. Iacobaeus, I. Kockum, E. Wallstrom, M. Khademi, A. Oturai, L. P. Ryder, J. Saarela, H. F. Harbo, E. G. Celius, H. Salter, T. Olsson and J. Hillert (2007). "Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis." Nat Genet **39**(9): 1108-1113.

Maher, B. (2008). "Personal genomes: The case of the missing heritability." Nature **456**(7218): 18-21.

Man, S., E. E. Ubogu and R. M. Ransohoff (2007). "Inflammatory cell migration into the central nervous system: a few new twists on an old tale." Brain Pathol **17**(2): 243-250.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll and P. M. Visscher (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Marian, A. J. (2012). "Elements of 'missing heritability'." Curr Opin Cardiol **27**(3): 197-201.

Marrie, R. A. (2004). "Environmental risk factors in multiple sclerosis aetiology." Lancet Neurol **3**(12): 709-718.

McClellan, J. and M. C. King (2010). "Genetic heterogeneity in human disease." Cell **141**(2): 210-217.

McDonald, W. I., A. Compston, G. Edan, D. Goodkin, H. P. Hartung, F. D. Lublin, H. F. McFarland, D. W. Paty, C. H. Polman, S. C. Reingold, M. Sandberg-Wollheim, W. Sibley, A. Thompson, S. van den Noort, B. Y. Weinschenker and J. S. Wolinsky (2001). "Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis." Ann Neurol **50**(1): 121-127.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.

Miller, D. T., Y. Shen, L. A. Weiss, J. Korn, I. Anselm, C. Bridgemohan, G. F. Cox, H. Dickinson, J. Gentile, D. J. Harris, V. Hegde, R. Hundley, O. Khwaja, S. Kothare, C. Luedke, R. Nasir, A. Poduri, K. Prasad, P. Raffalli, A. Reinhard, S. E. Smith, M. M. Sobeih, J. S. Soul, J. Stoler, M. Takeoka, W. H. Tan, J. Thakuria, R. Wolff, R. Yusupov, J. F. Gusella, M. J. Daly and B. L. Wu (2009). "Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders." J Med Genet **46**(4): 242-248.

Milo, R. and E. Kahana (2010). "Multiple sclerosis: geoepidemiology, genetics and the environment." Autoimmun Rev **9**(5): A387-394.

Miretti, M. M., E. C. Walsh, X. Ke, M. Delgado, M. Griffiths, S. Hunt, J. Morrison, P. Whittaker, E. S. Lander, L. R. Cardon, D. R. Bentley, J. D. Rioux, S. Beck and P. Deloukas (2005). "A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms." Am J Hum Genet **76**(4): 634-646.

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

Nylander, A. and D. A. Hafler (2012). "Multiple sclerosis." J Clin Invest **122**(4): 1180-1188.

Oksenberg, J. R., S. E. Baranzini, S. Sawcer and S. L. Hauser (2008). "The genetics of multiple sclerosis: SNPs to pathways to pathogenesis." Nat Rev Genet **9**(7): 516-526.

Oldstone, M. B. (1998). "Molecular mimicry and immune-mediated diseases." FASEB J **12**(13): 1255-1265.

Orton, S. M., B. M. Herrera, I. M. Yee, W. Valdar, S. V. Ramagopalan, A. D. Sadovnick, G. C. Ebers and G. Canadian Collaborative Study (2006). "Sex ratio of multiple sclerosis in Canada: a longitudinal study." Lancet Neurol **5**(11): 932-936.

Patsopoulos, N. A., F. Esposito, J. Reischl, S. Lehr, D. Bauer, J. Heubach, R. Sandbrink, C. Pohl, G. Edan, L. Kappos, D. Miller, J. Montalban, C. H. Polman, M. S. Freedman, H. P. Hartung, B. G. Arnason, G. Comi, S. Cook, M. Filippi, D. S. Goodin, D. Jeffery, P. O'Connor, G. C. Ebers, D. Langdon, A. T. Reder, A. Traboulsee, F. Zipp, S. Schimrigk, J. Hillert, M. Bahlo, D. R. Booth, S. Broadley, M. A. Brown, B. L. Browning, S. R. Browning, H. Butzkueven, W. M. Carroll, C. Chapman, S. J. Foote, L. Griffiths, A. G. Kermode, T. J. Kilpatrick, J. Lechner-Scott, M. Marriott, D. Mason, P. Moscato, R. N. Heard, M. P. Pender, V. M. Perreau, D. Perera, J. P. Rubio, R. J. Scott, M. Slee, J. Stankovich, G. J. Stewart, B. V. Taylor, N. Tubridy, E. Willoughby, J. Wiley, P. Matthews, F. M. Boneschi, A. Compston, J. Haines, S. L. Hauser, J. McCauley, A. Ivinson, J. R. Oksenberg, M. Pericak-Vance, S. J. Sawcer, P. L. De Jager, D. A. Hafler and P. I. de Bakker (2011). "Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci." Ann Neurol **70**(6): 897-912.

Pauwels, I., L. Cosemans, S. Boonen, B. Dubois and A. Goris (2013). "Rare MEFV variants are not associated with risk to develop multiple sclerosis and severity of disease." Mult Scler.

Pinto, D., A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams, J. Almeida, E. Bacchelli, G. D. Bader, A. J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bolte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S. E. Bryson, A. R. Carson, G. Casallo, J. Casey, B. H. Chung, L. Cochrane, C. Corsello, E. L. Crawford, A. Crossett, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, C. M. Freitag, J. Gilbert, C. Gillberg, J. T. Glessner, J. Goldberg, A. Green, J. Green, S. J. Guter, H. Hakonarson, E. A. Heron, M. Hill, R. Holt, J. L. Howe, G.

Hughes, V. Hus, R. Iglizzi, C. Kim, S. M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B. L. Leventhal, A. C. Lionel, X. Q. Liu, C. Lord, L. Lotspeich, S. C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C. R. Marshall, H. McConachie, C. J. McDougle, J. McGrath, W. M. McMahon, A. Merikangas, O. Migita, N. J. Minshew, G. K. Mirza, J. Munson, S. F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J. R. Parr, B. Parrini, T. Paton, A. Pickles, M. Pilorge, J. Piven, C. P. Ponting, D. J. Posey, A. Poustka, F. Poustka, A. Prasad, J. Ragoussis, K. Renshaw, J. Rickaby, W. Roberts, K. Roeder, B. Roge, M. L. Rutter, L. J. Bierut, J. P. Rice, J. Salt, K. Sansom, D. Sato, R. Segurado, A. F. Sequeira, L. Senman, N. Shah, V. C. Sheffield, L. Soorya, I. Sousa, O. Stein, N. Sykes, V. Stoppioni, C. Strawbridge, R. Tancredi, K. Tansey, B. Thiruvahindrapduram, A. P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J. B. Vincent, F. Volkmar, S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Wittmeyer, S. Wood, J. Wu, B. L. Yaspan, D. Zurawiecki, L. Zwaigenbaum, J. D. Buxbaum, R. M. Cantor, E. H. Cook, H. Coon, M. L. Cuccaro, B. Devlin, S. Ennis, L. Gallagher, D. H. Geschwind, M. Gill, J. L. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger, Jr., A. D. Paterson, M. A. Pericak-Vance, G. D. Schellenberg, P. Szatmari, A. M. Vicente, V. J. Vieland, E. M. Wijsman, S. W. Scherer, J. S. Sutcliffe and C. Betancur (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**(7304): 368-372.

Polman, C. H., P. W. O'Connor, E. Havrdova, M. Hutchinson, L. Kappos, D. H. Miller, J. T. Phillips, F. D. Lublin, G. Giovannoni, A. Wajgt, M. Toal, F. Lynn, M. A. Panzara and A. W. Sandrock (2006). "A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis." N Engl J Med **354**(9): 899-910.

Polman, C. H., S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. O'Connor, M. Sandberg-Wollheim, A. J. Thompson, E. Waubant, B. Weinshenker and J. S. Wolinsky (2011). "Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria." Ann Neurol **69**(2): 292-302.

Polychronakos, C. (2011). "Fine points in mapping autoimmunity." Nat Genet **43**(12): 1173-1174.

Poser, C. M., D. W. Paty, L. Scheinberg, W. I. McDonald, F. A. Davis, G. C. Ebers, K. P. Johnson, W. A. Sibley, D. H. Silberberg and W. W. Tourtellotte (1983). "New diagnostic criteria for multiple sclerosis: guidelines for research protocols." Ann Neurol **13**(3): 227-231.

Pritchard, J. K. and N. J. Cox (2002). "The allelic architecture of human disease genes: common disease-common variant...or not?" Hum Mol Genet **11**(20): 2417-2423.

Ramagopalan, S. V. and G. C. Ebers (2009). "Epistasis: multiple sclerosis and the major histocompatibility complex." Neurology **72**(6): 566-567.

Ramagopalan, S. V., N. J. Maugeri, L. Handunnetthi, M. R. Lincoln, S. M. Orton, D. A. Dyment, G. C. Deluca, B. M. Herrera, M. J. Chao, A. D. Sadovnick, G. C. Ebers and J. C. Knight (2009). "Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1\*1501 is regulated by vitamin D." PLoS Genet **5**(2): e1000369.

Ramanan, V. K., L. Shen, J. H. Moore and A. J. Saykin (2012). "Pathway analysis of genomic data: concepts, methods, and prospects for future development." Trends Genet **28**(7): 323-332.

Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-510.

Reich, T., J. W. James and C. A. Morris (1972). "The use of multiple thresholds in determining the mode of transmission of semi-continuous traits." Ann Hum Genet **36**(2): 163-184.

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-1517.

Rodgers, J. M. and S. D. Miller (2012). "Cytokine control of inflammation and repair in the pathology of multiple sclerosis." Yale J Biol Med **85**(4): 447-468.

Roxburgh, R. H., S. R. Seaman, T. Masterman, A. E. Hensiek, S. J. Sawcer, S. Vukusic, I. Achiti, C. Confavreux, M. Coustans, E. le Page, G. Edan, G. V. McDonnell, S. Hawkins, M. Trojano, M. Liguori, E. Cocco, M. G. Marrosu, F. Tesser, M. A. Leone, A. Weber, F. Zipp, B. Milterski, J. T. Epplen, A. Oturai, P. S. Sorensen, E. G. Celius, N. T. Lara, X. Montalban, P. Villoslada, A. M. Silva, M. Marta, I. Leite, B. Dubois, J. Rubio, H. Butzkueven, T. Kilpatrick, M. P. Mycko, K. W. Selmaj, M. E. Rio, M. Sa, G. Salemi, G. Savettieri, J. Hillert and D. A. Compston (2005). "Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity." Neurology **64**(7): 1144-1151.

Sanna, S., M. Pitzalis, M. Zoledziewska, I. Zara, C. Sidore, R. Murru, M. B. Whalen, F. Busonero, A. Maschio, G. Costa, M. C. Melis, F. Deidda, F. Poddie, L. Morelli, G. Farina, Y. Li, M. Dei, S. Lai, A. Mulas, G. Cuccuru, E. Porcu, L. Liang, P. Zavattari, L. Moi, E. Deriu, M. F. Urru, M. Bajorek, M. A. Satta, E. Cocco, P. Ferrigno, S. Sotgiu, M. Pugliatti, S. Traccis, A. Angius, M. Melis, G. Rosati, G. R. Abecasis, M. Uda, M. G. Marrosu, D. Schlessinger and F. Cucca (2010). "Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis." Nat Genet **42**(6): 495-497.

Sato, D., A. C. Lionel, C. S. Leblond, A. Prasad, D. Pinto, S. Walker, I. O'Connor, C. Russell, I. E. Drmic, F. F. Hamdan, J. L. Michaud, V. Endris, R. Roeth, R. Delorme, G. Huguet, M. Leboyer, M. Rastam, C. Gillberg, M. Lathrop, D. J. Stavropoulos, E. Anagnostou, R. Weksberg, E. Fombonne, L. Zwaigenbaum, B. A. Fernandez, W. Roberts, G. A. Rappold, C. R. Marshall, T. Bourgeron, P. Szatmari and S. W. Scherer (2012). "SHANK1 Deletions in Males with Autism Spectrum Disorder." Am J Hum Genet **90**(5): 879-887.

Sawcer, S. (2008). "The complex genetics of multiple sclerosis: pitfalls and prospects." Brain **131**(Pt 12): 3118-3131.

Sawcer, S., M. Ban, M. Maranian, T. W. Yeo, A. Compston, A. Kirby, M. J. Daly, P. L. De Jager, E. Walsh, E. S. Lander, J. D. Rioux, D. A. Hafler, A. Ivinson, J. Rimmler, S. G. Gregory, S. Schmidt, M. A. Pericak-Vance, E. Akesson, J. Hillert, P. Datta, A. Oturai, L. P. Ryder, H. F. Harbo, A. Spurkland, K. M. Myhr, M. Laaksonen, D. Booth, R. Heard, G.



Stewart, R. Lincoln, L. F. Barcellos, S. L. Hauser, J. R. Oksenberg, S. J. Kenealy and J. L. Haines (2005). "A high-density screen for linkage in multiple sclerosis." Am J Hum Genet **77**(3): 454-467.

Schumacker, G. A., G. Beebe, R. F. Kibler, L. T. Kurland, J. F. Kurtzke, F. McDowell, B. Nagler, W. A. Sibley, W. W. Tourtellotte and T. L. Willmon (1965). "Problems of Experimental Trials of Therapy in Multiple Sclerosis: Report by the Panel on the Evaluation of Experimental Trials of Therapy in Multiple Sclerosis." Ann N Y Acad Sci **122**: 552-568.

Sirota, M., M. A. Schaub, S. Batzoglou, W. H. Robinson and A. J. Butte (2009). "Autoimmune disease classification by inverse association with SNP alleles." PLoS Genet **5**(12): e1000792.

Sloka, J. S., W. E. Pryse-Phillips and M. Stefanelli (2008). "The relation of ultraviolet radiation and multiple sclerosis in Newfoundland." Can J Neurol Sci **35**(1): 69-74.

Sloka, S., C. Silva, W. Pryse-Phillips, S. Patten, L. Metz and V. W. Yong (2011). "A quantitative analysis of suspected environmental causes of MS." Can J Neurol Sci **38**(1): 98-105.

Smith, D. J. and A. J. Lusk (2002). "The allelic structure of common disease." Hum Mol Genet **11**(20): 2455-2461.

Sobocki, P., M. Pugliatti, K. Lauer and G. Kobelt (2007). "Estimation of the cost of MS in Europe: extrapolations from a multinational cost study." Mult Scler **13**(8): 1054-1064.

Sospedra, M. and R. Martin (2005). "Immunology of multiple sclerosis." Annu Rev Immunol **23**: 683-747.

Stefansson, H., D. Rujescu, S. Cichon, O. P. Pietilainen, A. Ingason, S. Steinberg, R. Fossdal, E. Sigurdsson, T. Sigmundsson, J. E. Buizer-Voskamp, T. Hansen, K. D. Jakobsen, P. Muglia, C. Francks, P. M. Matthews, A. Gylfason, B. V. Halldorsson, D. Gudbjartsson, T. E. Thorgeirsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, A. Bjornsson, S. Mattiasdottir, T. Blondal, M. Haraldsson, B. B. Magnusdottir, I. Giegling, H. J. Moller, A. Hartmann, K. V. Shianna, D. Ge, A. C. Need, C. Crombie, G. Fraser, N. Walker, J. Lonnqvist, J. Suvisaari, A. Tuulio-Henriksson, T. Paunio, T. Touloupoulou, E. Bramon, M. Di Forti, R. Murray, M. Ruggeri, E. Vassos, S. Tosato, M. Walshe, T. Li, C. Vasilescu, T. W. Muhleisen, A. G. Wang, H. Ullum, S. Djurovic, I. Melle, J. Olesen, L. A. Kiemenev, B. Franke, Group, C. Sabatti, N. B. Freimer, J. R. Gulcher, U. Thorsteinsdottir, A. Kong, O. A. Andreassen, R. A. Ophoff, A. Georgi, M. Rietschel, T. Werge, H. Petursson, D. B. Goldstein, M. M. Nothen, L. Peltonen, D. A. Collier, D. St Clair and K. Stefansson (2008). "Large recurrent microdeletions associated with schizophrenia." Nature **455**(7210): 232-236.

Steinman, L. (2001). "Multiple sclerosis: a two-stage disease." Nat Immunol **2**(9): 762-764.

Tourdias, T., B. Brochet, K. G. Petry and V. Dousset (2009). "[Magnetic resonance imaging of central nervous system inflammation]." Rev Neurol (Paris) **165 Suppl 3**: S77-87.

Trynka, G., K. A. Hunt, N. A. Bockett, J. Romanos, V. Mistry, A. Szperl, S. F. Bakker, M. T. Bardella, L. Bhaw-Rosun, G. Castillejo, E. G. de la Concha, R. C. de Almeida, K. R. Dias, C. C. van Diemen, P. C. Dubois, R. H. Duerr, S. Edkins, L. Franke, K. Fransen, J. Gutierrez, G. A. Heap, B. Hrdlickova, S. Hunt, L. Plaza Izurieta, V. Izzo, L. A. Joosten, C. Langford, M. C. Mazzilli, C. A. Mein, V. Midah, M. Mitrovic, B. Mora, M. Morelli, S. Nutland, C. Nunez, S. Onengut-Gumuscu, K. Pearce, M. Platteel, I. Polanco, S. Potter, C. Ribes-Koninckx, I. Ricano-Ponce, S. S. Rich, A. Rybak, J. L. Santiago, S. Senapati, A. Sood, H. Szajewska, R. Troncone, J. Varade, C. Wallace, V. M. Wolters, A. Zhernakova, D. Spanish Consortium on the Genetics of Coeliac, C. D. S. G. Prevent, C. Wellcome Trust Case Control, B. K. Thelma, B. Cukrowska, E. Urcelay, J. R. Bilbao, M. L. Mearin, D. Barisani, J. C. Barrett, V. Plagnol, P. Deloukas, C. Wijmenga and D. A. van Heel (2011). "Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease." Nat Genet **43**(12): 1193-1201.

van der Mei, I. A., A. L. Ponsonby, L. Blizzard and T. Dwyer (2001). "Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation." Neuroepidemiology **20**(3): 168-174.

Vermersch, P. (2009). "[Pragmatic approach to treatments in multiple sclerosis]." Rev Neurol (Paris) **165 Spec No 2**: F125-131.

Visscher, P. M., J. Yang and M. E. Goddard (2010). "A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010)." Twin Res Hum Genet **13**(6): 517-524.

Wang, J. H., D. Pappas, P. L. De Jager, D. Pelletier, P. I. de Bakker, L. Kappos, C. H. Polman, L. B. Chibnik, D. A. Hafler, P. M. Matthews, S. L. Hauser, S. E. Baranzini and J. R. Oksenberg (2011). "Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data." Genome Med **3**(1): 3.

Wang, K., M. Li and H. Hakonarson (2010). "Analysing biological pathways in genome-wide association studies." Nat Rev Genet **11**(12): 843-854.

Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd (2005). "Genome-wide association studies: theoretical and practical concerns." Nat Rev Genet **6**(2): 109-118.

Weiss, L. A., Y. Shen, J. M. Korn, D. E. Arking, D. T. Miller, R. Fossdal, E. Saemundsen, H. Stefansson, M. A. Ferreira, T. Green, O. S. Platt, D. M. Ruderfer, C. A. Walsh, D. Altshuler, A. Chakravarti, R. E. Tanzi, K. Stefansson, S. L. Santangelo, J. F. Gusella, P. Sklar, B. L. Wu, M. J. Daly and C. Autism (2008). "Association between microdeletion and microduplication at 16p11.2 and autism." N Engl J Med **358**(7): 667-675.

Weiss, N., F. Miller, S. Cazaubon and P. O. Couraud (2009). "[Biology of the blood-brain barrier: Part I]." Rev Neurol (Paris) **165**(11): 863-874.

Wellcome Trust Case Control, C., N. Craddock, M. E. Hurles, N. Cardin, R. D. Pearson, V. Plagnol, S. Robson, D. Vukcevic, C. Barnes, D. F. Conrad, E. Giannoulatou, C. Holmes, J. L. Marchini, K. Stirrups, M. D. Tobin, L. V. Wain, C. Yau, J. Aerts, T. Ahmad, T. D. Andrews,

H. Arbury, A. Attwood, A. Auton, S. G. Ball, A. J. Balmforth, J. C. Barrett, I. Barroso, A. Barton, A. J. Bennett, S. Bhaskar, K. Blaszczyk, J. Bowes, O. J. Brand, P. S. Braund, F. Bredin, G. Breen, M. J. Brown, I. N. Bruce, J. Bull, O. S. Burren, J. Burton, J. Byrnes, S. Caesar, C. M. Clee, A. J. Coffey, J. M. Connell, J. D. Cooper, A. F. Dominiczak, K. Downes, H. E. Drummond, D. Dudakia, A. Dunham, B. Ebbs, D. Eccles, S. Edkins, C. Edwards, A. Elliot, P. Emery, D. M. Evans, G. Evans, S. Eyre, A. Farmer, I. N. Ferrier, L. Feuk, T. Fitzgerald, E. Flynn, A. Forbes, L. Forty, J. A. Franklyn, R. M. Freathy, P. Gibbs, P. Gilbert, O. Gokumen, K. Gordon-Smith, E. Gray, E. Green, C. J. Groves, D. Grozeva, R. Gwilliam, A. Hall, N. Hammond, M. Hardy, P. Harrison, N. Hassanali, H. Hebaishi, S. Hines, A. Hinks, G. A. Hitman, L. Hocking, E. Howard, P. Howard, J. M. Howson, D. Hughes, S. Hunt, J. D. Isaacs, M. Jain, D. P. Jewell, T. Johnson, J. D. Jolley, I. R. Jones, L. A. Jones, G. Kirov, C. F. Langford, H. Lango-Allen, G. M. Lathrop, J. Lee, K. L. Lee, C. Lees, K. Lewis, C. M. Lindgren, M. Maisuria-Armer, J. Maller, J. Mansfield, P. Martin, D. C. Massey, W. L. McArdle, P. McGuffin, K. E. McLay, A. Mentzer, M. L. Mimmack, A. E. Morgan, A. P. Morris, C. Mowat, S. Myers, W. Newman, E. R. Nimmo, M. C. O'Donovan, A. Onipinla, I. Onyiah, N. R. Ovington, M. J. Owen, K. Palin, K. Parnell, D. Pernet, J. R. Perry, A. Phillips, D. Pinto, N. J. Prescott, I. Prokopenko, M. A. Quail, S. Rafelt, N. W. Rayner, R. Redon, D. M. Reid, Renwick, S. M. Ring, N. Robertson, E. Russell, D. St Clair, J. G. Sambrook, J. D. Sanderson, H. Schuilenburg, C. E. Scott, R. Scott, S. Seal, S. Shaw-Hawkins, B. M. Shields, M. J. Simmonds, D. J. Smyth, E. Somaskantharajah, K. Spanova, S. Steer, J. Stephens, H. E. Stevens, M. A. Stone, Z. Su, D. P. Symmons, J. R. Thompson, W. Thomson, M. E. Travers, C. Turnbull, A. Valsesia, M. Walker, N. M. Walker, C. Wallace, M. Warren-Perry, N. A. Watkins, J. Webster, M. N. Weedon, A. G. Wilson, M. Woodburn, B. P. Wordsworth, A. H. Young, E. Zeggini, N. P. Carter, T. M. Frayling, C. Lee, G. McVean, P. B. Munroe, A. Palotie, S. J. Sawcer, S. W. Scherer, D. P. Strachan, C. Tyler-Smith, M. A. Brown, P. R. Burton, M. J. Caulfield, A. Compston, M. Farrall, S. C. Gough, A. S. Hall, A. T. Hattersley, A. V. Hill, C. G. Mathew, M. Pembrey, J. Satsangi, M. R. Stratton, J. Worthington, P. Deloukas, A. Duncanson, D. P. Kwiatkowski, M. I. McCarthy, W. Ouwehand, M. Parkes, N. Rahman, J. A. Todd, N. J. Samani and P. Donnelly (2010). "Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls." *Nature* **464**(7289): 713-720.

Whitacre, C. C. (2001). "Sex differences in autoimmune disease." *Nat Immunol* **2**(9): 777-780.

Wolburg, H. and A. Lippoldt (2002). "Tight junctions of the blood-brain barrier: development, composition and regulation." *Vascul Pharmacol* **38**(6): 323-337.

Wray, N. R., J. Yang, B. J. Hayes, A. L. Price, M. E. Goddard and P. M. Visscher (2013). "Pitfalls of predicting complex traits from SNPs." *Nat Rev Genet* **14**(7): 507-515.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard and P. M. Visscher (2010). "Common SNPs explain a large proportion of the heritability for human height." *Nat Genet* **42**(7): 565-569.

Yang, Q., M. J. Khoury, J. Friedman, J. Little and W. D. Flanders (2005). "How many genes underlie the occurrence of common complex diseases in the population?" *Int J Epidemiol* **34**(5): 1129-1137.

Yaouanq, J., G. Semana, S. Eichenbaum, E. Quelvennec, M. P. Roth, M. Clanet, G. Edan and F. Clerget-Darpoux (1997). "Evidence for linkage disequilibrium between HLA-DRB1 gene and multiple sclerosis. The French Research Group on Genetic Susceptibility to MS." Science **276**(5313): 664-665.

Yednock, T. A., C. Cannon, L. C. Fritz, F. Sanchez-Madrid, L. Steinman and N. Karin (1992). "Prevention of experimental autoimmune encephalomyelitis by antibodies against alpha 4 beta 1 integrin." Nature **356**(6364): 63-66.

Zhernakova, A., C. C. van Diemen and C. Wijmenga (2009). "Detecting shared pathogenesis from the shared genetics of immune-related diseases." Nat Rev Genet **10**(1): 43-55.

Zhong, H., X. Yang, L. M. Kaplan, C. Molony and E. E. Schadt (2010). "Integrating pathway analysis and genetics of gene expression for genome-wide association studies." Am J Hum Genet **86**(4): 581-591.



# **GÉNÉTIQUE DE LA SCLÉROSE EN PLAQUES :**

## **Héritabilité manquante et Charge génétique**

### **Résumé**

La Sclérose En Plaques (SEP) est une maladie multifactorielle résultant de l'interaction de facteurs environnementaux, aujourd'hui inconnus, et de facteurs génétiques. Ces dernières années, la recherche génétique a connu de grandes avancées. En 2011, une importante étude internationale permis d'identifier 57 régions non-HLA de prédisposition à la SEP, expliquant ainsi 17% de sa composante génétique, les 83% restants étant appelés « héritabilité manquante ».

Au cours de mon doctorat, j'ai mené ou participé à différents projets dans le but d'identifier cette héritabilité manquante et dont les hypothèses reposaient sur l'augmentation de la taille des échantillons pour augmenter la puissance statistique des études, l'identification des variants causaux des régions génomiques mises en évidence lors des études précédentes, l'étude des régions non investiguées jusqu'à présent et des variants interagissant au sein d'un même réseau biologique et l'étude de charge génétique individuelle.

Un des projets auquel j'ai participé a permis d'augmenter le nombre de variants connus associés à la SEP à 110, hors HLA. Cependant, 77% de la composante génétique de la maladie reste encore à identifier.

D'autres hypothèses sont à explorer comme par exemple l'identification de variants de prédisposition spécifiques à une sous-population, l'identification de variants de risque localisés dans des régions déjà connues mais qui ne seraient présents que chez les individus non-porteurs des allèles de risque déjà identifiés dans ces régions.

Le défi majeur des années à venir sera de corréliser les variants de risque à leurs effets fonctionnels et des profils génétiques spécifiques à un phénomène SEP exhaustif.

**Mots clés :** Sclérose En Plaques, Maladie complexe, Héritabilité manquante, Charge génétique, Prédisposition, Sévérité, Association, Réseaux de gènes

### **Résumé en anglais**

Multiple Sclerosis (MS) is a multifactorial disease, resulting from an interaction between genetics and unknown environmental factors. During the last years, the identification of several predisposing genetic polymorphisms constituted a real breakthrough in the understanding of MS genetic architecture.

In 2011, a large international study identified 57 non-MHC genomic regions associated with MS susceptibility. However, the genetic part of the disease explained by these factors was evaluated to only be 17% of total MS heritability.

During my PhD, I leaded or participated to several projects aiming to identify the so-called « missing heritability », based on different strategies : to increase the sample sizes in order to amplify the statistical power of the studies, to identify the causative variants within previously highlighted genomic regions, to study un-investigated genomic regions and genetic interactions between variants of a same biological pathway, to study the individual genetic burden.

One of the projects extended the number of MS non-MHC-variants to 110, increasing the known part of heritability from 17% to 23%. In order to go further in MS genetics, other hypotheses should be explored, as (a) the identification of variants specific to one sub-population (French, English ...), (b) the identification of new risk variants localized in previously highlighted regions in populations of non-carriers of known risk alleles of these regions, (c) the study of the Copy-Number Variations.

Finally, the major challenge of next years will be to correlate genetic risk variants to functional effects and to comprehensive MS Phenome.

**Keywords :** Multiple sclerosis, Complex disease, Missing heritability, Genetic burden, Susceptibility, Severity, Association, Pathways