Epigenetics and Autoimmune Disease

Lisa F. Barcellos, PhD, MPH
Associate Professor
UC Berkeley
School of Public Health
QB3 Genetic Epidemiology and Genomics Laboratory

ACTRIMS, May 30, 2014
Dallas, TX

Outline for Today

• Some background: genetics, environment and MS
• Importance of considering epigenetics in complex diseases
• Technology for studying DNA methylation
• Issues related to study design and available tools
• Examples from MS and other autoimmune disease studies in progress
Evidence for Genetic Factors in MS

- Familial aggregation of MS cases (10%)
- Concordance in MZ twins: 25-35%; DZ twins: 3-5%
- Increased relative risk to siblings
- Ch. 6p21 Major Histocompatibility Complex (MHC) region
  Class II HLA association (DRB1*15:01)
- ~50% of MS cases have at least one copy of this genetic variant

Published Genome-Wide Associations through 3/2010, 779 published GWA at p≤5x10^{-8} for 148 traits

NHGRI GWA Catalog
www.genome.gov/GWASudies
Modest Effects of Many Genes in Multiple Sclerosis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Associated single nucleotide polymorphism</th>
<th>Location</th>
<th>Risk allele</th>
<th>Estimated odds ratio</th>
<th>Protein function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD58</td>
<td>rs2300747</td>
<td>1p13</td>
<td>A</td>
<td>1.18</td>
<td>Cell adhesion; Immunological co-stimulation</td>
</tr>
<tr>
<td>EM5</td>
<td>rs10735781</td>
<td>1p22.1</td>
<td>G</td>
<td>1.23</td>
<td>Rab GTPase activator; cell cycle</td>
</tr>
<tr>
<td>RGS1</td>
<td>rs2760524</td>
<td>1q31</td>
<td>G</td>
<td>1.13</td>
<td>GTPase activator</td>
</tr>
<tr>
<td>IL7R</td>
<td>rs8897932</td>
<td>5p13</td>
<td>C</td>
<td>1.04</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>IL2RA</td>
<td>rs2104286</td>
<td>10p15-p14</td>
<td>T</td>
<td>1.18</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>CD6</td>
<td>rs17824533</td>
<td>11q13</td>
<td>G</td>
<td>1.16</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>rs1800693</td>
<td>1p21.3.2</td>
<td>C</td>
<td>1.23</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>CLEC16a</td>
<td>rs12708716</td>
<td>16p13.3</td>
<td>A</td>
<td>1.17</td>
<td>Not determined</td>
</tr>
<tr>
<td>IRF8</td>
<td>rs17445836</td>
<td>16p24.1</td>
<td>G</td>
<td>1.33</td>
<td>Transcription</td>
</tr>
<tr>
<td>CD206</td>
<td>rs763361</td>
<td>18q22.3</td>
<td>T</td>
<td>1.04</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>TYK2</td>
<td>rs34356443</td>
<td>19p13.2</td>
<td>G</td>
<td>1.30</td>
<td>Intracellular signaling</td>
</tr>
<tr>
<td>CD40</td>
<td>p66074022</td>
<td>20q12</td>
<td>G</td>
<td>1.22</td>
<td>Immunological co-stimulatory receptor</td>
</tr>
</tbody>
</table>


LETTER

doi:10.1038/nature10251

Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis

The International Multiple Sclerosis Genetics Consortium & the Wellcome Trust Case Control Consortium 2

Multiple sclerosis is a common disease of the central nervous system in which the interplay between inflammatory and neurodegenerative processes typically results in intermittent neurological disturbance followed by progressive accumulation of disability. Epidemiological studies have shown that genetic factors are primarily responsible for the substantially increased frequency of the disease seen in the relatives of affected individuals, and systematic attempts to identify linkage in multiplex families have confirmed that variation within the major histocompatibility complex (MHC) exerts the greatest individual effect on risk. Moderately powered genome-wide association studies (GWAS) have enabled more than 20 additional risk loci to be identified and have shown that multiple variants exerting modest individual effects have a key role in disease susceptibility. Most of the genetic variation in the genome is found in non-MHC regions, and a recent study identified a modest effect of a non-MHC variant in the CD58 gene on multiple sclerosis risk. Here we report a genome-wide association study of 6,087 cases and 14,890 controls of European ancestry. We identified a risk allele in the CD58 gene that was associated with multiple sclerosis at a genome-wide significant level (P = 2.0 × 10^-12) and confirmed the association in an independent replication sample (P = 2.0 × 10^-10). This finding suggests that multiple genes with small individual effects contribute to the genetic risk of multiple sclerosis.
Genome wide association study:

> 50 MS variants identified and replicated in 2011

IMSGC and WTCCC2
Nature 2011

---

Some Lessons from GWAS

- Most associations do not involve previous candidate genes. New biological hypotheses!
- Many involve non-protein coding regions
- Associations are modest
- Common variants across diseases
Illumina ImmunoChip Studies

184 Fine-mapping Regions
108,966
592 SNPs per region

Deep replication & ‘Wild Cards’
49,801
(1,444 MS SNPs)

MHC Region (21-34Mb coverage)
(7,786 SNPs)

• ~196,000 SNPS total
• Designed in 2009
• 11 autoimmune and Inflammatory diseases

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

International Multiple Sclerosis Genetics Consortium (IMSGC)

Compute immune, org, ane genetic group, we estimate tos (x,y =Mapping Regions and 2,499,999 MS SNPs per region, the association
region 108,966 SNPs per region, including 592 SNPs per region.

Deep replication & ‘Wild Cards’

49,801 SNPs

(1,444 MS SNPs)

MHC Region (21-34Mb coverage)

(7,786 SNPs)

• ~196,000 SNPS total
• Designed in 2009
• 11 autoimmune and Inflammatory diseases

A total of 14,498 cases and 24,091 healthy controls

11 countries

Studied for 161,311 autosomal variants (post QC)

IMSGC, IBDGC, WTCCC2

promise and pitfalls of the Immunochip

M. C. Calabresi and M. Farina

Please find the associated variants on arrays for non-diseased — we should find that the number of immune variants that are present on the Immunochip for the autoimmune diseases and inflammatory diseases is much higher than the number of variants that are present on the Immunochip for non-diseased subjects. This may be due to the different biological processes involved in immune dysregulation in autoimmune diseases and inflammatory diseases compared to non-diseased conditions.
What do we know so far about genetics of MS?

- **HLA-DRB1*15** is strongly associated with MS risk
- Other **DRB1/classical HLA/MHC region alleles** are associated with MS risk
- More than **110 non-MHC risk variants** with modest effects have been identified
- Together explain ~50% of genetic component
- **Environment, rare variants, or epigenetics** explain missing heritability?
- Discordance in MZ twins indicates environmental risk factors
Explaining ‘Missing Heritability’

• Some possibilities:
  1) Many more common, very modestly MS-associated variants remain to be identified
  2) Rare variant(s) for MS have been missed
  3) A common MS-associated SNPs tags a rare variant(s) with much bigger effects
  4) Genetic variant confers greater risk of MS in the presence of environmental exposure or visa versa (so called ‘GxE’)
  5) A common MS-associated SNP tags a heritable epigenetic variant(s) with much bigger effects

Environmental Exposures of Importance in MS

• Tobacco Smoke
• Obesity/BMI
• Vitamin D/Sunlight
• Epstein-Barr Virus
• Others
• GxE
Smoking and Risk of Multiple Sclerosis
Evidence of Modification by NAT1 Variants

Farren B. S. Briggs,* Brigid Acoma,* Ling Shen,* Patriceia Ramnally,* Hong Quach,* Allison Bernstein,* Kalliepe H. Bellinos,* Ingrid S. Kostum,* Emma K. Hedström,* Lars Alfredsson,* Tomas Olsson,* Catherine Schoenk,* and Lisa F. Barcellos**

Background: Tobacco smoke is an established risk factor for multiple sclerosis (MS). The heritability of this variation in genes coding for nicotine metabolism of tobacco smoke constitutes an early MHC interaction.

Methods: A time-series genome-wide association study was initiated for NAT1, NAT2, and CYP2D6 variants. The ancestry was confirmed among 5808 white MS cases and controls from Kaiser Permanente Northern California, San Francisco (US).

Results: The replication analysis was carried out in 686 white MS cases and controls from London.

Hypothesis: Tobacco smoking exposure at the age of 19 years was associated with greater MS risk in both data sets for NAT1 alleles T (OR = 1.13, 95% CI: 0.99-1.28) and T (OR = 1.12, 95% CI: 0.98-1.27). A series of NAT1 variants not studied in the replication data set was not associated with MS (OR = 1.16, 95% CI: 0.99-1.36). A meta-analysis of the two data sets was statistically significant for T allele OR = 1.12 (95% CI: 0.95-1.32), and was stronger in the replication data set (OR = 1.16, 95% CI: 0.99-1.36).

Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis

QB3 Genetic Epidemiology and Genomics Laboratory: AD Research Program

Three Main Areas:

- MHC genetics and maternal-child histocompatibility relationships
  UC San Francisco Department of Medicine/Rheumatology Division, Blood Systems Research Institute, San Francisco

- Genetic and environmental risk factors in MS/autoimmune disease
  - Northern and Southern California Kaiser Permanente
  - UC San Francisco Departments of Neurology, Medicine/Rheumatology Division
  - University of Oslo, Karolinska Institute, Copenhagen University Hospital
  - Johns Hopkins University

- Epigenetic studies in autoimmune disease
  - UC San Francisco Department of Medicine/Rheumatology Division, Neurology
  - University of Oslo, Children’s Hospital Oakland Research Institute

ABSTRACT

Objective: To investigate potential interactions between human leukocyte antigen (HLA) genotypes and body mass index (BMI) status in relation to the risk of developing multiple sclerosis (MS).

Methods: We used 2 case-control studies, one with incident cases (MS:130 cases, 2927 controls) and one with prevalent cases (MS:118 cases, 1073 controls). We compared BMI at study entry with a history of smoking and BMI at study entry with a history of smoking and BMI at study entry with a history of smoking and BMI at study entry with a history of smoking. Potential interactions between genotypes and BMI were evaluated by calculating the attributable proportion due to interaction.

Results: In both cohorts, a significant interaction was observed between BMI/HLA-DQB1*06 and obesity, regardless of BMI/HLA-DQB1*06 status. Similarly, there was a significant interaction between absence of BMI/obesity in the incident cohort, obesity in the prevalent cohort, and obesity in the prevalent cohort. The corresponding SP in the prevalent cohort was 0.035 (4.3-4.6).

Conclusions: We observed significant interactions between BMI/obesity and HLA genotypes with regard to the risk of developing MS. These results highlight the importance of considering BMI in genetic susceptibility to the disease. Neurology® 2014;82:960-972
**Epigenetics**: the study of reversible and heritable influences on gene activity that are **not** due to changes in DNA sequence

- Human genome is comprised of more than 3 billion base pairs
- Smaller number are within known coding sequences
- Differences among individuals, including those that influence health and disease, are likely to include epigenetic changes such as DNA methylation (DNAm)
- Methylation of the 5’ carbon of cytosine which occurs in the context of CpG dinucleotides, does not affect the genetic sequence; however, it plays a critical role in gene regulation

---

DNA methylation:

‘CpG sites’ - Cytosines (C) that are followed by guanine (G)

Hypermethylated $\xrightarrow{}$ Lower gene expression

\[
\text{ATCGTAGCGGCCGTCGACGCAACG}
\]

Hypomethylated $\xrightarrow{}$ Higher gene expression

\[
\text{ATCGTAGCGCGCGTCGACGCAACG}
\]

Detection of CpG methylation

1. Bisulfite treatment of denatured DNA
   - $m$C $\rightarrow$ $m$C
   - C $\rightarrow$ U

2. PCR amplification
   - $m$C $\rightarrow$ C
   - U $\rightarrow$ T

3. Analysis
   - Original sequence: GGA$^m$C/CGTAG$^m$C/CAGCTGCTA
   - After bisulfite: GGA$^m$C/UGUTAG$^m$C/UUGAUTGUTA
   - After PCR: GGAC/TGTTAGC/TGATTTGA
Illumina Infinium* 450K array (450K)

- 485,512 CpG sites across >20K genes
- Coverage ranges from 1-1,293 CpG sites/gene (mean 24 sites/gene)

<table>
<thead>
<tr>
<th>Gene Region</th>
<th>CpGs Per Area</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>8.1</td>
<td>93.6</td>
</tr>
<tr>
<td>5’UTR</td>
<td>4.5</td>
<td>82.2</td>
</tr>
<tr>
<td>1st Exon</td>
<td>2.4</td>
<td>82.4</td>
</tr>
<tr>
<td>Gene Body</td>
<td>9.5</td>
<td>96.7</td>
</tr>
<tr>
<td>3’UTR</td>
<td>1.5</td>
<td>79.7</td>
</tr>
<tr>
<td>CpG Islands</td>
<td>5.3</td>
<td>97.6</td>
</tr>
<tr>
<td>CpG Islands Shores</td>
<td>5.5</td>
<td>94.8</td>
</tr>
<tr>
<td>CpG Islands Shelves</td>
<td>4.0</td>
<td>89.6</td>
</tr>
</tbody>
</table>

*Illumina Infinium HumanMethylation450 BeadChip (450K)
Measure of DNA Methylation from 450K BeadChip

- $\beta$ is the ratio of the intensities of fluorescent signals from methylated (M) and unmethylated (U) alleles, where
  \[ \beta = \frac{\text{Max}(M, 0)}{\text{Max}(M, 0) + \text{Max}(U, 0) + 100} \]
- $\beta$ is a continuous measure with values from 0 (completely unmethylated site) to 1 (completely methylated site)
- Intensities from a large number of negative controls on the chips are used to adjust for background
- *DNA methylation reduces gene expression*
“Positives” for 450K BeadChip

- Content derived from all designable (96%) human RefSeq genes, including promoter, 5’ and 3’ regions, all CpG islands and island shores, as well as sites outside of CpG islands and CpG islands outside of coding regions
- Gene-centric
- Only 750 ng of genomic DNA needed
- High-throughput, cost-efficient and robust approach to characterizing global methylation for each individual sample/cell type and for genes of interest
- Validation studies; correlation with gene expression
- Methods well-established for normalization, pre/post processing

Reproducible!

Figure 1. Correlation plot for methylation profiles derived from Jurkat controls (between chip variation)

Figure 2. Correlation plot for DNA methylation profiles derived from replicate samples (within chip variation)

\( r^2 = 0.997; \) each control sample was run on a different Illumina 450K BeadChip

\( r^2 = 0.998; \) each replicate sample was run on the same Illumina 450K BeadChip; DNA samples from B cells
Evidence for 450K BeadChip validation with independent platform

Quantitative cross-validation and content analysis of the 450K DNA methylation array from Illumina, Inc.


Abstract

Background: The newly released 450K DNA methylation array from Illumina, Inc., offers the possibility to analyze more than 48,500 individual CpG sites in a user-friendly, standardized format. In this study, we analyze the relationship between the p-values provided by the Illumina, Inc. array for each individual CpG site and the quantitative methylation levels obtained by pyrosequencing were analyzed. In addition, the representation of known DNA methylation epigenetic loci was assessed in detail. Genome-wide DNA methylation levels were analyzed on 450K DNA methylation array, the p-values for the individual CpG sites from 12 different genes were cross-validated using conventional quantitative pyrosequencing.

Findings: The newly released 450K methylation array from Illumina, Inc., shows a high concordance with pyrosequencing data (48,500 CpG sites analyzed in cell lines, p-value < 0.0001), which is somewhat reduced in primary tumor specimens (p-value < 0.05). The CpG sites show an excellent correlation with pyrosequencing data in cell lines (r = 0.85 for MCF7 breast cancer cells and r = 0.87 for A549 lung cancer cells) and primary tumors (r = 0.83 for colon and r = 0.87 in prostate tumors). The number of CpG sites representing microRNA genes and 62 genes is very heterogeneous (r = 0.7 to 1,000 sites for microRNA and r = 260 for reported loci).

Conclusions: The newly developed 450K methylation array from Illumina, Inc., provides a genome-wide quantitative representation of DNA methylation alterations in a convenient format. Overall, the concordance with pyrosequencing data is very high. However, for individual loci, one should be careful to validate the p-values directly into percent methylation levels.

Figure 1 Cross-validation of 450K methylation array and quantitative pyrosequencing. Shown are scatter plots for the methylation values obtained by pyrosequencing (y-axis) and the p-values obtained from the 450K array for identical CpG sites. A) cell lines (Spearman r = 0.88, slope 0.89, 95% confidence interval 0.86 - 0.90), B) primary human tumor tissues (Spearman r = 0.88, linear regression r = 0.77, slope 0.81, 95% confidence interval 0.71 - 0.91). Each dot in the figure represents the comparison of 1% methylation according to pyrosequencing versus the p-value for a single CpG site in one sample. C and D respectively, the horizontal and dashed lines. For the construction of these plots, the p-values have been multiplied by the factor 100 in order to obtain data sets of the same size range. The coordinates of the individual CpG sites analyzed in this comparison are compiled in Additional file 1.
“Negatives” for 450K BeadChip

- SNPs in probe regions
- Missing information not captured by chip
- Assay does not have allelic level discrimination
Researchers utilized an “experiment of nature” in rural Gambia, where the population’s dependence on own-grown foods and a markedly seasonal climate impose a large difference in people's dietary patterns between rainy and dry seasons.

Season of conception affects DNA methylation in infants

Figure 3 | Season of conception affects DNA methylation in infants. (a) Percent methylation at the six MREs in PBL of infants conceived in the dry or rainy season. Median % methylation is consistently higher in infants conceived in the rainy season. (b) Mean PBL methylation across the six MREs is significantly higher in infants conceived in the rainy season. (c) Percent methylation at the six MREs in HF of infants conceived in the dry or rainy season; the overall pattern of methylation is similar to that observed in PBL, as is the seasonal difference in mean methylation. (d) Box plots represent the median (horizontal line) and interquartile range (box) of the indicated distribution. The whiskers extend from the top/bottoms of the box to the highest/lowest data value that is within 1.5. Asterisk represents interquartile range of the box. Data beyond the whiskers are plotted as individual points.

PBL, peripheral blood lymphocyte. One-way analysis of variance: *p < 0.05, **p < 0.01. PBL, N_dry = 126 and PBL, N_rainy = 87 infant DNA samples.

Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation

Sonja Zeilingera,b, Brigitte Kühne,c, Norman Kloppe,a,d, Hans-Jörg Bauarcb,e, Anja Kleinschmidt,a, Christian Gieger,f, Stephan Weidinger,f, Eva Lattka,f, Jery Adamski,f, Annette Peters,e,f, Konstantin Strauch,f,g, Melanie Waldenberger,a,c, Thomas Illig1,f,g

1Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, 2Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, 3Hannover Medical School, Hannover, Germany, 4Department of Genomics, Biometry, and Genealogy, University Hospital Schleswig-Holstein Campus Kiel, Kiel, Germany, 5Graduate School of Information Science in Health (GSIH), Technische Universität München, Munich, Germany, 6Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany, 7Chair of Experimental Genetics, Technische Universität München, Munich, Germany, 8Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, 9Institute of Medical Informatics, Biometry and Epidemiology, UMIT, University Innsbruck, Innsbruck, Austria, 10University of Heidelberg, Heidelberg, Germany

Abstract

Environmental factors such as cigarette smoking may have long-lasting effects on DNA methylation patterns, which might lead to changes in gene expression and in turn affect the development and progression of various diseases. We conducted an epigenome-wide association study (EWAS) comparing current, former and never smokers from 1793 participants of the population-based KORA F4 panel, with replication in 479 participants from the KORA F3 panel, carried out by the IOM BeadChip with genomic DNA obtained from whole blood. We observed widespread differences in the degree of interindividual methylation (with methylation ranging from 9.37% to 24.46% in each of the 22 autosomes), with the percent of variance explained by smoking ranging from 1.31 to 41.02. Depending on covariates and the effects of smoking, methylation levels in former smokers were found to be close to the ones seen in never smokers. In addition, methylation-specific gene expression patterns were observed for cg0575921 within AHR, which had the highest level of detective changes in DNA methylation associated with tobacco smoking (p = 2.546×10⁻¹⁸), suggesting a regulatory role for gene expression. The results of our study confirm the broad effect of smoking on the human organism, but also show that quitting tobacco smoking probably allows regaining the DNA methylation state of never smokers.

DNA methylation and complex diseases

DNA methylation studies of AD

- **Rheumatoid Arthritis**
  - 60 RA cases and 30 controls (all females)
  - Risk factors and several clinical outcomes
  - Comparison of peripheral blood (whole) and immune cells to synovial cells to identify biomarkers

- **Multiple Sclerosis**
  - 30 MS cases and 30 controls (all females)
  - Risk factors and clinical outcomes
  - Comparison of peripheral blood (whole) to immune cells to identify biomarkers

*using sorted immune cells from peripheral blood
DNA methylation studies of AD

- **Systemic Lupus Erythematosus**
  - 325 SLE cases
  - Autoantibody phenotypes
  - Submitted for publication

- **Sjogren’s Syndrome**
  - 20 SS cases and 20 controls
  - Using peripheral blood (whole), immune cells, and labial salivary gland tissue to identify biomarkers

- **Type 1 Diabetes**
  - Discordant MZ twin and sibpairs
  - Validation/replication in progress

*using sorted cells

**Samples collected from study subjects:**
Whole blood for DNA extraction without sorting
Five EDTA vacutainer tubes of whole blood for sorting/DNA, **FACS performed on whole blood.**
Comparison of two cell sorting protocols demonstrates feasibility of overnight blood storage

Table 2. Results from Pilot Study: FACS on Day 1 and Day 2*

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Total (%)</th>
<th>Gated (%)</th>
<th>Total (%)</th>
<th>Gated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocyte</td>
<td>90.7</td>
<td>97.9</td>
<td>92.2</td>
<td>98.5</td>
</tr>
<tr>
<td>B Cell</td>
<td>93.8</td>
<td>99.9</td>
<td>93.6</td>
<td>99.7</td>
</tr>
<tr>
<td>CD4+ Naive</td>
<td>90.1</td>
<td>97.0</td>
<td>93.7</td>
<td>98.6</td>
</tr>
<tr>
<td>CD4+ Memory</td>
<td>92.5</td>
<td>96.7</td>
<td>94.4</td>
<td>96.8</td>
</tr>
</tbody>
</table>

*All results were derived from whole blood collected from one donor. Stained cells were sorted by the DEM Flow Core Laboratory on a FACS Aria (BD, Bioscience) to collect the four cell populations simultaneously.

Figure 1. Site-specific DNA methylation profile differences (450K) for CD19+ B cells sorted 24 hours apart at T₁ and T₂.

B cells were derived from the same donor. Differences for methylation at each of 450K sites were determined and plotted using R; r² =0.987. Results show profiles were nearly identical. X-axis refers to CpGs in map order.

Figure 2. Site-specific DNA methylation profile difference (450K) for monocytes and CD4+ Memory T cells collected at same time.

CD19+ B cells and CD4+ T memory cells were derived from the same donor and processed at T₂. Differences for methylation at each of 450K sites were determined and plotted using R; r² =0.888. X-axis refers to CpGs in map order.
MS Study: Preliminary Results

- Whole-genome DNA methylation profiles derived from whole blood, CD4+ and CD8+ T cells in treatment naive female MS patients and controls
- Whole genome SNP profiles, RNAseq
- 30 (60) individuals, 90 (180) samples
- Tested several hypotheses
Table 1. Characteristics of MS cases and controls

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=16)</th>
<th>Controls (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, years (SD; range)</td>
<td>38.9 (8.9; 25-63)</td>
<td>39.2 (8.4; 28-58)</td>
</tr>
<tr>
<td>Mean age at onset, years (SD; range)</td>
<td>30.1 (7.5; 22-46)</td>
<td>-</td>
</tr>
<tr>
<td>Mean disease duration, years (SD; range)</td>
<td>8.8 (7.7; 1-33)</td>
<td>-</td>
</tr>
<tr>
<td>Oligoclonal bands in the CSF</td>
<td>14/16 (87.5%)</td>
<td>-</td>
</tr>
<tr>
<td>Mean EDSS</td>
<td>1.7 (1.3; 0-5)</td>
<td>-</td>
</tr>
<tr>
<td>Mean MSSS</td>
<td>2.4 (2.1; 0.17-7.61)</td>
<td>-</td>
</tr>
<tr>
<td>Total number of MRI lesions &gt; 20</td>
<td>10/16 (62.5%)</td>
<td>-</td>
</tr>
<tr>
<td>Contrast enhancing MRI lesions</td>
<td>4/16 (25.0%)</td>
<td>-</td>
</tr>
<tr>
<td>Current smoker</td>
<td>5/16 (31.3%)</td>
<td>4/14 (28.6%)</td>
</tr>
</tbody>
</table>

Abbreviations: SD= standard deviation, CSF= Cerebrospinal Fluid, EDSS= Expanded Disability Status Scale, MSSS= Multiple Sclerosis Severity Score.

*Bos SD, Page CM, ...Harbo HF, Barcellos LF, in prep*
Supplementary Figure 2. SNPs in methylation probes influence reported beta-values.

A. Example of a SNP located directly in CpG site, this results in clear distinction of reported beta-values depending on genotype for a Type I probe (two-color assay).
B. As in A, SNP located in the probe sequence but not in the CpG site itself.
C. Example of a SNP located directly in CpG site, this results in clear distinction of reported beta-values depending on genotype for a Type II probe (two-color assay).
D. As in C, with SNP located in the probe sequence but not in the CpG site itself.
Results: Case-Control Comparison

- Next Table. Top 40 results of the linear regression analyses for each cell type. The results shown are restricted to methylation differences of at least 5% (absolute beta difference).

- Data sorted by nominal P-values for: (A) CD4+ T cells methylation only, and (B) CD8+ T cell methylation only, and (C) Whole blood samples

<table>
<thead>
<tr>
<th>CD4+ T cells</th>
<th>CD8+ T cells</th>
<th>Whole Blood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>P-value</td>
<td>Beta</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4+ T cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>P-value</td>
<td>Beta</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Red Bold: Occurs in all three data sets
- Blue: Occurs in data sets with 40% or more
- Highlighted: Occurs in all data sets but not 40%
- Standard deviation for specified probe
- Effect size
- More methylation in MS
More CpG sites from CD8+ T cells in MS patients are hypermethylated compared to CD4+ T cells and whole blood

Direction of methylation for CpG sites with >5% differences in methylation (p<0.05)

Not shown: CpG sites with p>0.50. Hyper:Hypo ~50:50

Results: MS Candidate Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>CD4</th>
<th>CD4_BH_adj</th>
<th>CD8</th>
<th>CD8_BH_adj</th>
<th>WB</th>
<th>WB_BH_adj</th>
<th>CpG_probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPK3</td>
<td>0.005</td>
<td>0.005</td>
<td>0.216</td>
<td>0.216</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>TNFSF14</td>
<td>0.010</td>
<td>0.010</td>
<td>0.588</td>
<td>0.588</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>ZBTB46</td>
<td>0.050</td>
<td>0.048</td>
<td>0.379</td>
<td>0.379</td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>ETL2</td>
<td>0.064</td>
<td>0.056</td>
<td>0.311</td>
<td>0.311</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>MAF</td>
<td>0.069</td>
<td>0.060</td>
<td>0.372</td>
<td>0.372</td>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>TSFM</td>
<td>0.072</td>
<td>0.081</td>
<td>0.644</td>
<td>0.644</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>MMEL1</td>
<td>0.082</td>
<td>0.090</td>
<td>0.783</td>
<td>0.783</td>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>VDAC1</td>
<td>0.089</td>
<td>0.070</td>
<td>0.226</td>
<td>0.226</td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>CARD11</td>
<td>0.090</td>
<td>0.084</td>
<td>0.704</td>
<td>0.704</td>
<td></td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>IFI30</td>
<td>0.103</td>
<td>0.092</td>
<td>0.791</td>
<td>0.791</td>
<td></td>
<td></td>
<td>13</td>
</tr>
</tbody>
</table>

- Candidate gene based analyses based on the IMSGC reported SNPs providing the number of CpG probes within each of the genes (CpG_probes). Nominal p-values shown. Adjusted p-values are NS.
MS Study Results So Far:

- No consistent single CpG or gene large-scale DNA methylation differences in female MS cases as compared to controls for CD4+ T cells, CD8+ T cells or whole blood
- CD8+ T cells of MS patients showed a much higher level of DNA methylation for those sites with the lowest p-values in this study (hypermethylation)

MS Study Results So Far:

- No strong evidence for association between CpG sites within candidate MS genes and MS
- Whole genome data for study participants verified genetic ancestry and was used to removed CpG probes with SNPs
- Temporality
- Potential for biomarker?
In our study, 18 out of 19 of the reported MHC methylation sites in Graves et al. were compromised by the presence of at least one SNP in the probe sequence.

We could not replicate MHC or non-MHC findings.

Sample differences due to gender, treatment, other?
Analysis of MHC region in CD4+ T cells before and after QC

- Our results are in agreement with this study
- Large-scale differences in our much larger sample of homogenous MS patients and controls were not observed
- Baranzini, et al. also reported that MS discordant siblings displayed differences for CpG sites close to the TMEM1 or PEX14 genes
- These CpG sites are not covered by the 450K methylation chip; therefore, we could not verify these findings
Summary (1):

- Whole genome and targeted DNA methylation studies in MS and other complex diseases are feasible
- Choice of platform is important
- Choice of cell/tissue type is important
- Integration of genetic and gene expression data
- Study design/sample is important
- Some CpG sites are stable across cell types, others are not
- Methods well established for QC and normalization of Illumina data
Summary (2):

- Strong evidence for DNA methylation of candidate MS genes explaining ‘missing heritability’ not shown in our study
- Some new (non-candidates) to follow
- Differences that are meaningful are likely to be of smaller scale
- CD8+ T cell hypermethylation is of interest
- Working on new approaches to ‘correction’ so that whole blood can potentially be used
- Assessing the relationship between genetic and epigenetic variation (specifically 450K)
- Targeted DNA methylation sequencing
- miRNA and miRSNPs in pediatric and adult onset MS

Acknowledgements

University of Oslo: Christian Page, Hanne Harbo, Steffan Bos, Bettina Andreassen

UC Berkeley: Hong Quach, Diana Quach, Xiaorong Shao, Emon Elboudwarej, Farren Briggs

Supported by: NMSS: FG 1847A1/1,
NIH/NIEHS R01 NS0495103
Cell mixture correlation between paired subjects in T1D study

<table>
<thead>
<tr>
<th>Discordant MZ Twins</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP1</td>
<td>0.987</td>
</tr>
<tr>
<td>TP2</td>
<td>0.921</td>
</tr>
<tr>
<td>TP3</td>
<td>0.918</td>
</tr>
<tr>
<td>TP4</td>
<td>0.999</td>
</tr>
<tr>
<td>TP5</td>
<td>0.991</td>
</tr>
<tr>
<td>TP6</td>
<td>0.969</td>
</tr>
<tr>
<td>TP7</td>
<td>0.994</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discordant Sibs</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>0.836</td>
</tr>
<tr>
<td>SP2</td>
<td>0.923</td>
</tr>
<tr>
<td>SP3</td>
<td>0.650</td>
</tr>
<tr>
<td>SP4</td>
<td>0.965</td>
</tr>
<tr>
<td>SP5</td>
<td>0.998</td>
</tr>
</tbody>
</table>