Supplementary Online Content

Xia Z, Steele SU, Bakshi A, et al. Assessment of early evidence of multiple sclerosis in a prospective study of asymptomatic high-risk family members. *JAMA Neurol*. Published online January 17, 2017. doi:10.1001/jamaneurol.2016.5056

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Supplementary Methods

Cohort Data Collection

The Genes and Environment in Multiple Sclerosis (GEMS) Study did not include participants older than 50, as the mean age at MS diagnosis is in the early thirties and the age-related appearance of nonspecific white matter lesions in older individuals could confound neuroimaging outcomes. The GEMS study has been recruiting participants from across the United States. As of September 2016, the study has enrolled 2687 participants with representation from every state.

Upon enrollment, each GEMS participant completes a detailed web-based questionnaire that captures demographics, medical history, family history, and environmental exposures. Each participant mails a saliva sample for extraction of DNA (OG-500 DNA Genotek, Ontario, Canada). Participants are asked to complete additional questionnaires annually and donate biological samples (blood, stool) with the goal to follow each for 20 years[1]. The institutional review boards of Partners Health, National Institute of Health, and University of Pittsburgh approved the study.

Individualized Risk Estimate and Stratification

Genotyping of MS risk variants and calculation of the Genetic and Environmental Risk Score for Multiple Sclerosis Susceptibility ($GERS_{MS}$) in GEMS participants have been previously described[1]. In brief, a list of 64 single nucleotide polymorphisms (SNPs) that are significantly associated with the risk of developing MS based on published genome-wide association studies of MS susceptibility[2-4], including five within the major histocompatibility complex (MHC) region, was the most up-to-date as of 2012 when targeted genotyping of GEMS participants first began. Each SNP was coded additively by the established risk allele and weighted by the natural log of the published odds ratio for MS susceptibility. For the validated non-genetic risk factors, we counted the presence or absence of the factor, each weighted by the natural log of a published odds or risk ratio: sex (which confers different hormonal milieu between women and men)[5], infectious mononucleosis[6], current smoking status[7]. For each participant with completed genotype and questionnaire data, we calculated a weighted $GERS_{MS}$, a summary estimate of one's risk for MS that incorporates genetic burden and environmental exposure history. $GERS_{MS}$ was analyzed as both continuous and categorical variables.

Study Location and Subject Selection for Deep Phenotyping

Between August 2012 and July 2015, one hundred neurologically asymptomatic participants in the GEMS cohort of first-degree family members traveled from various locations across the United States to NIH in Bethesda, MD, for detailed neuroimaging, laboratory, and neurological examination.

Participants were selected for detailed study based on their risk profile for MS susceptibility. The original plan was to examine equal numbers of participants with $GERS_{MS}$ at the top 10% and the bottom 10% of the GEMS cohort. Due to the variable availability of the participants and the complexity of travel arrangements, the final ratio of the higher and lower risk participants was non-uniform, ultimately reaching a ratio of approximately 3:2 (**Supplementary Table 1**). Given that 96% of the participants in the overall GEMS study are of non-Hispanic European descent[1], we focused this phase of the

study exclusively on non-Hispanic European descent to mitigate confounding due to inter-individual variability resulting from race and ethnicity. Inclusion of small numbers of minorities would potentially diminish the statistical power and would not allow a meaningful subset analysis of the minority population. We excluded participants with an existing diagnosis of MS or other significant central nervous system (CNS) diseases (e.g., neoplasm, cerebrovascular disease, neurodegenerative disease, or known CNS inflammatory disease) or any contraindications to MRI.

Biological Sample Collection

During the NIH visit, participants donated blood and saliva samples that have been banked. Total 25-hydroxy Vitamin D (25-OH Vit D) level was measured using a chemiluminescent immunoassay (DiaSorin Liaison XL: Stillwater, MN). Additional assays using biological samples will be reported separately in the future.

Standard Neurological Examination

Board-certified neurologists and experienced nurse practitioners with specialty training in MS performed the standard neurological examination, including the 128 Hz tuning fork test. As part of the exam, we recorded the Expanded Disability Status Scale (EDSS) score[8].

Quantitative Neurological Evaluation

Trained research assistants obtained the following quantitative clinical measures: (1) MS Functional Composite (MSFC)[9], which includes the Timed 25-Foot Walk, 9-Hole Peg Test, and Paced Auditory Serial Addition Test (PASAT); (2) the written form of the Symbol-Digit Modalities Test (SDMT)[10,11]; (3) Timed Up-and-Go (TUG)[12,13], (4) vibration sensitivity measured using two additional modalities in addition to the 128 Hz tuning fork (see below).

We performed quantitative assessment of the vibration sensitivity at the proximal interphalangeal joint of the great toe on both sides using three modalities: (1) 128 Hz tuning fork, (2) Rydel-Seiffert graduated tuning fork[14,15], and (3) Vibratron-II device[16-19]. For the 128 Hz tuning fork test, administered as part of the standard neurological exam, the vibration disappearance threshold was documented as the total elapsed time until a participant ceased to experience vibration. The duration was reported as a binary variable of either ≤25 seconds or >25 seconds. For the Rydel-Seiffert test, the vibration disappearance threshold was reported as the intersection of two virtual triangles that move on a scale from 0 to 8. For the Vibratron-II device, vibration sensation thresholds (vibration units) were quantified as previously reported in studies of MS patients[17-19]. Briefly, frequency is held constant while the examiner controls the applied amplitude and the sequence of intensities for each rod. For each trial, vibration stimulation was present in one of the two rods, and the participant was asked to report which rod was vibrating.

The same Vibratron-II device was used throughout the course of this study for all GEMS participants as well as 14 healthy women volunteers from a different study. The device remained intentionally uncalibrated during the study period to avoid amplifying batch effect. In a linear regression model, the average vibration sensitivity threshold in women from the GEMS study, as detected by Vibratron-II, drifted lower during the study period (beta-coefficient=-0.011, p=0.0024). This trend is primarily due to the original

goal of having equal ratio of higher- and lower-risk participants and the complexity of making travel arrangement of the GEMS study participants, such that more higher-risk women completed the study visit during the first half of the study period whereas more lower-risk women completed the study during the second half of the study period. More important, the average vibration sensitivity measure in healthy women during the overlapping period was not influenced by test date (beta-coefficient=0.0068, p=0.64). This points to the absence of substantial intrinsic drift in Vibratron-II measurement over time.

Visual Function Testing

Trained research assistants administered the monocular and binocular letter acuity test using retro-illuminated high- and low-contrast Sloan letter charts (100% and 1.25% contrast levels at 2 meters, Precision Vision, LaSalle, IL). Participants were asked to wear their usual distance glasses or contact lenses if applicable. Participants read off the letters from three charts. For each chart, we recorded the number of correctly identified letters (out of maximum score of 60 per chart).

Optical Coherence Tomography (OCT)

We performed OCT in both eyes using the Spectralis device (Heidelberg Engineering, Inc., Heidelberg, Germany). Retinal nerve fiber layer (RNFL) measures were acquired as described elsewhere[20]. Briefly, these scan protocols provided the mean RNFL thickness for 360 degrees around the optic disc (12 deg diameter circle scan, 100-ART) and total macular volumes (20x20 deg, 49 cross-sectional or B-scans, 16-ART). All scans passed quality control. Qualitative assessment did not reveal gross macular pathology in the study participants.

Brain Magnetic Resonance Imaging (MRI)

Image Acquisition: Participants underwent brain MRI scans on a 3-tesla (T) whole-body MR system (Skyra; Siemens; Erlangen, Germany), equipped with a 32-channel phased array receiver head coil and gradient coils. The MRI protocol included the following sequences: 3D isotropic volumetric T1-weighted MPRAGE, 3D-T2-weighted fluid-attenuated inversion recovery or FLAIR (1 mm isotropic), and 3D-segmented echo planar imaging T2*-weighted (0.65 mm isotropic). During imaging, a single dose (0.1 mmol/kg) of gadobutrol (Gadavist) was injected using a power injector (Medrad, Warrendale, PA) at a speed of 1 ml/second to assess for the presence of enhancing MS lesions. Due to technical difficulties with the power injector and scan time constraints, one out of 100 participants did not receive gadobutrol.

Image Processing: The images collected directly from the MR imager underwent post-processing through an automated pipeline that was built in-house using the Nipype data processing framework[21]. To correct for motion between acquisitions, a rigid registration was performed between different image volumes with normalized mutual information as the cost function and windowed sinc interpolation. All scans were also rigidly aligned to the Montreal Neurological Institute (MNI152) structural atlas[22] to provide a rough spatial alignment between participants and hence a regularized visual appearance. A FLAIR* volume[23] for perivenous lesion detection was calculated by the voxel-wise multiplication of the co-registered T2*-weighted and FLAIR volumes.

Image Analysis: For *lesion analysis*, the primary outcome was the presence or absence of hyperintense lesions on T2-weighted FLAIR sequences (for the upper cervical spinal cord, T1-weighted MPRAGE sequences). Two experienced reviewers (D.S.R., an attending neuroradiologist with 11 years of experience in MS imaging, and M.K.S., a neurologist with 5 years of experience in neurology and 2 years of subspecialty fellowship training in MS) evaluated all scans for the presence, appearance, and location of lesions, as well as the presence of absence of leptomeningeal enhancement[24], using standard radiology viewing software (Carestream Vue, Rochester, NY). In addition, lesions were analyzed for perivenous appearance [25,26]. A finding on brain MRI was defined as a "lesion" if it met the following three criteria: (1) ovoid, well-circumscribed, and homogeneous focus with or without involvement of the corpus callosum, (2) a T2 hyperintensity measuring $\geq$3mm, and (3) not consistent with a vascular pattern [27]. The two reviewers reviewed all scans in a blinded fashion without prior access to a participant's risk profile.

For *volumetric analysis*, the Lesion-TOADS (Topology-Preserving Anatomical Segmentation) method was deployed as previously described[28]. Briefly, this technique segments the brain into its component substructures while simultaneously delineating MS lesions, yielding volumes of the following regions of interest: cortical gray matter, cerebral white matter, cerebrospinal fluid, and thalamus. Images underwent nonparametric, non-uniform intensity normalization inhomogeneity correction and skull stripping before segmentation[29]. Raw volumes were normalized to intracranial volume. Brain parenchymal fraction (BPF) was calculated as total brain volume divided by intracranial volume. Spinal cord cross-sectional area, from the C1-C2 intervertebral disc level to the C5 vertebral body level, was derived from the 3D T1-weighted MPRAGE sequences using a semi-automatic algorithm based on edge detection[30] and a mean upper cervical spinal cord cross-sectional area was calculated.

Additional Statistical Analysis

In an omnibus test (**Fig. 1B**), we tested the hypothesis that the chi-squared statistic as calculated from a Fisher's combined probability test (combining the p-values of phenotypes) was higher in the observed data than expect by chance. To generate an empirical distribution of chi-squared statistics, we performed 10,000 permutations. For each permutation, we randomly assigned data to $GERS_{MS}$ risk category, maintaining the number of cases per category. We then evaluated each phenotype for association with risk group, and finally combined the resulting p-values using Fisher's combined probability test. Because we permuted the risk group and not the phenotype, the correlation structure of the phenotypes was maintained within each permutation. The omnibus p-value was the number of times we saw a chi-squared statistic in the empirical distribution that was greater than the observed chi-squared statistic, divided by the number of permutations (10,000). A quantile-quantile (qq)-plot was generated by plotting the observed p-value (-log10 [p-value]) for all measures against the expected p-value (-log10 [p-value]) with the assumption that there was no difference between higher and lower risk participants (i.e., null hypothesis). To create confidence intervals for the qq-plot, we used the extreme ranges (90% and 95%) of the randomly generated p-values.

To create a composite categorical trait in which there was an abnormal finding on *at least* one of the measured outcomes (brain MRI, OCT, or clinical), we included all dichotomous outcomes and continuous outcomes. For dichotomous outcomes, an abnormal finding was defined as meeting the criterion for abnormality. For continuous outcomes, the threshold for abnormality was set at one, two, or three standard deviations away from the mean of the study population toward the expected direction of abnormality. A participant was categorized as meeting the composite outcome if any one of the measured outcomes was abnormal.

eResults. Supplementary Results

We assessed the proportion of participants having an abnormal finding on *at least* one of the measured outcomes (brain MRI, OCT, or neurological exam). To identify outliers, we set the threshold at three standard deviations away from the mean in the direction of abnormal finding for a continuous trait or meeting the abnormal criterion for a dichotomous trait. Among the women who completed the entire battery of tests (including vibration sensitivity testing), we did not observe a statistically significant difference between higher and lower risk women (**eTable 3**), possibly due to the large variance for this composite categorical trait and the modest sample size. However, there was a suggestion of a greater proportion of higher-risk women to be outliers. Interestingly, 30% of the women who underwent the entire battery of testing (including vibration sensitivity) exhibited at least one abnormal finding as defined above (**eTable 3**).

Comparisons between the higher and lower risk group for the primary neuroimaging outcome (proportion of participants meeting the 2010 McDonald criteria for dissemination in space, **Table 1**) and the vibration sensitivity measure (the average of the measurement in the right and left great toe, **Table 2**) for the entire study cohort, without restriction to women, are shown in **eTable 5**. We note that these comparisons should not be construed as statistically meaningful due to the extremely skewed distribution of sex between the higher and lower risk groups

eTable 1. Overall Characteristics of the Neurologically Asymptomatic Participants From the Genes and Environment in Multiple Sclerosis Cohort of First-Degree Family Members Who Underwent Deep Phenotyping

Part A. All Participants

Parameter	**All Participants**	**Lower Risk Subgroup**	**Higher Risk Subgroup**	***p*-value ***
Number of Participants	100	59	41	N/A
Female, N (%)[♯]	65 (65%)	25 (42%)	40 (98%)	**1.5×10^{-9}**
Age [A], years Mean (SD)	35.1 (8.7)	34.8 (8.8)	35.8 (8.8)	0.58
Mono [B], N (%)[♯]	16 (16%)	7 (12%)	9 (22%)	0.27
Current Smoker [A, C], N (%)[♯]	4 (4%)	2 (3.3%)	2 (4.9%)	>0.99
Migraine [D], N (%)[♯]	30 (30%)	18 (31%)	12 (29%)	>0.99
BMI [A, E], kg/m^2, Mean (SD)	26.3 (5.4)	25.5 (4.3)	27.5 (6.5)	0.067
Height [A], meter, Mean (SD)	1.73 (0.11)	1.76 (0.12)	1.69 (0.07)	**0.001**
25-OH Vit D [A,F], ng/ml Mean (SD)	32.6 (12.5)	31.9 (11.6)	33.6 (13.9)	0.51
$GERS_{MS}$ [G] Mean (SD)	10.9 (2.1)	9.2 (0.5)	13.3 (0.6)	N/A

Part B. All Women

Parameter	**All Women**	**Lower Risk Subgroup**	**Higher Risk Subgroup**	***p*-value ***
Number of Participants	65	25	40	N/A
Age [A], years, Mean (SD)	35.3 (8.4)	34.6 (7.8)	35.7 (8.8)	0.61
Mono [B], N (%)[♯]	12 (18%)	3 (12%)	9 (23%)	0.34
Current Smoker [C], N (%)[♯]	3 (5%)	1 (4%)	2 (5%)	>0.99
Migraine [D], N (%)[♯]	24 (37%)	13 (52%)	11 (28%)	0.065
BMI [A, E],	26.3 (5.7)	24.5 (3.3)	27.4 (6.6)	**0.046**

kg/m^2, Mean (SD)				
Height [A], meter, Mean (SD)	1.67 (0.07)	1.65 (0.07)	1.68 (0.06)	0.052
25-OH Vit D [A,F], ng/ml Mean (SD)	33.5 (12.2)	33.3 (9.9)	33.6 (13.7)	0.92
GERS$_{MS}$ [G] Mean (SD)	11.8 (1.9)	9.5 (0.2)	13.2 (0.8)	N/A

Women Who Underwent Vibration Sensitivity Testing				
Parameter	**Women**	**Lower Risk Subgroup**	**Higher Risk Subgroup**	***p*-value ***
Number of Participants	47	20	27	N/A
Age [A], years, Mean (SD)	35.8 (8.2)	34.7 (7.5)	36.6 (8.8)	0.43
Mono [B], N (%)[♯]	14 (30%)	2 (10%)	5 (19%)	0.45
Current Smoker [C], N (%)[♯]	2 (4%)	0 (0%)	2 (7.4%)	0.50
Migraine [D], N (%)[♯]	20 (43%)	11 (55%)	9 (33%)	0.23
BMI [A, E], kg/m^2, Mean (SD)	26.4 (5.8)	24.2 (2.5)	28.0 (7.0)	**0.015**
Height [A], meter, Mean (SD)	1.67 (0.06)	1.65 (0.07)	1.69 (0.06)	0.052
25-OH Vit D [A,F], ng/ml Mean (SD)	33.2 (11.5)	34.2 (10.2)	32.3 (12.7)	0.60
GERS$_{MS}$ [G] Mean (SD)	11.6 (1.9)	9.5 (0.3)	13.2 (0.7)	N/A

Note:

A. At the time of neuroimaging and clinical evaluation
B. History of infectious mononucleosis
C. Current cigarette smoker
D. History of migraine
E. Body mass index
F. 25-hydroxy Vitamin D level
G. Genetic and environmental risk score (see reference xx)

* Comparing the lower-risk and higher-risk subgroup
♯ Percentage of the number of participants in the column
N/A Not applicable

eTable 2. Additional Quantitative Neurological Evaluation of the Neurologically Asymptomatic Women From the Genes and Environment in Multiple Sclerosis Cohort of First-Degree Family Members

Parameter	**All Female Participants**	**Lower Risk Subgroup**	**Higher Risk Subgroup**	***p*-value ***
Number of Participants	65	25	40	N/A
Visual Acuity, High-Contrast, Right [A], Number correct, Mean (SD)	56.7 (7.7)	57.5 (4.6)	56.7 (7.7)	0.61
Visual Acuity, High-Contrast, Left [A], Number correct, Mean (SD)	55.8 (6.0)	56.4 (4.8)	55.8 (6.0)	0.63
Visual Acuity, High-Contrast, Binocular [A], Number correct, Mean (SD)	60.8 (5.0)	60.0 (3.8)	60.8 (5.0)	0.47
Visual Acuity, Low-Contrast, Right [B], Number correct, Mean (SD)	19.5 (10.0)	19.8 (9.9)	19.5 (10.0)	0.89
Visual Acuity, Low-Contrast, Left [B], Number correct, Mean (SD)	18.5 (9.6)	19.3 (10.4)	18.5 (9.6)	0.76
Visual Acuity, Low-Contrast, Binocular [B], Number correct, Mean (SD)	31.1 (7.0)	31.4 (6.7)	31.1 (7.0)	0.85
Rydel, RLE [C], Median (IQR)	7 (2)	7 (1.75)	7 (2)	0.79
Rydel, LLE [C], Median (IQR)	7 (2)	7 (1.75)	7 (2)	0.56
128 Hz Duration, RLE [D], N (%)[♯] with ≤25 seconds	16 (34%)	4 (20%)	12 (44%)	0.12

128 Hz Duration, LLE [D], N (%)[♯] with ≤25 seconds	16 (34%)	4 (20%)	12 (44%)	0.12

Note:

A. High contrast (100%) visual acuity: number of correct answers
B. Low contrast (1.5%) visual acuity: number of correct answers
C. Rydel-Seiffert graduated tuning fork testing of vibratory sensation in the great toe: *RLE*, right great toe; *LLE*, left great toe; measured between 0-8 with a higher number indicating better vibratory sensation
D. Measure of 128 Hz tuning fork duration in the great toe: *Average*, average of the measurement in the right great toe and left great toe; measured as the proportion of participants with duration ≤25 seconds or >25 seconds (with a higher duration indicating better vibration sensation)

* Comparing the lower-risk and higher-risk subgroup

N/A Not applicable

eTable 3. The Burden of Subclinical Neurological Dysfunction in Asymptomatic Women From the Genes and Environment in Multiple Sclerosis Cohort of First-Degree Family Members Who Underwent Deep Phenotyping

Threshold for Abnormality [♯]	**All Women Participants (N=65)**	**Lower Risk Subgroup (N=25)**	**Higher Risk Subgroup (N=40)**	***p*-value ***
	N (%[♯♯])	N (%[♯♯])	N (%[♯♯])	
Any or 1 SD [A]	56 (86%)	23 (92%)	33 (83%)	0.46
Any or 2 SD [B]	34 (52%)	9 (36%)	25 (63%)	0.045
Any or 3 SD [C]	25 (38%)	6 (24%)	19 (48%)	0.071
	Women Who Underwent Vibratory Testing ** (N=47)	**Lower Risk Subgroup (N=20)**	**Higher Risk Subgroup (N=27)**	***p*-value ***
	N (%[♯♯])	N (%[♯♯])	N (%[♯♯])	N (%[♯♯])
Any or 1 SD [A]	38 (81%)	18 (90%)	20 (74%)	0.27
Any or 2 SD [B]	22 (47%)	7 (35%)	15 (56%)	0.23
Any or 3 SD [C]	14 (30%)	4 (20%)	10 (37%)	0.33

Note:

* Comparing the lower-risk and higher-risk subgroup

** Entire battery of testing includes all three measures of vibration sensitivity (Vibratron-II, Rydel-Seiffert graduated tuning fork, and measure of 128 Hz tuning fork duration in the great toe)

♯ A participant was defined as having any anomaly if the participant had *at least* one abnormal finding on any of the measured phenotypes brain magnetic resonance imaging scan (Table 2), optical coherence tomography (Table 2), or quantitative neurological exam (Table 2), including vibratron measure (Table 3). For a given dichotomous outcome, an abnormal finding was the presence thereof. For a given continuous outcome, an abnormal finding was defined as having greater than one, two, or three standard deviations away from the mean of the study population in the expected direction of the abnormality. We did not observe any statistically significant difference between higher and lower risk women, possibly due to the large variance for this composite categorical trait and the modest sample size.

♯♯ Percentage of the number of participants in the column

A. Meeting the criterion for abnormality for a dichotomous trait *or* having <u>one</u> standard deviation away from the mean in the expected direction of the abnormality for a continuous trait.

B. Meeting the criterion for abnormality for a dichotomous trait *or* having two standard deviations away from the mean in the expected direction of the abnormality for a continuous trait.
C. Meeting the criterion for abnormality for a dichotomous trait *or* having three standard deviations away from the mean in the expected direction of the abnormality for a continuous trait.

eTable 4. Power Calculation Based on Findings From the Current Study

Parameter	**Current Power**	**Number of Female Participants Per Risk Group for 80% Power**
McDonald 2010 [A]	12%	283
McDonald 2010 [A] + ≥40% T2 lesions Perivenous [B]	25%	97
Vibratron-II, Average [C]	99%	N/A

Note:

A. Frequency of meeting the 2010 McDonald criteria for dissemination in space [31]
B. Frequency of having >=40% of T2-hyperintense lesions that are exhibit perivenous appearance [25,26]
C. Vibratron measurement of vibratory sensation in the great toe in vibration unit (vu), which is the amplitude of the vibration (proportional to the square of applied voltage) and a lower unit of measurement indicates better vibratory sensation: *Average*, average of the measurement in the right great toe and left great toe

eTable 5. Comparison of the Primary Neuroimaging Outcome and Vibration Sensitivity Measure in All Participants

Parameter	**Total**	**Lower Risk Subgroup**	**Higher Risk Subgroup**	***p*-value** *****	***p*-value** ****** Adjusted
Neuroimaging Primary Outcome					
Number of Participants	100	59	41	N/A	N/A
Female, N (%)[♯]	65 (65%)	25 (42%)	40 (98%)	**1.5×10^{-9}**	N/A
McDonald 2010 [A], N (%)[♯]	9 (9%)	5 (8%)	4 (10%)	>0.99	N/A
Vibration Sensitivity Primary Outcome					
Number of Participants	73	45	28	N/A	N/A
Female, N (%)[♯]	47 (64%)	20 (44%)	27 (96%)	**1.1×10^{-4}**	N/A
Vibratron-II, Average [B], vu, Mean (SD)	2.24 (0.97)	2.18 (1.16)	2.35 (0.56)	0.47	0.14

Note:

* Comparing the lower-risk and higher-risk subgroup

** Comparing the lower-risk and higher-risk subgroup, after adjusting for age (at the time of evaluation), height, and testing date. Current smoking status was included in the Genetic and Environmental Risk Score ($GERS_{MS}$), which categorized a participant as higher versus lower risk.

♯ Percentage of the number of participants in the relevant section of the column

N/A Not applicable

A. Frequency of meeting the 2010 McDonald criteria for dissemination in space [31]
B. Vibratron-II measurement of vibration sensitivity in the great toe in vibration unit (vu), which is the amplitude of the vibration (proportional to the square of applied voltage) and a lower unit of measurement indicates better vibratory sensation: *Average*, average of the measurement in the right great toe and left great toe

eFigure. Correlation Architecture of the Measured Outcomes in Neurologically Asymptomatic Women Participants From the Genes and Environment in Multiple Sclerosis Cohort of First-Degree Family Members Who Underwent Deep Phenotyping

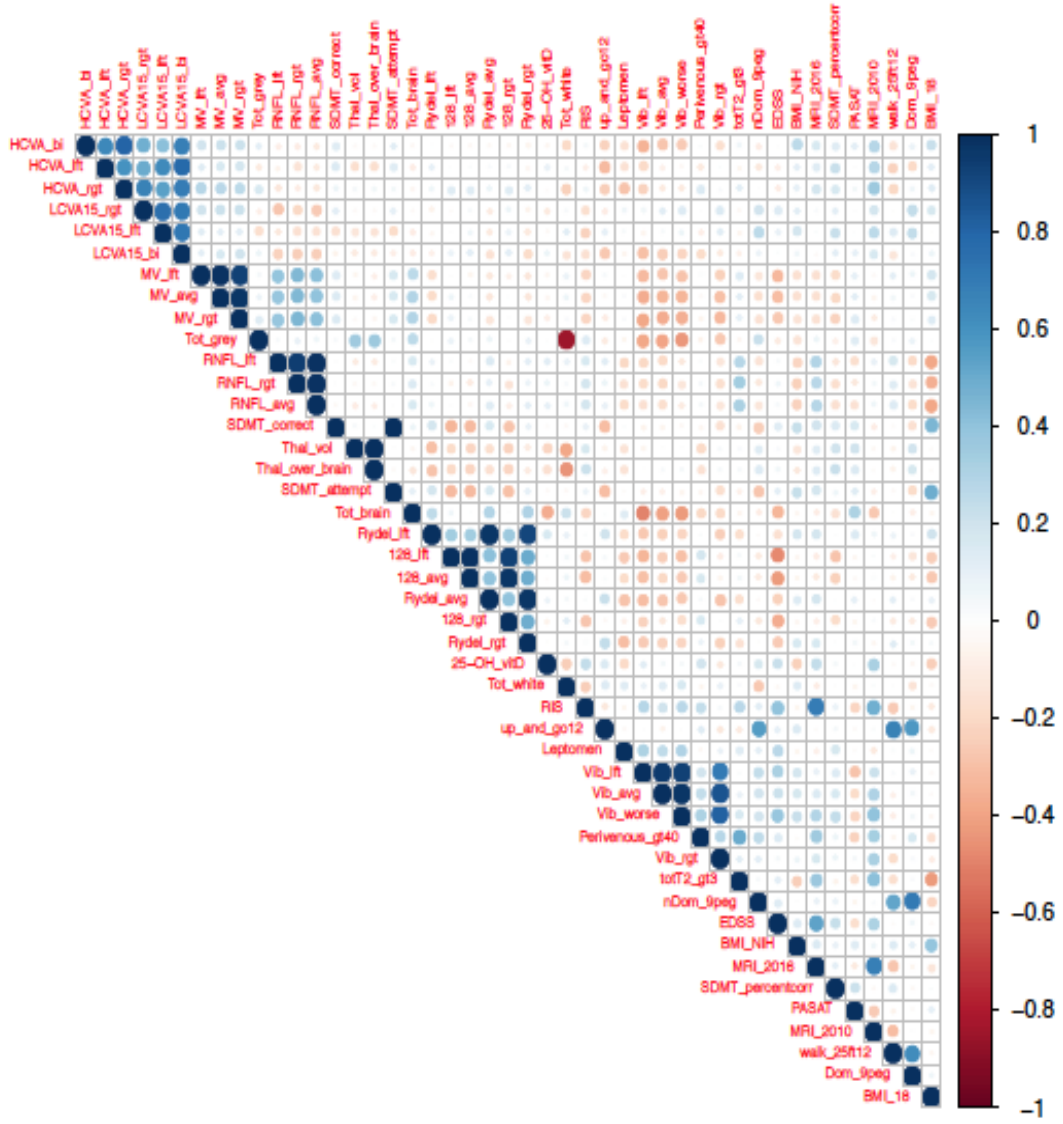
Variable Names	Description
128_avg	average 128-Hz duration between left and right great toe (seconds)
128_lft	128-Hz duration in left great toe (seconds)
128_rgt	128-Hz duration in right great toe (seconds)
25-OH_vitD	25-hydroxy vitamin D level (ng/ml)
bmi_18	body mass index (kg/m2) at age 18 years
bmi_nih	body mass index at the time of study visit to the National Institute of Health
Dom_9peg	9-hole peg test with dominant hand (average duration of two trials, seconds)
EDSS	Expanded Disability Severity Score
HCVA_bi	high-contrast (100%) visual acuity binocular
HCVA_lft	high-contrast (100%) visual acuity left
HCVA_rgt	high-contrast (100%) visual acuity right
LCVA15_bi	low-contrast (1.5%) visual acuity binocular
LCVA15_lft	low-contrast (1.5%) visual acuity left
LCVA15_rgt	low-contrast (1.5%) visual acuity right
Leptomen	having evidence of focal leptomeningeal enhancement on brain MRI
MRI_2010	meeting the 2010 McDonald criteria for dissemination in space
MRI_2016	meeting the 2016 proposed criteria for dissemination in space
MV_avg	macular volume, average between left and right side (m^3)
MV_lft	macular volume left side (m^3)
MV_rgt	macular volume right side (m^3)
nDom_9peg	9-hole peg test with non-dominant hand (average duration of two trials, seconds)
PASAT	Paced Auditory Serial Addition Test (number of correct response)
Perivenous_gt40	having >=40% of the T2-hyperintense lesions that exhibit perivenous appearance on brain MRI
RIS	meeting Okuda's criteria for Radiologically Isolated Syndrome
RNFL_avg	retinal nerve fiber layer thickness, average between left and right side (m)
RNFL_lft	retinal nerve fiber layer thickness left side (m)
RNFL_rgt	retinal nerve fiber layer thickness right side (m)
Rydel_avg	rydel test, the average bewteen left and right great toe
Rydel_lft	rydel test in the left great toe
Rydel_rgt	rydel test in the right great toe
SDMT_attempt	Symbol Digit Modalities Test (number attempted)
SDMT_correct	Symbol Digit Modalities Test (number correct)

SDMT_percentcorr	Symbol Digit Modalities Test (percentage of correct response)
Thal_over_brain	normalized thalamus volume divided by normalized total brain volume
Thal_vol	thalamus volume normalized by intracranial volume
Tot_brain	total brain volume normalized by intracranial volume
Tot_grey	total grey volume normalized by intracranial volume
Tot_white	total white matter volume normalized by intracranial volume
totT2_gt3	total number of T2-hyperintense lesions >=3mm on brain MRI
Up_and_go12	time of up-and-go (average duration of two trials, seconds)
Vib_avg	vibratron measure, average from left and right great toe
Vib_lft	vibratron measure in the left great toe
Vib_rgt	vibratron measure in the right great toe
Vib_worse	vibratron measure from the worse side between left and right great toe
Walk_25ft12	time of 25-foot walk (average duration of two trials, seconds)

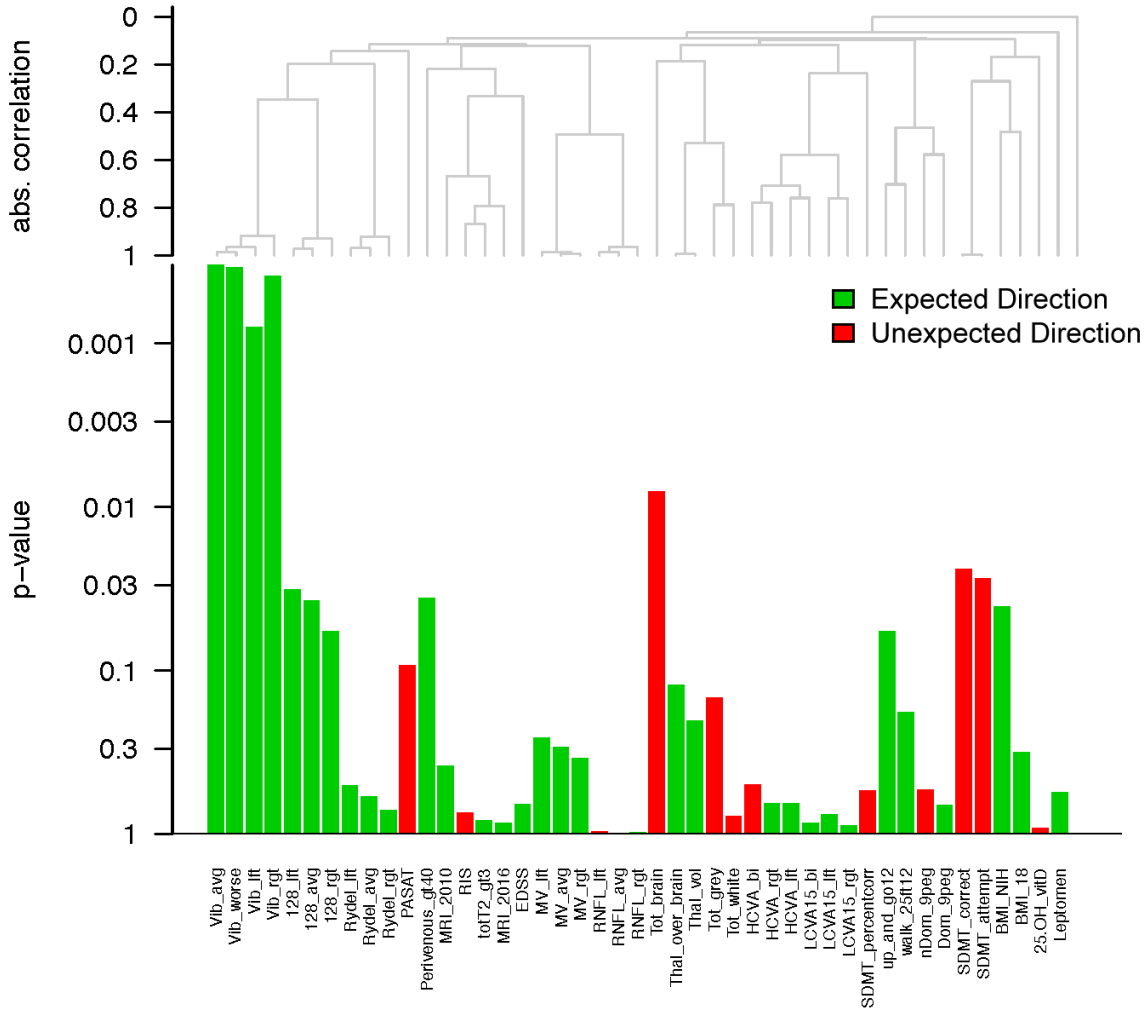
Abbreviations:

(**A**) Spearman's rank correlation coefficients (rho) between a given pair of phenotypes are shown on a color scale from -1 to 1. The size of each bubble indicates the significance level of the correlation. (**B**) The measured outcomes are ordered in a hierarchical clustering graph using average linkage. The distance measure used for the graph is the correlation distance. For a given outcome, the expected direction (green) indicates that a greater proportion of higher risk participants exhibited abnormality than the lower risk participants, whereas unexpected direction (red) indicates the opposite.

eFigure, A



eFigure, B



eReferences

1. Xia Z, White CC, Owen EK, et al. Genes and Environment in Multiple Sclerosis project: A platform to investigate multiple sclerosis risk. *Annals of Neurology*. 2016;79(2):178-189. doi:10.1002/ana.24560.

2. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011;476(7359):214-219. doi:10.1038/nature10251.

3. Patsopoulos NA, Barcellos LF, Hintzen RQ, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet*. 2013;9(11):e1003926. doi:10.1371/journal.pgen.1003926.

4. Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of Neurology*. 2011;70(6):897-912. doi:10.1002/ana.22609.

5. Wallin MT, Culpepper WJ, Coffman P, et al. The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service. *Brain*. 2012;135(6):1778-1785. doi:10.1093/brain/aws099.

6. Thacker EL, Mirzaei F, Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. *Annals of Neurology*. 2006;59(3):499-503. doi:10.1002/ana.20820.

7. Hernán MA, Jick SS, Logroscino G, Olek MJ, Ascherio A, Jick H. Cigarette smoking and the progression of multiple sclerosis. *Brain*. 2005;128(Pt 6):1461-1465. doi:10.1093/brain/awh471.

8. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444-1452.

9. Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. *Multiple Sclerosis*. 1999;5(4):244-250.

10. Drake A, Weinstock-Guttman B, Morrow S, Hojnacki D, Munschauer F, Benedict R. Psychometrics and normative data for the Multiple Sclerosis Functional Composite: replacing the PASAT with the Symbol Digit Modalities Test. *Multiple Sclerosis*. 2010;16(2):228-237. doi:10.1177/1352458509354552.

11. Van Schependom J, D'hooghe MB, Cleynhens K, et al. The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis. *Eur*

J Neurol. 2014;21(9):1219–25–e71–2. doi:10.1111/ene.12463.

12. Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of The Timed Up and Go as A Measure of Functional Mobility in Persons with Multiple Sclerosis. *Archives of Physical Medicine and Rehabilitation*. March 2016. doi:10.1016/j.apmr.2015.12.031.

13. Bethoux F, Bennett S. Evaluating walking in patients with multiple sclerosis: which assessment tools are useful in clinical practice? *Int J MS Care*. 2011;13(1):4-14. doi:10.7224/1537-2073-13.1.4.

14. Rydel A, SeiVer W. Untersuchungen über das vibrationsge- fühl oder die sog. "Knochensensibilität" (Pallästhesie). *Archiv fur Psychiatrie und Nervenkrankenheiten*. 1903;37:488-536.

15. Martina IS, van Koningsveld R, Schmitz PI, van der Meché FG, van Doorn PA. Measuring vibration threshold with a graduated tuning fork in normal aging and in patients with polyneuropathy. European Inflammatory Neuropathy Cause and Treatment (INCAT) group. *Journal of Neurology, Neurosurgery & Psychiatry*. 1998;65(5):743-747. doi:10.1136/jnnp.65.5.743.

16. Arezzo JC. *Quantitative Sensory Testing of Vibration Threshold: Vibratron II (Rationale and Methods)*. Clifton, NJ: Physitemp Instruments, Inc; 1993.

17. Newsome SD, Wang JI, Kang JY, Calabresi PA, Zackowski KM. Quantitative measures detect sensory and motor impairments in multiple sclerosis. *Journal of the Neurological Sciences*. 2011;305(1-2):103-111. doi:10.1016/j.jns.2011.03.003.

18. Fritz NE, Newsome SD, Eloyan A, Marasigan RER, Calabresi PA, Zackowski KM. Longitudinal relationships among posturography and gait measures in multiple sclerosis. *Neurology*. 2015;84(20):2048-2056. doi:10.1212/WNL.0000000000001580.

19. Fritz NE, Marasigan RER, Calabresi PA, Newsome SD, Zackowski KM. The impact of dynamic balance measures on walking performance in multiple sclerosis. *Neurorehabil Neural Repair*. 2015;29(1):62-69. doi:10.1177/1545968314532835.

20. Saidha S, Syc SB, Ibrahim MA, et al. Primary retinal pathology in multiple sclerosis as detected by optical coherence tomography. *Brain*. 2011;134(Pt 2):518-533. doi:10.1093/brain/awq346.

21. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*. 2011;5:13. doi:10.3389/fninf.2011.00013.

22. Fonov V, Evans AC, Botteron K, et al. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. 2011;54(1):313-327.

doi:10.1016/j.neuroimage.2010.07.033.

23. Sati P, George IC, Shea CD, Gaitán MI, Reich DS. FLAIR*: a combined MR contrast technique for visualizing white matter lesions and parenchymal veins. *Radiology*. 2012;265(3):926-932. doi:10.1148/radiol.12120208.

24. Absinta M, Vuolo L, Rao A, et al. Gadolinium-based MRI characterization of leptomeningeal inflammation in multiple sclerosis. *Neurology*. 2015;85(1):18-28. doi:10.1212/WNL.0000000000001587.

25. Solomon AJ, Schindler MK, Howard DB, et al. "Central vessel sign" on 3T FLAIR* MRI for the differentiation of multiple sclerosis from migraine. *Ann Clin Transl Neurol*. 2016;3(2):82-87. doi:10.1002/acn3.273.

26. Tallantyre EC, Dixon JE, Donaldson I, et al. Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. *Neurology*. 2011;76(6):534-539. doi:10.1212/WNL.0b013e31820b7630.

27. Okuda DT, Mowry EM, Beheshtian A, et al. Incidental MRI anomalies suggestive of multiple sclerosis: the radiologically isolated syndrome. *Neurology*. 2009;72(9):800-805. doi:10.1212/01.wnl.0000335764.14513.1a.

28. Shiee N, Bazin P-L, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*. 2010;49(2):1524-1535. doi:10.1016/j.neuroimage.2009.09.005.

29. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908.

30. Liu W, Nair G, Vuolo L, et al. In vivo imaging of spinal cord atrophy in neuroinflammatory diseases. *Annals of Neurology*. 2014;76(3):370-378. doi:10.1002/ana.24213.

31. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*. 2011;69(2):292-302. doi:10.1002/ana.22366.