OPEN

Genes From a Translational Analysis Support a Multifactorial Nature of White Matter Hyperintensities

Lorna M. Lopez, PhD; W. David Hill, MSc; Sarah E. Harris, PhD; Maria Valdes Hernandez, PhD; Susana Munoz Maniega, PhD; Mark E. Bastin, PhD; Emma Bailey, PhD; Colin Smith, MD; Martin McBride, PhD; John McClure, PhD; Delyth Graham, PhD; Anna Dominiczak, MD; Qiong Yang, PhD; Myriam Fornage, PhD; M. Arfan Ikram, MD, PhD; Stephanie Debette, MD, PhD; Lenore Launer, PhD; Joshua C. Bis, PhD; Reinhold Schmidt, MD; Sudha Seshadri, MD; David J. Porteous, PhD; John Starr, MD; Ian J. Deary, PhD; Joanna M. Wardlaw, MD

Background and Purpose—White matter hyperintensities (WMH) of presumed vascular origin increase the risk of stroke and dementia. Despite strong WMH heritability, few gene associations have been identified. Relevant experimental models may be informative.

Methods—We tested the associations between genes that were differentially expressed in brains of young spontaneously hypertensive stroke–prone rats and human WMH (using volume and visual score) in 621 subjects from the Lothian Birth Cohort 1936 (LBC1936). We then attempted replication in 9361 subjects from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE). We also tested the subjects from LBC1936 for previous genome-wide WMH associations found in subjects from CHARGE.

Results—Of 126 spontaneously hypertensive stroke–prone rat genes, 10 were nominally associated with WMH volume or score in subjects from LBC1936, of which 5 (*AFP*, *ALB*, *GNAI1*, *RBM8a*, and *MRPL18*) were associated with both WMH volume and score ($P<0.05$); 2 of the 10 (*XPNPEP1*, $P=6.7\times10^{-5}$; *FARP1*, $P=0.024$) plus another spontaneously hypertensive stroke–prone rat gene (*USMG5*, $P=0.00014$), on chromosomes 10, 13, and 10 respectively, were associated with WMH in subjects from CHARGE. Gene set enrichment showed significant associations for downregulated spontaneously hypertensive stroke–prone rat genes with WMH in humans. In subjects from LBC1936, we replicated CHARGE's genome-wide WMH associations on chromosomes 17 (*TRIM65* and *TRIM47*) and, for the first time, 1 (*PMF1*).

Conclusions—Despite not passing multiple testing thresholds individually, these genes collectively are relevant to known WMH associations, proposed WMH mechanisms, or dementia: associations with Alzheimer's disease, late-life depression, ATP production, osmotic regulation, neurodevelopmental abnormalities, and cognitive impairment. If replicated further, they suggest a multifactorial nature for WMH and argue for more consideration of vascular contributions to dementia. **(*Stroke*. 2015;46:341-347. DOI: 10.1161/STROKEAHA.114.007649.)**

Key Words: genetics ■ humans ■ leukoencephalopathies ■ magnetic resonance imaging

Received October 5, 2014; final revision received November 15, 2014; accepted December 3, 2014.

From the Centre for Cognitive Ageing and Cognitive Epidemiology (L.M.L., M.V.H., S.M.M., M.E.B., J.S., I.J.D., J.M.W.), Division of Neuroimaging Sciences, Brain Research Imaging Centre, (M.V.H., S.M.M., M.E.B., J.M.W.) and Academic Neuropathology (C.S.), Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom; Centre for Cognitive Ageing and Cognitive Epidemiology, Medical Genetics Section, University of Edinburgh Centre for Genomics and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom (W.D.H., S.E.H., D.J.P.); Department of Bioengineering, Imperial College London, London, United Kingdom (E.B.); BHF Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom (M.M., J.M., D.G., A.D.); Department of Biostatistics, Boston University School of Public Health, MA (Q.Y.); The Framingham Heart Study, Boston, MA (Q.Y., S.S.); The Human Genetics Center and Institute of Molecular Medicine, The University of Texas Health Science Center, Houston (M.F.); Departments of Epidemiology, Radiology and Neurology, Erasmus Medical Center, Rotterdam, The Netherlands (M.A.I.); Netherlands Consortium for Healthy Aging, Leiden, The Netherlands (M.A.I.); 12 INSERM U740 (Paris 7 University) and U708 (Bordeaux University), Bordeaux, France (S.D.); Department of Neurology, Lariboisière Hospital, 7 University, DHU Neurovasc Paris Sorbonne, Paris, France (S.D.); University of Versailles Saint-Quentin-en-Yvelines, Versailles, France (S.D.); Department of Neurology, Boston University School of Medicine, MA (S.D., S.S.); Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD (L.L.); Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle (J.C.B.); and Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria (R.S.).

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.114.007649/-/DC1.

Correspondence to Joanna M. Wardlaw, MD, Brain Research Imaging Centre, Division of Neuroimaging Sciences, University of Edinburgh, Bramwell Dott Bldg, Western General Hospital, Crewe Rd, Edinburgh, EH4 2XU, United Kingdom. E-mail joanna.wardlaw@ed.ac.uk

White matter hyperintensities (WMH) of presumed vascular origin, a major component of cerebral small vessel disease (SVD), double the risk of stroke and dementia.[1] Despite considerable societal effect, the causes of WMH and SVD are poorly understood.[2] Conventional vascular risk factors explain little of the WMH variance.[3] Family studies,[4] several rare monogenic SVD disorders,[5] and epidemiology[6] suggest that genetic predisposition is important.

Identification of genetic factors for SVD has been challenging. Several replicable single-nucleotide polymorphisms (SNPs) associated with WMH have been identified in 1 locus on chromosome 17q25,[7,8] although the exact gene(s) and biological pathways to WMH are unclear. Few other replicable genes have been found in genome-wide association studies (GWAS),[9,10] and little is known of their functional significance.

Experimental SVD models might provide insight into human SVD. The spontaneously hypertensive stroke–prone rat (SHRSP) is a relevant model of spontaneous SVD.[11] It was selectively crossbred (1974) from Wistar-Kyoto (WKY) rats via the spontaneously hypertensive rat (SHR, 1963).[12] Hypertension, established in SHRSP rats by 10 weeks of age, is considered to be the main cause of their brain disease. However, differences in protein and gene expression in SHRSP rats versus WKY rats at 5 weeks of age (before measurable blood pressure rises) suggest underlying susceptibilities to SVD.[13] Compared with WKY controls, 5-week-old SHRSP rats have reduced claudin 5 (tight junction) and myelin basic protein and increased microglia (IBA1) and glial activation (GFAP)[13]; at 16 and 21 weeks, increase in smooth muscle actin was seen, thought to reflect arteriolar smooth muscle hyperplasia secondary to hypertension. SHRSP gene expression differences at 5 weeks of age were more numerous than at 16 or 21 weeks of age and included downregulation of *Mmp14*, *Mbp*, *GFAP*, *AVP*, *Alb*, and *Igf2*, upregulation of *Gucy1A3*, *Rps9*, *Fos*, and *JunB*, early-growth response, cell-signaling genes, and overexpression of genes involved in neurological diseases (stroke, depression, and blood–brain barrier leakage),[14] rather than just hypertension. Recent gene sequencing of SHRSP rats (and 26 other rat models of common human diseases)[15] revealed that genes that were either shared between or uniquely mutated in these rat models were significantly over-represented in human GWAS hits for hypertension or metabolism-related phenotypes, suggesting coevolution of these genes and their role in common diseases in models and humans.[15]

In a hypothesis-driven collaborative approach, we tested for associations between genes that were differentially expressed in the brains of 5-week-old SHRSP rats[14] and WMH in humans. We used data from 5-week-old rats because gene expression differences were more frequent at that age than at 16 or 21 weeks, and we wanted to minimize the confounding of tissue changes by secondary effects of hypertension and to optimize the chances of detecting genes related to WMH susceptibility. We focused on WMH as the most frequent feature of SVD with the most data available in replication cohorts. We first tested the subjects from Lothian Birth Cohort 1936 (LBC1936)[16,17] and then attempted replication in subjects from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.[7] To provide confidence in the relevance of subjects from LBC1936, we also sought CHARGE's[7] previously reported WMH-gene associations in the subjects from LBC1936.

Methods

Subjects

The subjects from LBC1936 are community-dwelling individuals living in South East Scotland who underwent detailed cognitive, biomedical, genetic assessments, and detailed brain MRI at ≈73 years of age (n=866).[16,17] The MRI acquisition, methods for assessing WMH burden[17] qualitatively[18] and quantitatively,[19] and proportions with WMH by either method[20] have been reported. This study was approved by the Lothian (REC 07/MRE00/58) and Scottish Multicentre

Table 1. Genes Associated With Cerebral Small Vessel Disease in Rats That Are Associated With WMH in Older Humans: 126 Differentially Expressed Genes Between Spontaneously Hypertensive Stroke Prone and Wild-Type Rats Were Tested for Association With WMH in Subjects From LBC1936 and 10 Genes Were Significantly Associated (*P*<0.05) With Either WMH Volume or Fazekas Score

				Discovery: LBC1936						Replication: CHARGE	
				Genotyped SNPs			Imputed SNPs			Imputed SNPs	
					WMH Volume	Fazekas Score		WMH Volume	Fazekas Score		WMH Volume
Chromosome	Gene	Start Position	Stop Position	nSNPs	*P* Value	*P* Value	nSNPs	*P* Value	*P* Value	nSNPs	*P* Value
4	*AFP*	74 520 796	74 540 356	13	0.0021	0.00090	77	0.0037	0.0037	67	0.841
4	*ALB*	74 488 869	74 505 834	11	0.0026	0.0017	61	0.0063	0.0068	53	0.718
7	*GNAI1*	79 602 075	79 686 661	42	0.034	0.033	181	0.014	0.015	166	0.767
1	*RBM8A*	144 218 994	144 222 801	13	0.038	0.057	26	0.029	0.024	21	0.539
2	*INPP5D*	233 633 279	233 781 288	69	0.041	0.78	198	0.044	0.87	162	0.989
10	*XPNPEP1*	111 614 513	111 673 192	18	0.042	0.14	130	0.15	0.23	120	6.7×10^{-5}
9	*NR4A3*	101 623 957	101 668 994	13	0.045	0.16	62	0.11	0.25	56	0.484
13	*FARP1*	97 593 434	97 900 024	154	0.049	0.25	550	0.18	0.51	468	0.024
6	*MRPL18*	160 131 481	160 139 451	24	0.059	0.039	89	0.16	0.048	76	0.224
1	*SIPA1L2*	230 600 334	230 717 866	80	0.087	0.0093	340	0.20	0.018	285	0.885

nSNPs is the number of SNPs considered in the gene test. CHARGE indicates Cohorts for Heart and Aging Research in Genomic Epidemiology; LBC1936, Lothian Birth Cohort 1936; SNP, single-nucleotide polymorphism; and WMH, white matter hyperintensities.

(MREC/01/0/56) Research Ethics Committees; all subjects gave written informed consent.

The subjects from LBC1936 had genome-wide SNP data on 542 050 SNPs,[21] imputed to 2.5 million SNPs with HapMap2.[22] There were 621 participants (392 men) from LBC1936 with both MRI and genetic data (mean age, 72.67 years; SD=0.73 years; Table I and Methods in the online-only Data Supplement). We excluded 48 subjects from LBC1936 with a history of stroke or dementia.

Gene Analysis

In the 5-week-old SHRSP rats, 162 genes were differentially expressed compared with 5-week-old WKY rats in frontal and midcoronal brain sections (Table II in the online-only Data Supplement).[14] We used the following databases to match the SHRSP Illumina IDs to human genes (Materials and Table II in the online-only Data Supplement): Ensembl—http://www.ensembl.org, GeneCards—http://www.genecards.org, Illumina ID search—http://www.genscript.com, NCBI—http://www.ncbi.nlm.nih.gov, and Rat Genome Database—http://www.rgd.mcw.edu. Of the 162 SHRSP genes, 132 had an equivalent human gene, 8 transcripts were mapped to the same gene, 20 were uncharacterized in humans, and 2 had no human homologue. Of the 132 genes, 126 were available for association testing using the Versatile Gene-based Association Study (VEGAS) test.[23] We first performed a genome-wide association analysis on subjects from LBC1936 using PLINK software[24] to test the genetic association between 542 050 genotyped SNPs and 2 WMH measurements using a linear regression analysis: (1) log transformed WMH volume (mL), with age, sex, intracranial volume, and first 4 multiple dimension scaling components for population stratification as covariates; and (2) summed Fazekas score of periventricular and deep WMH, with age, sex, and the first 4 multiple dimension scaling population stratification components as covariates. We used both WMH volume and Fazekas score[20] to increase the reliability of the results. We did not stratify by vascular risk factors because hypertension (although it was the strongest vascular risk factor) explained <2% of WMH variance in subjects from LBC1936.[3] The VEGAS software summarized evidence for association with WMH in subjects from LBC1936 per gene by considering the *P* values of all 543 050 SNPs that were located within 17 681 unique autosomal genes (including SNPs±50 kb outside of genes to include regulatory regions). For a more direct comparison with CHARGE (which used imputed data), we also performed a gene-based test on LBC1936's 2 447 226 HapMap2 derived *P* values (after removing SNPs with a minor allele frequency of <0.01 and imputation quality of <0.3) with VEGAS software as above.

Replication in Subjects From CHARGE

We then tested whether any of the 126 SHRSP genes were also associated with WMH in subjects from CHARGE by using data from CHARGE's published genome-wide meta-analysis of WMH in 9361 stroke-free individuals from 7 community-based cohorts.[7] We performed a gene-based test using VEGAS software, which summarized the evidence for association with WMH burden on a per gene basis, as above, by considering the associated *P* values of all HapMap2 SNPs located within 17 787 autosomal genes (including SNPs±50 kb outside of genes to include regulatory regions).

Gene Set Enrichment

We performed a gene set enrichment analysis[25] to investigate the enrichment of the 126 SHRSP genes in the LBC1936 and CHARGE data associated with WMH, accounting for whether these were upregulated or downregulated (online-only Data Supplement),[26] corrected for multiple testing using a false discovery rate (FDR) method.[27]

Replication of Previous CHARGE Findings in Subjects From LBC1936

To demonstrate our ability to detect WMH-gene associations in subjects from LBC1936, we attempted replication of CHARGE's genome-wide associations with WMH[7,8] in the subjects from the LBC1936 Cohort in a genome-wide association analysis using the 2 534 887 SNPs imputed to HapMap2, with WMH (volume and Fazekas score) in Mach2QTL software.[28]

We applied Bonferroni correction for multiple testing (P=0.05/126 genes=0.0004). We did not include the 2 WMH phenotypes in the Bonferroni correction as they are highly correlated (r^2=0.77). Because of the overconservative nature of Bonferroni correction for multiple testing,[29] a nominal significance threshold of P value of <0.05 was required for replication efforts.

Results

SHRSP Genes in Subjects From LBC1936

Of the 126 candidate SHRSP-derived genes, 10 were nominally associated with WMH in subjects from LBC1936 (P<0.05; Table 1). Using imputed or genotyped data, 5 genes were associated with WMH volume (*AFP*, *ALB*, *GNAI1* [*RBM8A* and *INPP5D*, both borderline]); 3 of these (*AFP*, *ALB*, and *GNAI1*) and 2 others (*MRPL18* and *SIPA1L2*) were associated with WMH Fazekas scores. Three other genes were associated with WMH volume using genotyped data only (*XNXPEP1*, *NR4A3*, and *FARP1*). None of these genes individually passed Bonferroni correction in subjects from LBC1936 (all were P>0.0004), in part, reflecting the LBC1936 sample size.

SHRSP Genes in Subjects From CHARGE

Two of these 10 genes were also associated with WMH in subjects from CHARGE (*XPNPEP1*, $P=6.7\times10^{-5}$; and *FARP1*, P=0.024; Table 1). Full details of all 126 SHRSP to LBC1936 to CHARGE gene associations are given in Table III in the online-only Data Supplement. Several other of the 126 SHRSP genes (outside the 10/126 described above) showed significance at P<0.05 in subjects from CHARGE (eg, *USMG5*, *MED17*, *ZNF461*, *C20orf7*, *EGR1*, *ARC*, *NUDT14*, and *MMP14*) of which 1 (*USMG5*, P<0.000142) passed Bonferroni correction (P<0.0004).

Gene Set Enrichment

Using gene set enrichment analysis, all 126 SHRSP candidate genes were not enriched in subjects from LBC1936 for association with WMH in the 17 681 genes tested here (WMH volume, P=0.34; Fazekas score, P=0.81), but this would not preclude the possibility that in either upregulated or downregulated gene sets, there was an abundance of genes showing an enriched association. We tested the upregulated (n=76) and downregulated (n=50) SHRSP genes separately and found significant enrichment for Fazekas scores in SHRSP downregulated genes (P=0.035; FDR, 0.046) but not SHRSP upregulated genes (P=0.921; FDR, 0.899). WMH volume showed significant enrichment in downregulated (P=0.018; FDR, 0.025) but not upregulated (P=0.802; FDR, 0.780) genes. In the CHARGE consortium, there was no significant enrichment for either the total set of 126 genes (P=0.0514), the upregulated (P=0.109; FDR, 0.266) or the downregulated genes (P=0.173; FDR, 0.149).

Replication of CHARGE's Previous Genome-Wide Association in Subjects From LBC1936

We sought CHARGE's previous genome-wide association results for WMH[7] in subjects from LBC1936. Of CHARGE's

15 SNPs ($P<1\times10^{-5}$) associated with WMH (Table 2),[7] 3 SNPs replicated in subjects from LBC1936 with both WMH volume and Fazekas score at $P<0.05$ (rs3744028, rs1055129, and rs1052053); rs1052053, a miss-sense variant on chromosome 1 in the polyamine-modulated factor 1 gene (*PMF1*), has not replicated previously.

Discussion

We used a clinically relevant translational approach[15] to identify potential new gene associations for WMH, a common cause of cognitive impairment, stroke, and dementia. We found parallels between differentially expressed genes in a young spontaneous SVD model and WMH-gene associations in older humans. Two novel genes on chromosome 10 derived from SHRSP rats were associated with WMH, *XPNPEP1* in both LBC1936 and CHARGE and *USMG5* in CHARGE only. Several other genes were nominally associated with WMH in LBC1936 or CHARGE although none passed multiple testing. We replicated 3 of CHARGE's WMH-gene associations in subjects from LBC1936: 2 (rs3744028 and rs1055129) on chromosome 17q25 and 1 previously unreplicated SNP (rs1052053) on chromosome 1, a miss-sense variant in the polyamine-modulated factor 1 gene, *PMF1*, that has a role in the cell cycle. Jointly, these approaches yielded 6 genes (3 from the SHRSP rats and 3 replicates of a GWAS finding) and 5 further rat-derived genes based on the LBC1936 sample alone, which despite not passing multiple testing thresholds individually, as a group they are notable for their involvement in biological pathways relevant to WMH pathogenesis.[2]

Of the 2 SHRSP genes found in LBC1936 and CHARGE, *XPNPEP1* is X-prolyl aminopeptidase (aminopeptidase P) 1, soluble, associated with biliary atresia, and located in a region on chromosome 10 that is associated with Alzheimer's disease.[30] *FARP1* is Pleckstrin domain protein 1, associated with brain volume differences,[31] and important in synapse development.[32] The SHRSP-CHARGE–associated gene *USMG5* is upregulated during skeletal muscle growth 5 homolog (also known as diabetes mellitus–associated protein in insulin sensitive tissues, or *DAPIT*), sits on chromosome 10, and maintains ATP synthase populations in mitochondria.[33] All 5 SHRSP genes associated with both WMH volume and Fazekas score in subjects from LBC1936 (*AFP*, *ALB*, *GNAI1*, *RBM8A*, and *MRPL18)* are associated with white matter–relevant diseases in humans. Despite not surviving correction for multiple testing, there was a notable consistency in their association with 2 separate WMH measures. *AFP* encodes α-fetoprotein, a major plasma protein produced in the yolk sac and liver during fetal life. Abnormally, high amounts of α-fetoprotein are found in ataxia telangiectasia,[34] also associated with abnormal white matter.[35] *ALB* encodes albumin, a soluble monomeric protein important for maintaining plasma oncotic pressure found in cerebral WMH,[36] and cerebrospinal fluid as blood–brain barrier function deteriorates with ageing and dementia.[2,37] *GNAI1* encodes guanine nucleotide–binding protein (G protein), alpha-inhibiting activity polypeptide 1, implicated with Alzheimer's disease.[38] *RBM8A* is an RNA binding protein that has differential expression in Alzheimer's disease,[39] associations with a range of intellectual disabilities in humans and anxiety-related behavior in mice,[40] with schizophrenia, several neurodevelopmental intellectual disabilities, anxiety behavior and may target neuronal genes to regulate behaviors. WMH in old age are known associates of late-onset depression,[41] and they are also associated with lower age 11 IQ.[42] *MRPL18* is the mitochondrial ribosomal protein L18, previously associated

Table 2. Association of SNPs Previously Associated With WMH in CHARGE in Subjects From LBC1936 and the Corresponding SNP Association Results Are Given for LBC1936 WMH Volume and Fazekas Score

			CHARGE			LBC1936			WMH Volume		Fazekas Score	
SNP	Chromosome	Nearest Gene	Risk Allele	Allele Freq	*P* Value	Effect Allele	Allele Freq	r^2	β	*P* Value	β	*P* Value
rs3744028	17	*TRIM65*	C	0.18	4.0×10^{-9}	T	0.81	0.99	−0.217	0.00287	−0.287	0.000511
rs1055129	17	*TRIM47*	G	0.30	4.1×10^{-8}	G	0.28	0.97	0.286	9.5×10^{-6}	0.305	3.34×10^{-5}
rs7894407	10	*PDCD11*	T	0.63	6.1×10^{-7}	T	0.63	0.99	−0.026	0.662	−0.029	0.665
rs1892525	1	*RP11-518D3.1*	G	0.69	7.2×10^{-7}	G	0.73	0.99	0.070	0.269	0.107	0.135
rs10814323	9	*RECK*	A	0.21	1.7×10^{-6}	G	0.77	1.00	0.056	0.390	0.034	0.651
rs6992136	8	*RPL32P19*	G	0.85	3.2×10^{-6}	G	0.85	0.81	0.101	0.259	0.075	0.458
rs11731436	4	*AC097110.1*	C	0.64	3.3×10^{-6}	G	0.35	0.91	−0.035	0.565	−0.041	0.549
rs1052053	1	*PMF1*	A	0.62	5×10^{-6}	G	0.39	1.00	−0.112	0.047	−0.127	0.048
rs2167089	3	*AC098970.2*	G	0.73	6×10^{-6}	T	0.26	0.97	0.061	0.342	0.044	0.545
rs10012573	4	*COL25A1*	A	0.94	6×10^{-6}	C	0.06	0.85	0.097	0.481	0.029	0.855
rs11625623	14	*PTGDR*	G	0.23	7.7×10^{-6}	G	0.23	1.00	−0.051	0.460	−0.021	0.792
rs16901064	5	*RNASEN*	C	0.84	7.8×10^{-6}	C	0.85	0.99	0.030	0.695	0.055	0.532
rs6945846	7	*FOXP2*	C	0.2	7.9×10^{-6}	T	0.78	0.90	−0.036	0.625	0.113	0.175
rs11629135	14	*MTHFD1*	G	0.93	8.6×10^{-6}	G	0.92	0.99	0.034	0.749	−0.057	0.641
rs9410016	9	*C9orf62*	G	0.41	9.7×10^{-6}	G	0.39	0.99	−0.029	0.603	−0.041	0.521

Allele frequency is the frequency of the effect allele. r^2 is a measure of the imputation quality to HapMap2. β is the regression coefficient. CHARGE indicates Cohorts for Heart and Aging Research in Genomic Epidemiology; LBC1936, Lothian Birth Cohort 1936; SNP, single-nucleotide polymorphism; and WMH, white matter hyperintensities.

with multiple sclerosis.[43] These 7 SHRSP-derived genes are related to pathologies (ataxia telangiectasia, blood–brain barrier impairment, Alzheimer's disease, multiple sclerosis, depression, developmental intellectual disabilities, and brain size) that display white matter abnormalities or affect intellectual function. Impaired ATP production because of defects in *USMG5*, the gene that replicated from SHRSP to CHARGE, could increase susceptibility to WMH via ischemia.

The genes that were downregulated in the SHRSP were significantly enriched in subjects from LBC1936 for WMH. This may be because, in a complex disease such as SVD/WMH, several individually modest genetic defects in different components of key pathways, when present in combination, increase disease risk. This interpretation is consistent with differential protein expression seen in SHRSP[13] and the absence, so far, of individual major human gene defects explaining either sporadic WMH or lacunar stroke.[9]

The lack of consistent replication from SHRSP to LBC1936 to CHARGE requires caution. The power and required significance threshold of the LBC1936 was modest for GWAS, hence our hypothesis-driven approach. Genes associated with WMH in subjects from LBC1936 but not CHARGE could be false positives; other factors include greater heterogeneity of WMH assessment and greater age range in subjects from CHARGE. The narrow age range of subjects from LBC1936 minimizes the effect of age, possibly helping to expose relevant genes. CHARGE-contributing studies used several methods of quantifying WMH, different MR scanner field strengths, and generations of technology and sequences. However, WMH volume and visual scores are highly correlated,[20] and our replication of 3 findings from CHARGE in subjects from LBC1936 suggests that our approach has some validity. The CHARGE cohorts may have used different imputation platforms or more SNPS may have failed quality assurance in subjects from LBC1936, contributing to differences between the imputation results. There are several limitations to gene-based analysis, including the omission of nonautosomal genes, the effect of noncausal SNPs to dilute association (in particular, in the presence of a strong genetic association with a single locus within or in the regulatory region of a given gene, thus missing important associations), the lack of knowledge on (and overlap of) gene boundaries, the possibility that an SNP variant may influence a gene distal to its site, thus not corresponding to a gene that it is located next to it, and the potential of the genetic data not to tag causative genetic variants. Power may have been limited (despite CHARGE's large sample size) to detect associations with some genes. We did not stratify the human cohorts by risk factors as these explained <2% of WMH variance in subjects from LBC1936,[3] and risk-stratified genetic data were unavailable for CHARGE. We did not test gene associations with other SVD features in addition to WMH because a total SVD burden score was not available for CHARGE. Although it is a relevant model of spontaneous SVD[11,12] and of human hypertension and metabolic disorders,[15] like any model, the SHRSP has translational limitations, arguing for additional studies at different ages and brain regions, with or without environmental stressors.

This work has the following strengths: accurate LBC1936 WMH phenotyping[17] and genetic information in this relatively large narrow age-range older population.[16] The Glasgow SHRSP colony is long established, with carefully controlled environments. The mRNA data were obtained from the same rats that provided protein expression data.[13] Replication in other SHRSP colonies and examination of related strains (eg, SHR's) may be informative. The genomes of SHRSP and 26 other complex disease phenotype models were recently sequenced,[15] showing associations between genes in rat models of hypertension and human GWAS hits for hypertension phenotypes.[15] This provides support for our reverse-translational discovery approach, suggesting that genes in disease models have coevolved and may contribute to disease-related phenotypes in humans.

Our findings require validation. The selection of candidate genes for investigation could be widened by examining more genes from the 5-week-old SHRSP rats (Table II in the online-only Data Supplement), other models,[15] and in larger samples of well-phenotyped humans, such as from METASTROKE and the Wellcome Trust Case-Control Consortium. This translational analysis of experimental models and human disease suggests some aspects of the genetic architecture underlying SVD, stroke, and dementia and argues for greater awareness of vascular contributions to neurodegeneration.

Figure I and Tables IV and V in the online-only Data Supplement provide the top SNP ($P<1\times10^{-5}$) and gene ($P<0.001$) associations with WMH variables in subjects from LBC1936 for further reference.

Acknowledgments

We thank the Lothian Birth Cohort 1936 participants and research team members, Wellcome Trust Clinical Research Facility (http://www.wtcrf.ed.ac.uk, subject testing and genotyping), and Brain Research Imaging Centre (http://www.bric.ed.ac.uk, brain imaging and analyses). Cohorts for Heart and Aging Research in Genomic Epidemiology thanks the staff and participants of the Aging Gene-Environment Susceptibility-Reykjavik Study, Atherosclerosis Risk in Community Study (ARIC), Austrian Stroke Prevention Study (ASPS), Cardiovascular Health Study, and Framingham Heart and Rotterdam Studies for their important contributions. ASPS thanks Birgit Reinhart for her long-term administrative commitment and Ing Johann Semmler for the technical assistance at creating the DNA bank. Drs Wardlaw, Bailey, McBride, Graham, Dominiczak, Deary, Starr, Seshadri, Fornage, Ikram, Debette, Launer, Bis, and Schmidt contributed to data collection. Dr Lopez, Harris, Hill, Yang, Bailey, McClure, McBride, Smith, Hernandez, Maniega, Bastin, and Wardlaw contributed to data analysis. Drs Wardlaw, Deary, and Seshadri contributed to study design, co-ordination, and funding. Lopez, Wardlaw, Seshadri, and Deary contributed to article preparation. Lopez, Harris, Hill, Porteous, Smith, Deary, Starr, Seshadri, Yang, Fornage, Ikram, Debette Launer, Bis, Schmidt, Bailey, McBride, Graham, McClure, Dominiczak, Hernandez, Maniega, Bastin, and Wardlaw contributed to article review. Wardlaw was the guarantor and provided the overall concept.

Sources of Funding

Lothian Birth Cohort 1936 was funded by Age UK's Disconnected Mind programme (http://www.disconnectedmind.ed.ac.uk) and by Research Into Ageing (references 251 and 285). Whole-genome association was funded by Biotechnology and Biological Sciences Research Council (reference BB/F019394/1), brain image analysis was funded by Medical Research Council (G1001401 and 8200), and imaging was funded by Brain Research Imaging Centre (http://www.bric.ed.ac.uk), The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (http://www.

ccace.ed.ac.uk, G0700704/84698). The SHRSP analysis was funded by Medical Research Council (Dr Bailey), British Neuropathological Society, Newby Fund, Scottish Funding Council Scottish Imaging Network A Platform for Scientific Excellence Collaboration (http://www.sinapse.ac.uk), and European Union Community's FP7/2007–2013, grant agreement HEALTH-F4-2010-241504 EURATRANS. The Aging Gene-Environment Susceptibility-Reykjavik Study was funded by National Institutes of Ageing (NIA), contract N01-AG-12100, National Eye Institute, National Institute on Deafness and Other Communication Disorders, National Heart, Lung, and Blood Institute (NHLBI), the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The Atherosclerosis Risk in Communities Study is a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, and HHSN268201100011C, HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health (NIH) contract HHSN268200625226C; NIH grant number UL1RR025005, and NIH Roadmap for Medical Research and grant number HL093029. The Austrian Stroke Prevention Study was funded by the Austrian Science Fond grant numbers P20545-P05 and P13180. The Medical University of Graz supports the databank of the ASPS. The Cardiovascular Health Study was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, HHSN268200960009C, and HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from National Institute of Neurological Disorders and Stroke, AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA (http://www.chs-nhlbi.org/pi.htm). DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR), the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, Clinical and Translational Science Institute grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Framingham Heart Study was supported by the NHLBI (Contract no. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract no. N02-HL-6-4278); some of this research used the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center; and grants from the National Institute of Neurological Disorders and Stroke (R01 NS17950), and the National Institute of Aging (R01s AG08122, AG16495, and AG033193), and the NHLBI (U01 HL096917 and R01 HL093029). The Rotterdam Study Genome-Wide Association database was funded by the Netherlands Organization of Scientific Research (NOW; no. 175.010.2005.011); the study was further supported by the Netherlands Genomics Initiative/NWO project no. 050-060-810; the Erasmus Medical Center and Erasmus University, Rotterdam; NWO, the Netherlands Organization for the Health Research and Development, the Research Institute for Diseases in the Elderly, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (Dr Graham, XII), and the Municipality of Rotterdam and Dr Ikram was supported by the Nederlandse Hartstichting grant 2009B102 and ZonMW Veni-grant 916.13.054.

Disclosures

None.

References

1. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ*. 2010;341:c3666. doi: 10.1136/bmj.c3666.
2. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurol*. 2013;12:483–497. doi: 10.1016/S1474-4422(13)70060-7.
3. Wardlaw JM, Allerhand M, Doubal FN, Valdes Hernandez M, Morris Z, Gow AJ, et al. Vascular risk factors, large artery atheroma, and brain white matter hyperintensities. *Neurology*. 2014;82:1331–1338. doi: 10.1212/WNL.0000000000000312.
4. Kochunov P, Glahn D, Winkler A, Duggirala R, Olvera RL, Cole S, et al. Analysis of genetic variability and whole genome linkage of whole-brain, subcortical, and ependymal hyperintense white matter volume. *Stroke*. 2009;40:3685–3690. doi: 10.1161/STROKEAHA.109.565390.
5. Yamamoto Y, Craggs L, Baumann M, Kalimo H, Kalaria RN. Review: molecular genetics and pathology of hereditary small vessel diseases of the brain. *Neuropathol Appl Neurobiol*. 2011;37:94–113. doi: 10.1111/j.1365-2990.2010.01147.x.
6. Jackson CA, Hutchison A, Dennis MS, Wardlaw JM, Lindgren A, Norrving B, et al. Differing risk factor profiles of ischemic stroke subtypes: evidence for a distinct lacunar arteriopathy? *Stroke*. 2010;41:624–629. doi: 10.1161/STROKEAHA.109.558809.
7. Fornage M, Debette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann Neurol*. 2011;69:928–939. doi: 10.1002/ana.22403.
8. Verhaaren BF, de Boer R, Vernooij MW, Rivadeneira F, Uitterlinden AG, Hofman A, et al. Replication study of chr17q25 with cerebral white matter lesion volume. *Stroke*. 2011;42:3297–3299. doi: 10.1161/STROKEAHA.111.623090.
9. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al; Australian Stroke Genetics Collaborative; Wellcome Trust Case Control Consortium 2 (WTCCC2); International Stroke Genetics Consortium. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951–962. doi: 10.1016/S1474-4422(12)70234-X.
10. Adib-Samii P, Rost N, Traylor M, Devan W, Biffi A, Lanfranconi S, et al; Australian Stroke Genetics Collaborative; Wellcome Trust Case-Control Consortium-2 (WTCCC2); METASTROKE; International Stroke Genetics Consortium. 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke*. 2013;44:1609–1615. doi: 10.1161/STROKEAHA.113.679936.
11. Bailey EL, McCulloch J, Sudlow C, Wardlaw JM. Potential animal models of lacunar stroke: a systematic review. *Stroke*. 2009;40:e451–e458. doi: 10.1161/STROKEAHA.108.528430.
12. Bailey EL, Smith C, Sudlow CM, Wardlaw JM. Is the spontaneously hypertensive stroke prone rat a pertinent model of subcortical ischemic stroke? A systematic review. *Int J Stroke*. 2011;6:434–444. doi: 10.1111/j.1747-4949.2011.00659.x.
13. Bailey EL, Wardlaw JM, Graham D, Dominiczak AF, Sudlow CL, Smith C. Cerebral small vessel endothelial structural changes predate hypertension in stroke-prone spontaneously hypertensive rats: a blinded, controlled immunohistochemical study of 5- to 21-week old rats. *Neuropathol Appl Neurobiol*. 2011;37:711–726. doi: 10.1111/j.1365-2990.2011.01170.x.
14. Bailey EL, McBride MW, Beattie W, McClure JD, Graham D, Dominiczak AF, et al. Differential gene expression in multiple neurological, inflammatory and connective tissue pathways in a spontaneous model of human small vessel stroke. *Neuropathol Appl Neurobiol*. 2014;40:855–872. doi: 10.1111/nan.12116.
15. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, et al. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell*. 2013;154:691–703. doi: 10.1016/j.cell.2013.06.040.
16. Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr*. 2007;7:28. doi: 10.1186/1471-2318-7-28.
17. Wardlaw JM, Bastin ME, Valdés Hernández MC, Maniega SM, Royle NA, Morris Z, et al. Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *Int J Stroke*. 2011;6:547–559. doi: 10.1111/j.1747-4949.2011.00683.x.
18. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol*. 1987;149:351–356. doi: 10.2214/ajr.149.2.351.
19. Hernández Mdel C, Ferguson KJ, Chappell FM, Wardlaw JM. New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. *Eur Radiol*. 2010;20:1684–1691. doi: 10.1007/s00330-010-1718-6.

20. Valdés Hernández Mdel C, Morris Z, Dickie DA, Royle NA, Muñoz Maniega S, Aribisala BS, et al. Close correlation between quantitative and qualitative assessments of white matter lesions. *Neuroepidemiology*. 2013;40:13–22. doi: 10.1159/000341859.
21. Houlihan LM, Davies G, Tenesa A, Harris SE, Luciano M, Gow AJ, et al. Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *Am J Hum Genet*. 2010;86:626–631. doi: 10.1016/j.ajhg.2010.02.016.
22. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al; LifeLines Cohort Study; EchoGen consortium; AortaGen Consortium; CHARGE Consortium Heart Failure Working Group; KidneyGen consortium; CKDGen consortium; Cardiogenics consortium; CardioGram. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet*. 2011;43:1005–1011. doi: 10.1038/ng.922.
23. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al; AMFS Investigators. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010;87:139–145. doi: 10.1016/j.ajhg.2010.06.009.
24. 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;81:559–575. doi: 10.1086/519795.
25. Hill WD, Davies G, van de Lagemaat LN, Christoforou A, Marioni RE, Fernandes CP, et al. Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl Psychiatry*. 2014;4:e341. doi: 10.1038/tp.2013.114.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–15550. doi: 10.1073/pnas.0506580102.
27. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet*. 2007;81:1278–1283. doi: 10.1086/522374.
28. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34:816–834. doi: 10.1002/gepi.20533.
29. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316:1236–1238. doi: 10.1136/bmj.316.7139.1236.
30. Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, et al. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. *Am J Hum Genet*. 2006;78:78–88. doi: 10.1086/498851.
31. Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al; Alzheimer's Disease Neuroimaging Initiative. Voxelwise genome-wide association study (vGWAS). *Neuroimage*. 2010;53:1160–1174. doi: 10.1016/j.neuroimage.2010.02.032.
32. Cheadle L, Biederer T. The novel synaptogenic protein Farp1 links postsynaptic cytoskeletal dynamics and transsynaptic organization. *J Cell Biol*. 2012;199:985–1001. doi: 10.1083/jcb.201205041.
33. Ohsakaya S, Fujikawa M, Hisabori T, Yoshida M. Knockdown of DAPIT (diabetes-associated protein in insulin-sensitive tissue) results in loss of ATP synthase in mitochondria. *J Biol Chem*. 2011;286:20292–20296. doi: 10.1074/jbc.M110.198523.
34. Waldmann TA, McIntire KR. Serum-alpha-fetoprotein levels in patients with ataxia-telangiectasia. *Lancet*. 1972;2:1112–1115. doi: 10.1016/S0140-6736(72)92717-1.
35. Ciemins JJ, Horowitz AL. Abnormal white matter signal in ataxia telangiectasia. *AJNR Am J Neuroradiol*. 2000;21:1483–1485.
36. Grinberg LT, Thal DR. Vascular pathology in the aged human brain. *Acta Neuropathol*. 2010;119:277–290. doi: 10.1007/s00401-010-0652-7.
37. Farrall AJ, Wardlaw JM. Blood-brain barrier: ageing and microvascular disease–systematic review and meta-analysis. *Neurobiol Aging*. 2009;30:337–352. doi: 10.1016/j.neurobiolaging.2007.07.015.
38. Silver M, Janousova E, Hua X, Thompson PM, Montana G; Alzheimer's Disease Neuroimaging Initiative. Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *Neuroimage*. 2012;63:1681–1694. doi: 10.1016/j.neuroimage.2012.08.002.
39. Wong J. Altered expression of RNA splicing proteins in Alzheimer's disease patients: evidence from two microarray studies. *Dement Geriatr Cogn Dis Extra*. 2013;3:74–85. doi: 10.1159/000348406.
40. Alachkar A, Jiang D, Harrison M, Zhou Y, Chen G, Mao Y. An EJC factor RBM8a regulates anxiety behaviors. *Curr Mol Med*. 2013;13:887–899. doi: 10.2174/15665240113139990019.
41. Herrmann LL, Le Masurier M, Ebmeier KP. White matter hyperintensities in late life depression: a systematic review. *J Neurol Neurosurg Psychiatry*. 2008;79:619–624. doi: 10.1136/jnnp.2007.124651.
42. Valdés Hernández Mdel C, Booth T, Murray C, Gow AJ, Penke L, Morris Z, et al. Brain white matter damage in aging and cognitive ability in youth and older age. *Neurobiol Aging*. 2013;34:2740–2747. doi: 10.1016/j.neurobiolaging.2013.05.032.
43. Fischer MT, Sharma R, Lim JL, Haider L, Frischer JM, Drexhage J, et al. NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury. *Brain*. 2012;135(pt 3):886–899. doi: 10.1093/brain/aws012.

Stroke
Journal of the American Heart Association

Genes From a Translational Analysis Support a Multifactorial Nature of White Matter Hyperintensities

Lorna M. Lopez, W. David Hill, Sarah E. Harris, Maria Valdes Hernandez, Susana Munoz Maniega, Mark E. Bastin, Emma Bailey, Colin Smith, Martin McBride, John McClure, Delyth Graham, Anna Dominiczak, Qiong Yang, Myriam Fornage, M. Arfan Ikram, Stephanie Debette, Lenore Launer, Joshua C. Bis, Reinhold Schmidt, Sudha Seshadri, David J. Porteous, John Starr, Ian J. Deary and Joanna M. Wardlaw

Stroke. 2015;46:341-347; originally published online January 13, 2015;
doi: 10.1161/STROKEAHA.114.007649
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0039-2499. Online ISSN: 1524-4628

SUPPLEMENTAL MATERIAL

Genes from a translational analysis support a multifactorial nature of white matter hyperintensities

Lorna M. Lopez, PhD[1], W. David Hill, MSc[1], Sarah E. Harris, PhD[2], Maria Valdez Hernandez, PhD,[1,3] Susana Munoz Maniega, PhD,[1,3] Mark E Bastin, PhD,[1,3] Emma Bailey, PhD,[4] Colin Smith, MD[5], Martin McBride, PhD[6], John McClure, PhD[6], Delyth Graham, PhD[6], Anna Dominiczak, MD[6], Qiong Yang, PhD[7,8], Myriam Fornage, PhD[9], M. Arfan Ikram, MD, PhD[10-12], Stephanie Debette, MD, PhD,[13-16] Lenore Launer, PhD[17], Joshua C. Bis, PhD,[18] Reinhold Schmidt, MD[19], Sudha Seshadri, MD[8,15], David J. Porteous, PhD[2], John Starr, MD[1], Ian J. Deary, PhD[1], Joanna M. Wardlaw, MD[3*].

Index

Supplementary Methods

Genotyping

A detailed description of the genotyping method is described elsewhere.[1] Briefly, genotyping was performed using Illumina Human 610-Quadv1 arrays on blood-extracted DNA at the WTCRF Genetics Core. All individuals were checked for disagreement between genetic and reported gender. Relatedness between subjects was investigated and, for any related pair of individuals, one was removed. Samples with a call rate ≤ 0.95, and those showing evidence of non-Caucasian ascent by multidimensional scaling (MDS), were also removed. SNPs were included in the analyses if they met the following conditions: call rate ≥ 0.98, minor allele frequency ≥ 0.01, and Hardy-Weinberg Equilibrium test with $P \geq 0.001$. The final number of genotyped SNPs included in the study was 542,050 in 1,005 individuals.

Genetic imputation

~2.5M common SNPs included in HapMap, using the HapMap phase II CEU data as the reference sample were imputed. NCBI build 36 (UCSC hg18) was used and genotype data were imputed using MACH software. Prior to imputation SNPs were removed that diverged from HWE with a significance $p < 1x10^{-3}$ and SNPs with a minor allele frequency < 0.01.

Gene mapping.

Listed below are the databases used to match the Illumina IDs from the SHRSP study to human genes as shown in Supplementary Table II.

Ensembl – www.ensembl.org., GeneCards – www.genecards.org, Illumina ID search - www.genscript.com, NCBI - http://www.ncbi.nlm.nih.gov, Rat Genome Database - rgd.mcw.edu.

Gene set enrichment analysis

A gene set enrichment analysis was performed to investigate the enrichment of 126 SHRSP genes in WMH gene associations. First, the gene based statistics from VEGAS were rank ordered before being –log(10) transformed. Gene set enrichment analysis (GSEA) uses a set of candidate gene identifiers and a genome wide set of genes, ranked based on their association with a phenotype. Next, a weighted Kolmogorov-Smirnov type statistic, walks down the genome wide ranked set of genes and increases the test statistic each time it finds a gene that matches one from the candidate gene set and decreases it when it does not.[2, 3] The magnitude of the increase is proportional to its p value, allowing for information regarding rank and distance between ranks to be used in the calculation of enrichment. The maximum deviation from zero is assigned to the candidate gene set (this is the enrichment score or ES). The gene set is then permuted before the ES being re-calculated. The p-value describes the proportion of the 5,000 permuted enrichment scores that the observed enrichment score was greater than.

Table I: **Description of LBC1936 WMH variables.** Fazekas scale for periventricular lesions, Fazekas scale for deep lesions and the sum of these Fazeskas scores are described. WMH as percentage of true WMH alone in intracranial volume (ICV) are listed.

Trait	Median	Mean	Standard deviation	Minimum	Maximum
Age in years	72.72	72.67	0.73	71.04	74.22
Fazekas Peri	1	1.34	0.635	0	3
Fazekas Deep	1	1.08	0.655	0	3
Fazekas Sum	2	2.42	1.12	0	6
ICV mm^3	1,448,490	1,452,235.57	141,870.37	1,059,966	1,876,420
WMH volume mm^3	7,554	11,885.97	12,826.30	0	98,378
WMH transformed	2.1464	2.1135	0.9920	0	4.60
WMH in ICV	0.5245	0.8203	0.8999	0	7.47

Table II: Candidate genes from SHRSP rat model. This is a list of the 162 transcripts differentially expressed between SHRSP and WKY at 5 weeks in two relevant brain regions from Bailey et al[4] and the corresponding human genes.

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
PROBE_ID	**ILMN_GENE**	**CHROM-OSOME**	**WKY vs SHRSP frontal (FDR)**	**WKY vs SHRSP mid-coronal (FDR)**	**Rat Gene Symbol**	**Human Gene**	**Human Gene Symbol, Human Gene Description**	**Notes**
ILMN_1365113	RGD1564649_PREDICTED		0	0	*Rps9*	*RPS9*	RPS9 Homo sapiens ribosomal protein S9 (RPS9), mRNA.	similar to 40S ribosomal protein S9 (Pseudo gene)
ILMN_2038795	RPS9	1	0	0	*Rps9*	*RPS9*	As above	
ILMN_1371357	LOC497757		0	0	*Gucy1a3*	*GUCY1A3*	GUCY1A3 Homo sapiens guanylate cyclase 1, soluble, alpha 3 (GUCY1A3), transcript variant 1, mRNA.	
ILMN_2038796	RPS9	1	0	0	*Rps9*	*RPS9*	As above	
ILMN_1359040	RGD1561110_PREDICTED	2	0	0	*Fam151b*	*FAM151B*	FAM151B Homo sapiens family with sequence similarity 151, member B (FAM151B), mRNA.	
ILMN_1368735	RGD1311103_PREDICTED		0	0		*OOEP*	OOEP Homo sapiens oocyte expressed protein homolog (dog) (OOEP), mRNA.	OOEP is an ortholog
ILMN_2039673	ARC	7	0	0	*Arc*	*ARC*	ARC Homo sapiens activity-regulated cytoskeleton-associated protein (ARC), mRNA.	
ILMN_1361865	ZFP597	10	0	0	*Zfp597*	*ZNF597*	ZNF597 Homo sapiens zinc finger protein 597 (ZNF597), mRNA.	
ILMN_1372230	RNF149	9	4.13E-05	0	*Rnf149*	*RNF149*	RNF149 Homo sapiens ring finger protein 149 (RNF149), mRNA.	
ILMN_1350784	JUNB	19	0.00010101	0.00030303	*Junb*	*JUNB*	JUNB Homo sapiens jun B proto-oncogene (JUNB), mRNA.	
ILMN_1351340	LOC500950		7.27E-05	0.000187166	*Zfp317*	*ZNF317*	ZNF317 Homo sapiens zinc finger protein 317 (ZNF317), transcript variant 1, mRNA.	similar to zinc finger protein 75
ILMN_1368356	FOS	6	0.000117302	0.00030303	*Fos*	*FOS*	FOS Homo sapiens FBJ murine osteosarcoma viral oncogene homolog (FOS), mRNA.	
ILMN_1367486	DUSP1	10	0.00010101	0	*Dusp1*	*DUSP1*	DUSP1 Homo sapiens dual specificity phosphatase 1 (DUSP1), mRNA.	
ILMN_1371004	RPS16		0.000117302	0	*Rps16*	*RPS16*	RPS16 Homo sapiens ribosomal protein S16 (RPS16), mRNA.	
ILMN_1367530	LOC497727		0.000160428	0.000146628	*Sipa1l2*	*SIPA1L2*	SIPA1L2 Homo sapiens signal-induced proliferation-associated 1 like 2 (SIPA1L2), mRNA.	hypothetical protein XP_579313
ILMN_1372711	LOC502316		0.000160428	0.000146628	*Zfp566*	*ZNF566*	ZNF566 Homo sapiens zinc finger protein 566 (ZNF566), transcript variant 1, mRNA.	similar to Zinc finger protein 566
ILMN_1367467	PER2	9	0.000383838	0.000465632	*Per2*	*PER2*	PER2 Homo sapiens period homolog 2 (Drosophila) (PER2), mRNA.	
ILMN_1367162	GPM6A	16	0.000317125	0.000386364	*Gpm6a*	*GPM6A*	GPM6A Homo sapiens glycoprotein M6A (GPM6A), transcript variant 1, mRNA.	
ILMN_1352667	NAB1		0.000317125	0.000890538	*Nab1*	*NAB1*	NAB1 Homo sapiens NGFI-A binding protein 1 (EGR1 binding protein 1) (NAB1), mRNA.	
ILMN_1366713	ZNF575_PREDICTED		0.000227273	0.001603306	*Zfp575*	*ZNF575*	ZNF575 Homo sapiens zinc finger protein 575 (ZNF575), mRNA.	
ILMN_1353766	RGD1566136_PREDICTED	X	0.000317125	8.26E-05	*Rps9*	*RPS9*	As above	similar to 40S ribosomal protein S9
ILMN_1353839	PER1		0.000317125	0.00030303	*Per1*	*PER1*	PER1 Homo sapiens period homolog 1 (Drosophila) (PER1), mRNA.	
ILMN_1360786	FKBP8	16	0.000317125	0.000125392	*Fkbp8*	*FKBP8*	FKBP8 Homo sapiens FK506 binding protein 8, 38kDa (FKBP8), mRNA.	
ILMN_1363342	SLC1A3	2	0.000618182	0.001018182	*Slc1a3*	*SLC1A3*	SLC1A3 Homo sapiens solute carrier family 1 (glial high affinity glutamate transporter), member 3 (SLC1A3), transcript variant 1, mRNA.	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
ILMN_1359704	LOC307332		0.000371901	0.00161442	*Rnf149*	*RNF149*	As above	similar to goliath-related E3 ubiquitin ligase 4
ILMN_1364120	POLL	1	0.000714286	0.000847107	*Poll*	*POLL*	POLL Homo sapiens polymerase (DNA directed), lambda (POLL), transcript variant 1, mRNA.	
ILMN_1362834	DUSP6	7	0.001124807	0.01020475	*Dusp6*	*DUSP6*	DUSP6 Homo sapiens dual specificity phosphatase 6 (DUSP6), transcript variant 1, mRNA.	
ILMN_1369573	LOC688712	19	0.001792208	0.006753247	*Rpl22*	*RPL22*	RPL22 Homo sapiens ribosomal protein L22 (RPL22), mRNA.	similar to ribosomal protein L22 like 1
ILMN_1375922	NR4A3	5	0.000482375	0.007217069	*Nr4a3*	*NR4A3*	NR4A3 Homo sapiens nuclear receptor subfamily 4, group A, member 3 (NR4A3), transcript variant 1, mRNA.	
ILMN_1349793	LOC684139		0.000651801	0.002306649		*ZNF582*	ZNF582 Homo sapiens zinc finger protein 582 (ZNF582), mRNA.	similar to zinc finger protein 582
ILMN_1358205	RGD1560975_PREDICTED		0.008009404	0.028941878	*Tuba1b*	*TUBA1B*	TUBA1B Homo sapiens tubulin, alpha 1b (TUBA1B), mRNA.	similar to Tubulin alpha-2 chain
ILMN_1360210	LOC499068		0.001805378	0.01180303	*Zfp583*	*ZNF583*	ZNF583 Homo sapiens zinc finger protein 583 (ZNF583), transcript variant 1, mRNA.	similar to Zfp583 protein
ILMN_1376530	RT1-A3	20	0.000861244	0		*HLA-B*	HLA-B (major histocompatibility complex, class I, B)	RT1-A1 is HLA-B
ILMN_1369562	LOC499096		0.00119697	0.007554545		*ZNF45*	ZNF45 Homo sapiens zinc finger protein 45 (ZNF45), mRNA.	similar to Zinc finger protein 45
ILMN_1368305	LOC499613		0.001124807	0.000277778	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1373217	ADPGK		0.001633729	0.000847107	*Adpgk*	*ADPGK*	ADPGK Homo sapiens ADP-dependent glucokinase (ADPGK), transcript variant 1, mRNA.	
ILMN_1355124	GALNT2_PREDICTED		0.000714286	0.008354978	*Galnt2*	*GALNT2*	GALNT2 Homo sapiens UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2) (GALNT2), mRNA.	
ILMN_1368809	BTG2	13	0.01057352	0.013769231	*Btg2*	*BTG2*	BTG2 Homo sapiens BTG family, member 2 (BTG2), mRNA.	
ILMN_1364113	CTGF	1	0.002249417	0.004427391	*Ctgf*	*CTGF*	CTGF Homo sapiens connective tissue growth factor (CTGF), mRNA.	
ILMN_1367428	ZFP189_PREDICTED		0.004554637	0.013664773	*Zfp189*	*ZNF189*	ZNF189 Homo sapiens zinc finger protein 189 (ZNF189), transcript variant 1, mRNA.	
ILMN_1370045	TRAPPC2	X	0.002683983	0.008354978	*Trappc2*	*TRAPPC2*	TRAPPC2 Homo sapiens trafficking protein particle complex 2 (TRAPPC2), transcript variant 1, mRNA.	
ILMN_1364821	LOC500720		0.009307057	0.028941878	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1350533	RGD1563551_PREDICTED	4	0.017192118	0.00161442	*Rpl31*	*RPL31*	RPL31 Homo sapiens ribosomal protein L31 (RPL31), transcript variant 1, mRNA.	
ILMN_1367827	LOC298998		0.003996212	0.005117845	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1352722	LOC316550		0.003839458	0.025268474	*Rab18*	*RAB18*	RAB18 Homo sapiens RAB18, member RAS oncogene family (RAB18), transcript variant 1, mRNA.	similar to Rab18
ILMN_1349422	PTGS2	13	0.004818182	0.016441558	*Ptgs2*	*PTGS2*	PTGS2 Homo sapiens prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) (PTGS2), mRNA.	
ILMN_1359441	PLA2G2A	5	0.009950187	0.024345238	*Pla2g2a*	*PLA2G2A*	PLA2G2A Homo sapiens phospholipase A2, group IIA (platelets, synovial fluid) (PLA2G2A), transcript variant 1, mRNA.	
ILMN_1357461	ZFP61	1	0.003237998	0.002540107	No Human Homologue	No Human Homologue	No Human Homologue	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
ILMN_1368780	CLIC2	20	0.004415584	0.000847107	*Clic2*	*CLIC2*	CLIC2 Homo sapiens chloride intracellular channel 2 (CLIC2), mRNA.	
ILMN_1359630	LOC679663	16|NW_047479.1	0.008343109	0.045305577	*Tceb1*	*TCEB1*	TCEB1 Homo sapiens transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C) (TCEB1), transcript variant 1, mRNA.	similar to transcription elongation factor B (SIII), polypeptide 1
ILMN_1372466	ZCCHC9	2	0.008783107	0.001607143	*Zcchc9*	*ZCCHC9*	ZCCHC9 Homo sapiens zinc finger, CCHC domain containing 9 (ZCCHC9), transcript variant 1, mRNA.	
ILMN_1353304	RGD1561287_PREDICTED		0.009307057	0.004159402	*Fam32a*	*FAM32A*	FAM32A Homo sapiens family with sequence similarity 32, member A (FAM32A), mRNA.	family with sequence similarity 32, member A
ILMN_1374612	TM9SF4_PREDICTED		0.01004329	0.023941242	*Tm9sf4*	*TM9SF4*	TM9SF4 Homo sapiens transmembrane 9 superfamily protein member 4 (TM9SF4), mRNA.	
ILMN_1368506	LOC497770		0.012253444	0.021175449	*Scn3a*	*SCN3A*	SCN3A Homo sapiens sodium channel, voltage-gated, type III, alpha subunit (SCN3A), transcript variant 1, mRNA.	hypothetical protein XP_579373 , - change gene name
ILMN_1362561	LOC498378		0.023521336	0.029085968	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1359795	LOC499555		0.009918495	0.024214876	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1356628	NFKBIA	6	0.01039312	0.046209617	*Nfkbia*	*NFKBIA*	NFKBIA Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (NFKBIA), mRNA.	
ILMN_1374199	GIOT1	7	0.006464646	0.013769231		*ZNF461*	ZNF461 Homo sapiens zinc finger protein 461 (ZNF461), mRNA.	GIOT1 is ZNF461
ILMN_1353935	FARP1_PREDICTED		0.009307057	0.027335423	*Farp1*	*FARP1*	FARP1 Homo sapiens FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived) (FARP1), transcript variant 1, mRNA.	
ILMN_1362409	BAI2_PREDICTED	5	0.008701299	0.002714097	*Bai2*	*BAI2*	BAI2 Homo sapiens brain-specific angiogenesis inhibitor 2 (BAI2), mRNA.	
ILMN_1370369	EGR2	20	0.009307057	0.003713188	*Egr2*	*EGR2*	EGR2 Homo sapiens early growth response 2 (EGR2), transcript variant 1, mRNA.	
ILMN_1372919	CYR61	2	0.003996212	0.006986166	*Cyr61*	*CYR61*	CYR61 Homo sapiens cysteine-rich, angiogenic inducer, 61 (CYR61), mRNA.	
ILMN_1355401	RGD1563543_PREDICTED		0.021554002	0.0188	*Rpl31*	*RPL31*	As above	
ILMN_1362451	RGS2	13	0.007420147	0.0072147	*Rgs2*	*RGS2*	RGS2 Homo sapiens regulator of G-protein signaling 2, 24kDa (RGS2), mRNA.	
ILMN_1365095	DHX40	10	0.011252914	0.030909091	*Dhx40*	*DHX40*	DHX40 Homo sapiens DEAH (Asp-Glu-Ala-His) box polypeptide 40 (DHX40), transcript variant 1, mRNA.	
ILMN_1376434	PGRMC1	X	0.008343109	0.033276328	*Pgrmc1*	*PGRMC1*	PGRMC1 Homo sapiens progesterone receptor membrane component 1 (PGRMC1), mRNA.	
ILMN_1349269	SGK	1	0.011998871	0.002164502	*Sgk1*	*SGK1*	SGK1 Homo sapiens serum/glucocorticoid regulated kinase 1 (SGK1), transcript variant 1, mRNA.	
ILMN_1361423	RKHD2_PREDICTED		0.011998871	0.036649595	*Mex3c*	*MEX3C*	MEX3C Homo sapiens mex-3 homolog C (C. elegans) (MEX3C), mRNA.	RKDHD2 is MEX3C
ILMN_1649981	STRN3	6	0.01487781	0.042800325	*Strn3*	*STRN3*	STRN3 Homo sapiens striatin, calmodulin binding protein 3 (STRN3), transcript variant 1, mRNA.	
ILMN_1361139	TMPRSS8	10	0.015932282	0.033140909	*Prss30*	*PRSS30P*	PRSS30P Homo sapiens protease, serine, 30 homolog (mouse), pseudogene (PRSS30P), non-coding RNA	TMPRSS8 is PRSS30P
ILMN_1357368	LOC497841		0.03056229	0.030909091	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1356949	COL6A1_PREDICTED	20	0.01375383	0.006753247	*Col6a1*	*COL6A1*	COL6A1 Homo sapiens collagen, type VI, alpha 1 (COL6A1), mRNA.	
ILMN_1368116	LOC367398		0.017802335	0.015773059	*Rpl17*	*RPL17*	RPL17 Homo sapiens ribosomal protein L17 (RPL17), transcript variant 1, mRNA.	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
ILMN_1359375	LOC499418	20	0.012542188	0.013913949	*Spatc1l*	*C21orf56*	C21orf56 Homo sapiens chromosome 21 open reading frame 56 (C21orf56), transcript variant 1, mRNA.	
ILMN_1352529	IER2	19	0.01724612	0.043680416	*Ier2*	*IER2*	IER2 Homo sapiens immediate early response 2 (IER2), mRNA.	
ILMN_1354120	LOC691762	Un|NW_047990.1	0.012963959	0.027361039	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1373383	TIPARP_PREDICTED		0.009131615	0.004731935		*TIPARP*	TIPARP Homo sapiens TCDD-inducible poly(ADP-ribose) polymerase (TIPARP), transcript variant 1, mRNA.	
ILMN_1364683	NECAB2		0.017802335	0.002306649	*Necab2*	*NECAB2*	NECAB2 Homo sapiens N-terminal EF-hand calcium binding protein 2 (NECAB2), mRNA.	
ILMN_1360868	RNF40	1	0.021361686	0.015614973	*Rnf40*	*RNF40*	RNF40 Homo sapiens ring finger protein 40 (RNF40), transcript variant 1, mRNA.	
ILMN_1373434	RAB28	14	0.01620753	0.043683386	*Rab28*	*RAB28*	RAB28 Homo sapiens RAB28, member RAS oncogene family (RAB28), transcript variant 1, mRNA.	
ILMN_1354535	ZNF386	6	0.014258893	0.029966683	*Zfp386*	*ZNF519*	ZNF519 Homo sapiens zinc finger protein 519 (ZNF519), transcript variant 1, mRNA.	ZNF386 is ZNF519
ILMN_1374435	C1GALT1C1	X	0.024523282	0.030423928	*C1galt1c1*	*C1GALT1C1*	C1GALT1C1 Homo sapiens C1GALT1-specific chaperone 1 (C1GALT1C1), transcript variant 1, mRNA.	
ILMN_1362029	LOC502490		0.04383255	0.022875929	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1351127	PLCL1	9	0.019282511	0.020445748	*Plcl1*	*PLCL1*	PLCL1 Homo sapiens phospholipase C-like 1 (PLCL1), mRNA.	
ILMN_1361932	MLL5	4	0.037223421	0.029966683	*Mll5*	*MLL5*	MLL5 Homo sapiens myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila) (MLL5), transcript variant 1, mRNA.	
ILMN_1369005	EGR1	18	0.035286665	0.010340909	*Egr1*	*EGR1*	EGR1 Homo sapiens early growth response 1 (EGR1), mRNA.	
ILMN_1360758	RGD1308626	15	0.044931617	0.043680416	*Slain1*	*SLAIN1*	SLAIN1 Homo sapiens SLAIN motif family, member 1 (SLAIN1), transcript variant 1, mRNA.	similar to SLAIN motif-containing protein 1
ILMN_1359043	EGR4	4	0.030741362	0.027467899	*Egr4*	*EGR4*	EGR4 Homo sapiens early growth response 4 (EGR4), mRNA.	
ILMN_1368493	RGD1562629_PREDICTED		0.038248485	0.034655248		*NBEA*	NBEA Homo sapiens neurobeachin (NBEA), transcript variant 1, mRNA.	similar to neurobeachin
ILMN_1349772	FTCD	20	0.039682259	0.009592184	*Ftcd*	*FTCD*	FTCD Homo sapiens formiminotransferase cyclodeaminase (FTCD), transcript variant A, mRNA.	
ILMN_1361370	RYBP_PREDICTED	4	0.017995019	0.023480201		*RYBP*	RYBP Homo sapiens RING1 and YY1 binding protein (RYBP), mRNA.	
ILMN_1373798	PLCB1	3	0.024981672	0.020053129	*Plcb1*	*PLCB1*	PLCB1 Homo sapiens phospholipase C, beta 1 (phosphoinositide-specific) (PLCB1), transcript variant 1, mRNA.	
ILMN_1372236	ZFP36L1	6	0.038386514	0.040504587	*Zfp36l1*	*ZFP36L1*	ZFP36L1 Homo sapiens zinc finger protein 36, C3H type-like 1 (ZFP36L1), transcript variant 1, mRNA.	
ILMN_1349546	MDGA2	6	0.047363636	0.017687075	*Mdga2*	*MDGA2*	MDGA2 Homo sapiens MAM domain containing glycosylphosphatidylinositol anchor 2 (MDGA2), transcript variant 1, mRNA.	
ILMN_1353576	GPR149	2	0.04383255	0.024593868	*Gpr149*	*GPR149*	GPR149 Homo sapiens G protein-coupled receptor 149 (GPR149), mRNA.	
ILMN_1354065	LOC686053	Un|NW_047874.1	0.042841958	0.014360902	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1357903	RGD1564171_PREDICTED	2	0.028060375	0.033276328	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1361346	RGD1564940_PREDICTED	5	0.031886608	0.033533363	Uncharacterised	Uncharacterised	Uncharacterised	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
ILMN_1370101	LOC499058		0.046880878	0.022875929	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1352135	CEACAM10	1	0.045470541	0.003838384	*Ceacam1*	*CEACAM1*	CEACAM1 Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), transcript variant 4, mRNA.	CEACAM10 not in RefSeq but overlaps with CEACAM1
ILMN_1371662	PPP2R1A	1	0.048310249	0.024218835	*Ppp2r1a*	*PPP2R1A*	PPP2R1A Homo sapiens protein phosphatase 2, regulatory subunit A, alpha (PPP2R1A), transcript variant 1, mRNA.	
ILMN_1363414	LOC365025		0.043867069	0.005140693	*Tuba1b*	*TUBA1B*	As above	similar to alpha tubulin subunit
ILMN_1370033	LOC498604		0.037413057	0.04984	*Fam129b*	*FAM129B*	FAM129B Homo sapiens family with sequence similarity 129, member B (FAM129B), transcript variant 1, mRNA.	similar to B-cell novel protein 1
ILMN_1359027	RGD1563482_PREDICTED	12	0.047780599	0.019192584	*RGD1563482*	*C12orf65*	C12orf65 Homo sapiens chromosome 12 open reading frame 65 (C12orf65), transcript variant 1, mRNA.	similar to hypothetical protein FLJ38663
ILMN_1350094	SDPR	9	0.019282511	0.011280632	*Sdpr*	*SDPR*	SDPR Homo sapiens serum deprivation response (SDPR), mRNA.	
ILMN_2039346	HLA-DMA	20	0.017802335	0.00802139	*RT1-DMa*	*HLA-DMA*	HLA-DMA Homo sapiens major histocompatibility complex, class II, DM alpha (HLA-DMA), mRNA.	
ILMN_1349530	CRSP6		0.02584	0.007749775	*Med17*	*MED17*	MED17 Homo sapiens mediator complex subunit 17 (MED17), mRNA.	CRSP6 = MED17
ILMN_1369447	RGD1565673_PREDICTED	3	0.015577014	0.043683386	*Ttf1*	*TTF1*	TTF1 Homo sapiens transcription termination factor, RNA polymerase I (TTF1), transcript variant 1, mRNA. (NKX2-1 Homo sapiens NK2 homeobox 1 (NKX2-1), transcript variant 1, mRNA.)	
ILMN_1355984	TPX2_PREDICTED		0.036723549	0.016383399	*Tpx2*	*TPX2*	TPX2 Homo sapiens TPX2, microtubule-associated, homolog (Xenopus laevis) (TPX2), mRNA.	
ILMN_1350525	XPNPEP1	1	0.011273885	0.036497199	*Xpnpep1*	*XPNPEP1*	XPNPEP1 Homo sapiens X-prolyl aminopeptidase (aminopeptidase P) 1, soluble (XPNPEP1), transcript variant 1, mRNA.	
ILMN_1360785	LOC361929	2	0.01569697	0.004775604	*Atp11b*	*ATP11B*	ATP11B Homo sapiens ATPase, class VI, type 11B (ATP11B), mRNA.	similar to Potential phospholipid-transporting ATPase IF
ILMN_1651182	ANKRD15	1	0.013436679	0.029966683	*Kank1*	*KANK1*	KANK1 Homo sapiens KN motif and ankyrin repeat domains 1 (KANK1), transcript variant 1, mRNA.	ANKRD15 = KANK1
ILMN_1360040	NUDT14_PREDICTED		0.007355372	0.01180303	*Nudt14*	*NUDT14*	NUDT14 Homo sapiens nudix (nucleoside diphosphate linked moiety X)-type motif 14 (NUDT14), mRNA.	
ILMN_1354941	COL3A1	9	0.008265852	0.007226814		*COL3A1*	COL3A1 Homo sapiens collagen, type III, alpha 1 (COL3A1), mRNA.	
ILMN_1374043	ERAF_PREDICTED		0.011252914	0.012742299	*Ahsp*	*AHSP*	AHSP Homo sapiens alpha hemoglobin stabilizing protein (AHSP), mRNA.	
ILMN_1351156	GNAI1	4	0.038758971	0.006236786	*Gnai1*	*GNAI1*	GNAI1 Homo sapiens guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1 (GNAI1), transcript variant 1, mRNA.	
ILMN_1352524	USMG5	1	0.012053872	0.004427391	*Usmg5*	*USMG5*	USMG5 Homo sapiens up-regulated during skeletal muscle growth 5 homolog (mouse) (USMG5), transcript variant 1, mRNA.	
ILMN_1374825	LOC294789		0.00608658	0.029223587	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1369541	TCF4	18	0.011252914	0.000871212	*Tcf7l2*	*TCF7L2*	TCF7L2 Homo sapiens transcription factor 7-like 2 (T-cell specific, HMG-box) (TCF7L2), transcript variant 1, mRNA. (TCF4 Homo sapiens transcription factor 4 (TCF4), transcript variant 1, mRNA.)	
ILMN_1357413	LOC360443		0.007355372	0.004427391	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1650955	IFI27L		0.028212577	0.01020475		*IFI27L2*	IFI27L2 Homo sapiens interferon, alpha-inducible protein 27-like 2 (IFI27L2), mRNA.	
ILMN_1375028	HTATIP2_PREDICTED		0.023521336	0.036142857	*Htatip2*	*HTATIP2*	HTATIP2 Homo sapiens HIV-1 Tat interactive protein 2, 30kDa (HTATIP2), transcript variant 1, mRNA.	
ILMN_1348843	SLC17A6	1	0.009307057	0.016441558	*Slc17a6*	*SLC17A6*	SLC17A6 Homo sapiens solute carrier family 17 (sodium-dependent	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
							inorganic phosphate cotransporter), member 6 (SLC17A6), mRNA.	
ILMN_1361722	RBM8_PREDICTED		0.011111111	0.001194296	*Rbm8a*	*RBM8A*	RBM8A Homo sapiens RNA binding motif protein 8A (RBM8A), mRNA.	
ILMN_1359627	LOC360919	14	0.002249417	0.01020475	*Afp*	*AFP*	AFP Homo sapiens alpha-fetoprotein (AFP), mRNA.	Similar to alpha-fetoprotein
ILMN_1358978	RGD1306126	10	0.002584885	0.007217069	*Fam173a*	*FAM173A*	FAM173A Homo sapiens family with sequence similarity 173, member A (FAM173A), mRNA.	
ILMN_1350985	INPP5D	9	0.002191781	0.004427391	*Inpp5d*	*INPP5D*	INPP5D Homo sapiens inositol polyphosphate-5-phosphatase, 145kDa (INPP5D), transcript variant 1, mRNA.	
ILMN_1367329	PPP1R16A_PREDICTED	7	0.002684492	0.00848199	*Ppp1r16a*	*PPP1R16A*	PPP1R16A Homo sapiens protein phosphatase 1, regulatory subunit 16A (PPP1R16A), mRNA.	
ILMN_1361915	PDE10A	1	0.001931818	0.008296558	*Pde10a*	*PDE10A*	PDE10A Homo sapiens phosphodiesterase 10A (PDE10A), transcript variant 1, mRNA.	
ILMN_1365885	MMP14	15	0.002472284	0.043683386	*Mmp14*	*MMP14*	MMP14 Homo sapiens matrix metallopeptidase 14 (membrane-inserted) (MMP14), mRNA.	
ILMN_1371064	RPS18	20	0.002236842	0.01076555	*Rps18*	*RPS18*	RPS18 Homo sapiens ribosomal protein S18 (RPS18), mRNA.	
ILMN_1372988	FHL2	9	0.002472284	0.007204301	*Fhl2*	*FHL2*	FHL2 Homo sapiens four and a half LIM domains 2 (FHL2), transcript variant 1, mRNA.	
ILMN_1651148	POLR2I_PREDICTED	1	0.0015427	0.001276224	*Polr2i*	*POLR2I*	POLR2I Homo sapiens polymerase (RNA) II (DNA directed) polypeptide I, 14.5kDa (POLR2I), mRNA.	
ILMN_1350743	PRMT5_PREDICTED	15	0.001505682	0.002164502	*Prmt5*	*PRMT5*	PRMT5 Homo sapiens protein arginine methyltransferase 5 (PRMT5), transcript variant 1, mRNA.	
ILMN_1350481	LOC499790		0.000482375	0.014687924	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1352269	RT1-149		0.0015427	0.016747759	No Human Homologue	No Human Homologue	No Human Homologue	
ILMN_1362726	HMGN3	8	0.001564171	0.011285266	*Hmgn3*	*HMGN3*	HMGN3 Homo sapiens high mobility group nucleosomal binding domain 3 (HMGN3), transcript variant 1, mRNA.	
ILMN_1357163	SYMPK		0.001251863	0.000868687	*Sympk*	*SYMPK*	SYMPK Homo sapiens symplekin (SYMPK), mRNA.	
ILMN_1357432	CYP11B1	7	0.000641711	0.003168831	*Cyp11b1*	*CYP11B1*	CYP11B1 Homo sapiens cytochrome P450, family 11, subfamily B, polypeptide 1 (CYP11B1), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.	
ILMN_1359619	LYZL4_PREDICTED	8	0.001564171	0.049642205	*Lyzl4*	*LYZL4*	LYZL4 Homo sapiens lysozyme-like 4 (LYZL4), mRNA.	
ILMN_1360418	RGD1302996	20	0.000714286	0.001320755	*RGD1302996*	*C6orf136*	C6orf136 Homo sapiens chromosome 6 open reading frame 136 (C6orf136), transcript variant 3, mRNA.	uncharacterized protein C6orf136 homolog
ILMN_1649821	HAGHL		0.00040619	0.001603306		*HAGHL*	HAGHL Homo sapiens hydroxyacylglutathione hydrolase-like (HAGHL), transcript variant 2, mRNA.	
ILMN_1349624	KIF5C_PREDICTED		0.002223587	0.002306649	*Kif5c*	*KIF5C*	KIF5C Homo sapiens kinesin family member 5C (KIF5C), mRNA.	
ILMN_1358541	IGFBP6	7	0.000117302	0.002201705	*Igfbp6*	*IGFBP6*	IGFBP6 Homo sapiens insulin-like growth factor binding protein 6 (IGFBP6), mRNA.	
ILMN_1363262	CSNK2A1		0.000395257	0.000187166	*Csnk2a1*	*CSNK2A1*	CSNK2A1 Homo sapiens casein kinase 2, alpha 1 polypeptide (CSNK2A1), transcript variant 1, mRNA.	
ILMN_1359301	IGF2		0.000181818	0.011375291	*Igf2*	*IGF2*	IGF2 Homo sapiens insulin-like growth factor 2 (somatomedin A) (IGF2), transcript variant 1, mRNA.	
ILMN_1358234	RGD1562351_PREDICTED		0.000142045	0.000233766	Uncharacterised	Uncharacterised	Uncharacterised	

**SHRSP versus WKY differential gene expression at five weeks of age.					*** Human equivalent of SHRSP differentially expressed genes			
ILMN_1364214	LOC497864		0.000117302	0.000170455	Uncharacterised	Uncharacterised	Uncharacterised	
ILMN_1351068	RGD1309829_PREDICTED		7.27E-05	9.74E-05	*Ndufaf5*	*NDUFAF5*	NDUFAF5 Homo sapiens NADH dehydrogenase (ubiquinone) complex I, assembly factor 5 (NDUFAF5), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.	
ILMN_1368752	LOC499378		7.27E-05	9.74E-05	Uncharacterised	Uncharacterised	Uncharacterised	similar to oocyte-testis gene 1 (I can't find a record of this gene)
ILMN_1351487	RT1-A1	20	0	9.74E-05	*RT1-CE7*	*HLA-B*	As above	RT1-A1 is HLA-B
ILMN_1353260	PXMP4	3	0	4.33E-05	*Pxmp4*	*PXMP4*	PXMP4 Homo sapiens peroxisomal membrane protein 4, 24kDa (PXMP4), transcript variant 1, mRNA.	
ILMN_1353156	COLQ		0	9.74E-05	*Colq*	*COLQ*	COLQ Homo sapiens collagen-like tail subunit (single strand of homotrimer) of asymmetric acetylcholinesterase (COLQ), transcript variant I, mRNA.	
ILMN_1651096	RGD1560364_PREDICTED		0	9.74E-05		*VPS13C*	VPS13C Homo sapiens vacuolar protein sorting 13 homolog C (S. cerevisiae) (VPS13C), transcript variant 1A, mRNA.	
ILMN_2038792	ALB	14	0	0.000871212	*Alb*	*ALB*	ALB Homo sapiens albumin (ALB), mRNA.	
ILMN_1650062	RGD1563903_PREDICTED		0	0		Uncharacterised	Uncharacterised	
ILMN_1370031	LOC362068		0	0	*Gpr98*	*GPR98*	GPR98 Homo sapiens G protein-coupled receptor 98 (GPR98), transcript variant 1, mRNA.	
ILMN_1371684	LOC499103		0	0		Uncharacterised	Uncharacterised	
ILMN_1358480	LOC365566		0	0		*UBE2SP2*	ZNF286B Homo sapiens zinc finger protein 286B	similar to Ubiquitin-conjugating enzyme E2S
ILMN_1376663	LOC287167	10	0	0	*Hba1*	*HBA1*	HBA1 Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA.	HBA2 = HBA1
ILMN_1349205	RGD1562905_PREDICTED	1	0	0	*Rpl17*	*RPL17*	As above	similar to 60S ribosomal protein L17 (L23) (predicted)
ILMN_1359180	MRPL18_PREDICTED	1	0	0	*Mrpl18*	*MRPL18*	MRPL18 Homo sapiens mitochondrial ribosomal protein L18 (MRPL18), nuclear gene encoding mitochondrial protein, mRNA.	

Table III All candidate SHRSP gene-based results from LBC1936 and CHARGE with WMH variables. Chr is chromosome. nSNPs is the number of SNPs in the gene (+/- 50kb). Please note that the gene boundaries are overlapping as SNPs can be allocated to multiple genes, so the same SNP could be driving the signal in different genes. The results are ordered by significance.

				Discovery: LBC1936						Replication: CHARGE	
				Genotyped SNPs			Imputed SNPs			Imputed SNPs	
					WMH Volume	Fazekas Score		WMH Volume	Fazekas Score		WMH Volume
Chr	Gene	Start Position	Stop Position	nSNPs	Pvalue	Pvalue	nSNPs	Pvalue	Pvalue	nSNPs	Pvalue
4	*AFP*	74,520,796	74,540,356	13	***0.0021***	***0.0009***	77	***0.0037***	***0.0037***	67	*0.83*
4	*ALB*	74,488,869	74,505,834	11	***0.0027***	***0.0017***	61	***0.0063***	***0.0068***	53	*0.75*
7	*GNAI1*	79,602,075	79,686,661	42	***0.034***	***0.033***	181	***0.014***	***0.015***	166	*0.79*
1	*RBM8A*	144,218,994	144,222,801	13	***0.038***	*0.057*	26	***0.029***	***0.024***	21	*0.54*
2	*INPP5D*	233,633,279	233,781,288	69	***0.041***	*0.78*	198	***0.044***	*0.868*	162	*0.98*
10	*XPNPEP1*	111,614,513	111,673,192	18	***0.042***	*0.14*	130	*0.145*	*0.23*	120	***0.000101***
9	*NR4A3*	101,623,957	101,668,994	13	***0.045***	*0.16*	62	*0.106*	*0.245*	56	*0.48*
13	*FARP1*	97,593,434	97,900,024	154	***0.049***	*0.25*	550	*0.18*	*0.507*	468	***0.025***
18	*RPL17*	45,268,853	45,272,904	16	*0.058*	*0.13*	79	*0.076*	*0.218*	74	*0.28*
6	*MRPL18*	160,131,481	160,139,451	24	*0.059*	***0.039***	89	*0.16*	***0.048***	76	*0.24*
4	*RAB28*	12,978,479	13,095,054	11	*0.076*	*0.49*	84	*0.264*	*0.61*	40	*0.18*
8	*CYP11B1*	143,950,774	143,958,238	30	*0.080*	*0.49*	103	*0.15*	*0.613*	87	*0.13*
2	*SDPR*	192,407,280	192,420,226	15	*0.081*	*0.10*	68	*0.24*	*0.314*	55	*0.25*
1	*SIPA1L2*	230,600,334	230,717,866	80	*0.087*	***0.00934***	340	*0.20*	***0.018***	285	*0.87*
6	*PDE10A*	165,664,524	165,995,575	159	*0.094*	*0.065*	594	*0.15*	*0.168*	534	*0.57*
13	*SLAIN1*	77,170,470	77,236,378	26	*0.10*	*0.30*	137	*0.19*	*0.341*	124	*0.51*
2	*KIF5C*	149,349,288	149,591,519	31	*0.10*	*0.66*	148	*0.14*	*0.711*	128	*0.35*
12	*IGFBP6*	51,777,702	51,782,395	16	*0.13*	*0.08*	67	*0.17*	*0.065*	60	*0.24*
21	*COL6A1*	46,226,090	46,249,391	27	*0.14*	*0.28*	96	*0.080*	*0.20*	93	*0.66*
2	*FHL2*	105,343,714	105,421,392	47	*0.15*	*0.56*	179	*0.11*	*0.55*	167	*0.37*
5	*SLC1A3*	36,642,213	36,724,193	71	*0.15*	*0.45*	211	*0.27*	*0.74*	186	*0.16*
19	*RPS16*	44,615,686	44,618,458	20	*0.16*	*0.43*	60	*0.38*	*0.69*	53	*0.61*

2	*SCN3A*	165,652,275	165,768,823	33	*0.17*	*0.83*	135	*0.11*	*0.88*	124	*0.90*
21	*C21orf56*	46,405,497	46,428,729	38	*0.17*	*0.12*	113	*0.070*	***0.043***	100	*0.36*
16	*RNF40*	30,681,130	30,694,039	5	*0.17*	*0.58*	11	*0.21*	*0.66*	11	*0.30*
14	*FOS*	74,815,283	74,818,665	28	*0.17*	*0.37*	78	*0.58*	*0.46*	56	*0.14*
1	*BTG2*	201,541,286	201,545,352	19	*0.18*	*0.33*	98	*0.37*	*0.55*	81	*0.93*
20	*TM9SF4*	30,160,969	30,218,722	16	*0.20*	*0.29*	86	*0.28*	*0.17*	72	*0.44*
9	*TTF1*	134,240,757	134,272,042	37	*0.20*	*0.23*	96	*0.61*	*0.40*	86	*0.42*
1	*CYR61*	85,819,047	85,821,978	32	*0.20*	*0.24*	107	*0.28*	*0.45*	84	*0.73*
14	*NUDT14*	104,710,320	104,718,685	19	*0.21*	*0.088*	67	*0.28*	*0.10*	56	***0.039***
19	*RPS9*	59,396,537	59,403,327	22	*0.21*	*0.77*	115	***0.011***	*0.49*	61	*0.051*
2	*RPL31*	100,985,122	101,002,587	28	*0.21*	*0.27*	129	*0.066*	*0.16*	124	*0.11*
12	*DUSP6*	88,265,967	88,270,427	20	*0.22*	*0.16*	86	*0.35*	*0.18*	75	*0.12*
11	*SLC17A6*	22,316,242	22,357,619	39	*0.23*	*0.78*	159	*0.12*	*0.86*	150	*0.90*
3	*TIPARP*	157,875,072	157,907,228	24	*0.29*	*0.92*	144	*0.30*	*0.47*	81	*0.69*
15	*VPS13C*	59,931,881	60,139,939	56	*0.30*	*0.76*	341	*0.57*	*0.80*	290	*0.72*
6	*CTGF*	132,311,009	132,314,211	33	*0.30*	*0.16*	138	*0.78*	*0.48*	129	*0.14*
6	*HLA-DMA*	33,024,372	33,028,831	95	*0.31*	*0.54*	176	*0.41*	*0.73*	151	*0.94*
6	*C6orf136*	30,722,779	30,728,961	38	*0.32*	*0.92*	61	*0.39*	*0.94*	27	*0.56*
4	*GUCY1A3*	156,807,327	156,871,226	55	*0.33*	*0.49*	147	*0.51*	*0.49*	114	*0.29*
16	*ZNF597*	3,426,110	3,433,491	19	*0.33*	*0.70*	83	*0.20*	*0.62*	80	*0.45*
16	*HBA1*	166,678	167,520	16	*0.34*	*0.26*	27	*0.37*	*0.34*	23	*0.24*
5	*EGR1*	137,829,079	137,832,903	16	*0.34*	*0.91*	42	*0.62*	*0.94*	33	***0.031***
19	*ZNF575*	48,729,168	48,732,124	18	*0.34*	*0.080*	69	*0.232*	*0.089*	62	*0.25*
17	*DHX40*	54,997,667	55,040,484	10	*0.36*	*0.19*	61	*0.55*	*0.27*	59	*0.17*
17	*PER1*	7,984,512	7,996,478	27	*0.38*	*0.49*	77	*0.55*	*0.57*	64	*0.14*
20	*C20orf7*	13,713,681	13,745,874	28	*0.39*	*0.11*	147	*0.83*	*0.17*	130	***0.024***
5	*ZCCHC9*	80,633,177	80,644,872	22	*0.39*	*0.14*	81	*0.57*	*0.30*	74	*0.92*
9	*KANK1*	494,702	736,103	169	*0.40*	*0.19*	613	*0.52*	*0.43*	580	*0.069*
18	*MEX3C*	46,954,917	46,977,688	15	*0.41*	*0.22*	78	*0.38*	*0.31*	76	*0.51*
10	*TCF7L2*	114,699,998	114,916,060	49	*0.42*	*0.73*	166	*0.79*	*0.87*	145	*0.12*
20	*PXMP4*	31,754,210	31,771,797	9	*0.43*	*0.44*	34	*0.40*	*0.59*	32	*0.38*
12	*C12orf65*	122,283,415	122,308,459	4	*0.43*	*0.73*	64	*0.62*	*0.39*	51	*0.11*

1	*PLA2G2A*	20,174,517	20,179,496	26	*0.44*	*0.46*	75	*0.29*	*0.54*	68	*0.31*
16	*NECAB2*	82,559,737	82,593,880	54	*0.44*	*0.48*	259	*0.71*	*0.79*	215	*0.30*
19	*FKBP8*	18,503,567	18,515,383	9	*0.44*	*0.74*	49	*0.37*	*0.87*	44	*0.57*
19	*ZNF582*	61,586,459	61,596,701	22	*0.46*	*0.22*	82	*0.27*	*0.27*	71	*0.22*
3	*RYBP*	72,506,438	72,578,464	29	*0.46*	*0.52*	128	*0.49*	*0.50*	122	*0.57*
14	*STRN3*	30,432,755	30,565,358	39	*0.48*	*0.44*	166	*0.44*	*0.46*	131	*0.12*
19	*SYMPK*	51,010,539	51,058,388	23	*0.48*	*0.64*	79	*0.36*	*0.36*	78	*0.24*
14	*PRMT5*	22,459,572	22,468,501	17	*0.51*	*0.99*	68	*0.58*	*1.00*	61	*0.46*
15	*ADPGK*	70,830,760	70,863,179	19	*0.51*	*0.76*	73	*0.49*	*0.71*	73	*0.089*
16	*FAM173A*	711,158	712,591	16	*0.51*	*0.68*	39	*0.24*	*0.70*	32	*0.91*
21	*FTCD*	46,380,603	46,399,909	31	*0.52*	*0.33*	104	*0.32*	*0.19*	87	*0.47*
6	*HLA-B*	31,429,627	31,432,968	84	*0.52*	*0.67*	383	*0.45*	*0.53*	353	*0.77*
12	*TUBA1B*	47,807,832	47,811,571	3	*0.52*	*0.71*	30	*0.16*	*0.47*	29	*0.21*
19	*POLR2I*	41,296,450	41,298,046	12	*0.53*	*0.80*	42	*0.22*	*0.87*	32	*0.47*
6	*HMGN3*	79,967,680	80,001,174	29	*0.54*	*0.44*	124	*0.55*	*0.35*	84	*0.65*
3	*GPR149*	155,538,154	155,630,198	32	*0.54*	*0.50*	167	*0.68*	*0.59*	156	*0.23*
8	*TCEB1*	75,021,187	75,046,900	24	*0.54*	*0.79*	91	*0.50*	*0.89*	82	*0.81*
16	*HAGHL*	716,958	719,716	15	*0.55*	*0.45*	42	*0.31*	*0.63*	31	*0.90*
2	*COL3A1*	189,547,343	189,585,717	27	*0.56*	*0.18*	143	*0.62*	*0.17*	124	*0.83*
2	*EGR4*	73,371,564	73,374,181	8	*0.57*	*0.43*	43	*0.72*	*0.22*	40	*0.39*
16	*ERAF*	31,446,703	31,447,625	11	*0.57*	*0.75*	48	*0.31*	*0.53*	47	*0.71*
14	*NFKBIA*	34,940,466	34,943,711	23	*0.57*	*0.65*	95	*0.71*	*0.76*	90	*0.65*
19	*ZNF583*	61,607,529	61,628,212	28	*0.58*	*0.41*	88	*0.32*	*0.42*	78	*0.33*
6	*SGK1*	134,532,076	134,537,727	21	*0.59*	*0.79*	78	*0.78*	*0.87*	67	*0.62*
9	*FAM129B*	129,307,438	129,381,089	31	*0.59*	*0.37*	119	*0.92*	*0.42*	113	*1.00*
2	*PLCL1*	198,377,777	198,721,365	49	*0.60*	*0.99*	270	*0.86*	*0.99*	256	*0.93*
11	*MED17*	93,157,052	93,186,144	15	*0.60*	*0.46*	65	*0.86*	*0.60*	62	***0.0095***
5	*FAM151B*	79,819,555	79,873,962	17	*0.60*	*0.51*	96	*0.83*	*0.72*	74	*0.26*
20	*TPX2*	29,790,564	29,853,264	16	*0.62*	*0.35*	112	*0.60*	*0.37*	108	*0.37*
6	*RPS18*	33,347,829	33,352,259	14	*0.64*	*0.66*	63	*0.81*	*0.72*	58	*0.37*
19	*CEACAM1*	47,703,297	47,724,479	4	*0.64*	*0.55*	29	*0.82*	*0.70*	27	*0.92*
3	*COLQ*	15,466,643	15,538,262	37	*0.65*	*0.38*	145	*0.88*	*0.41*	135	*0.47*

10	*RAB18*	27,833,254	27,869,105	22	*0.66*	*0.70*	103	*0.79*	*0.74*	96	*0.80*
19	*IER2*	13,122,281	13,126,718	15	*0.67*	*0.71*	33	*0.45*	*0.66*	21	*0.19*
3	*LYZL4*	42,413,578	42,427,069	28	*0.68*	*0.72*	138	*0.88*	*0.88*	112	*0.99*
3	*ATP11B*	183,993,984	184,122,115	26	*0.69*	*0.21*	135	*0.89*	*0.37*	134	*0.50*
5	*DUSP1*	172,127,706	172,130,809	34	*0.69*	*0.98*	107	*0.79*	*0.98*	90	*0.29*
11	*HTATIP2*	20,341,806	20,361,905	24	*0.69*	*0.44*	118	*0.86*	*0.53*	112	*0.38*
2	*RNF149*	101,258,984	101,291,584	16	*0.72*	*0.41*	88	*0.77*	*0.47*	85	*0.30*
19	*FAM32A*	16,157,234	16,163,857	16	*0.72*	*0.72*	80	*0.68*	*0.94*	77	*0.17*
19	*ZNF317*	9,112,085	9,135,089	30	*0.72*	*0.96*	74	*0.72*	*0.95*	71	*0.11*
1	*GALNT2*	228,269,650	228,484,569	99	*0.74*	*0.18*	328	*0.89*	*0.45*	284	*0.31*
19	*ZNF566*	41,630,421	41,672,177	16	*0.76*	*0.52*	59	*0.91*	*0.82*	59	*0.26*
14	*MDGA2*	46,378,577	47,213,738	164	*0.76*	*0.63*	1036	*0.75*	*0.73*	967	*0.89*
19	*ZNF45*	49,108,620	49,121,398	16	*0.76*	*0.88*	90	*0.67*	*0.79*	81	*0.061*
1	*RPL22*	6,167,666	6,182,266	19	*0.78*	*0.89*	44	*0.87*	*0.98*	30	*0.46*
1	*PTGS2*	184,907,591	184,916,179	16	*0.79*	*0.44*	75	*0.99*	*0.73*	68	*0.98*
7	*MLL5*	104,441,872	104,541,768	21	*0.79*	*0.39*	98	*0.61*	*0.66*	97	*0.94*
10	*USMG5*	105,138,803	105,146,213	10	*0.79*	*0.88*	31	*0.92*	*0.96*	28	***0.000142***
14	*MMP14*	22,375,632	22,386,643	40	*0.79*	*0.90*	99	*0.87*	*0.97*	75	***0.046***
19	*PPP2R1A*	57,385,045	57,421,483	47	*0.80*	*0.66*	141	*0.94*	*0.84*	132	*0.41*
6	*OOEP*	74,135,000	74,136,236	9	*0.80*	*0.93*	65	*0.76*	*0.92*	55	*0.87*
8	*PPP1R16A*	145,692,916	145,698,312	6	*0.81*	*0.11*	24	*0.78*	*0.25*	20	*0.27*
8	*ARC*	143,689,411	143,692,835	17	*0.81*	*0.94*	81	*0.22*	*0.98*	78	***0.033***
5	*GPR98*	89,890,372	90,495,789	142	*0.82*	*0.75*	657	*0.91*	*0.88*	591	*0.74*
9	*ZNF189*	103,200,983	103,212,763	25	*0.82*	*0.30*	135	*0.72*	*0.28*	128	*0.98*
10	*POLL*	103,328,628	103,337,963	12	*0.83*	*0.90*	79	*0.90*	*0.76*	73	*0.69*
10	*EGR2*	64,241,762	64,246,133	21	*0.84*	*0.93*	100	*0.88*	*0.95*	97	*0.84*
2	*NAB1*	191,222,092	191,265,737	17	*0.85*	*1.00*	88	*0.87*	*1.00*	86	*0.12*
14	*ZFP36L1*	68,324,127	68,329,538	31	*0.86*	*0.47*	87	*0.76*	*0.41*	84	*0.20*
2	*PER2*	238,817,417	238,861,946	18	*0.87*	*0.96*	92	*0.20*	*0.66*	70	*0.37*
14	*FAM14A*	93,663,870	93,665,710	34	*0.87*	*0.98*	96	*0.98*	*1.00*	89	*0.35*
18	*ZNF519*	14,094,723	14,122,429	13	*0.88*	*0.75*	49	*0.87*	*0.83*	41	*0.86*
20	*CSNK2A1*	411,337	472,482	34	*0.90*	*0.53*	115	*0.94*	*0.29*	100	*0.41*

4	*GPM6A*	176,791,081	177,160,642	95	*0.90*	*0.95*	383	*0.94*	*0.98*	315	*0.053*
13	*NBEA*	34,414,455	35,144,873	114	*0.93*	*0.72*	621	*0.99*	*0.82*	571	*0.32*
11	*IGF2*	2,106,922	2,127,409	21	*0.94*	*0.64*	83	*0.92*	*0.64*	67	*0.32*
1	*RGS2*	191,044,793	191,048,026	11	*0.94*	*0.82*	84	*0.88*	*0.63*	74	*0.77*
1	*BAI2*	31,965,304	32,002,235	17	*0.95*	*0.65*	61	*0.91*	*0.61*	54	*0.12*
20	*PLCB1*	8,061,295	8,813,547	290	*0.95*	*0.84*	1123	*0.94*	*0.94*	1027	*0.61*
19	*ZNF461*	41,820,122	41,849,579	9	*0.98*	*0.55*	54	*0.90*	*0.61*	52	***0.016***
19	*JUNB*	12,763,309	12,765,125	5	*0.99*	*0.48*	29	*1.00*	*0.58*	13	*0.49*

Table IV Top hits from genome wide association study with WMH variables in LBC1936 ($P < 1x10^{-5}$).

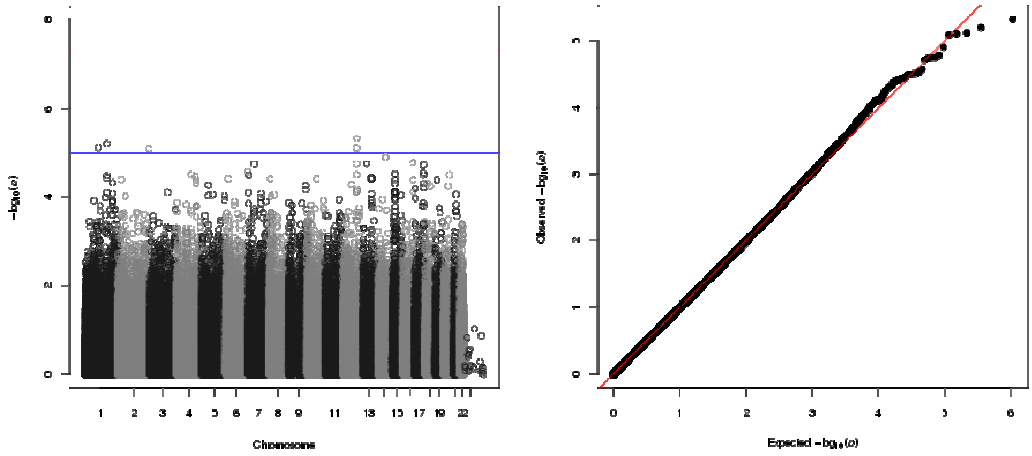
SNP	Chromosome	Position	Nearest Gene	Risk Allele	Allele Frequency	Beta	*P*
WMH							
rs7312545	12	104,877,937	*NUAK1*	A	0.48	0.1775	$4.70x10^{-6}$
rs1908311	1	164,698,889	*FMO9P*	T	0.34	0.1761	$6.27\ x10^{-6}$
rs1778193	1	94,886,005	*SLC44A3/F3*	T	0.10	-0.1733	$7.50\ x10^{-6}$
rs1344567	12	104,869,369	*NUAK1*	T	0.49	0.1735	$7.71\ x10^{-6}$
rs10439220	2	235,653,271	*SH3BP4*	C	0.27	0.1735	$8.05\ x10^{-6}$
Fazekas score							
rs1156440	4	106,244,538	*TET2*	A	0.26	0.1931	$1.45\ x10^{-6}$
rs764275	4	106,236,076	*TET2*	G	0.26	0.1858	$3.67\ x10^{-6}$
rs1991979	4	147,368,197	*LSM6/SLC10A7*	C	0.37	0.1794	$8.27\ x10^{-6}$
rs9905906	17	7,626,473	*DNAH2*	T	0.35	0.1778	$8.58\ x10^{-6}$
rs1923416	6	89,348,510	*RNGTT*	C	0.11	0.1771	$8.92\ x10^{-6}$
rs13250792	8	16,784,710	*FGF20*	G	0.27	0.1773	$9.50\ x10^{-6}$
rs6561615	13	50,595,795	*GUGY1B2*	A	0.13	0.1773	$9.87\ x10^{-6}$

Table V Top gene-based results for the LBC1936 WMH variables analysed in Vegas (P<0.001). Chr is chromosome. nSNPs is the number of SNPs in the gene (+/- 50kb). Please note that the gene boundaries are overlapping as SNPs can be allocated to multiple genes, so the same SNP could be driving the signal in different genes. The results are ordered by significance.

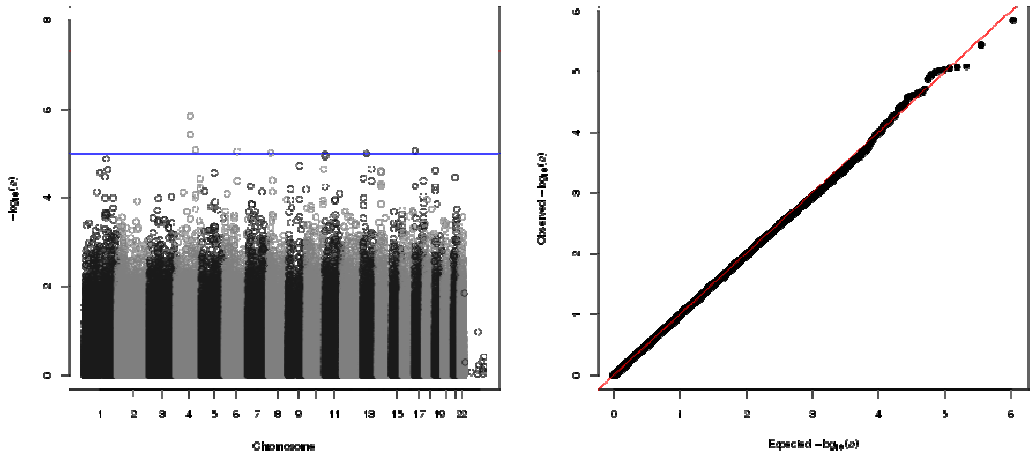
			WMH						**Fazekas Score**		
Chr	Gene	nSNPs	Start	Stop	*P*	Chr	Gene	nSNPs	Start	Stop	*P*
17	*FBF1*	17	71,418,212	71,448,714	$8.00x10^{-6}$	17	*TRIM65*	10	71,396,635	71,404,649	$5.7x10^{-5}$
17	*MRPL38*	10	71,406,318	71,413,069	$1.20\ x10^{-5}$	17	*FBF1*	17	71,418,212	71,448,714	$6.10x10^{-5}$
17	*TRIM65*	10	71,396,635	71,404,649	$1.80\ x10^{-5}$	19	*HDGF2*	21	4,423,254	4,453,222	$6.70x10^{-5}$
17	*TRIM47*	10	71,381,839	71,386,251	$1.90\ x10^{-5}$	17	*MRPL38*	10	71,406,318	71,413,069	$6.70\ x10^{-5}$
17	*WBP2*	11	71,353,374	71,363,096	$2.20\ x10^{-5}$	17	*TRIM47*	10	71,381,839	71,386,251	$7.40\ x10^{-5}$
17	*UNC13D*	13	71,334,901	71,352,393	$3.50\ x10^{-5}$	19	*UBXD1*	16	4,396,260	4,408,790	$8.30\ x10^{-5}$
17	*ACOX1*	26	71,449,186	71,487,039	$5.00\ x10^{-5}$	4	*LSM6*	17	147,316,284	147,330,663	$1.00\ x10^{-5}$
15	*SLC12A1*	27	46,285,789	46,383,568	$7.10\ x10^{-5}$	19	*LRG1*	22	4,488,226	4,491,036	*0.00010*
15	*DUT*	20	46,410,912	46,422,862	$8.40\ x10^{-5}$	1	*MRPL24*	17	154,973,717	154,977,547	*0.00011*
17	*UNK*	21	71,292,275	71,333,481	*0.00012*	1	*HDGF*	17	154,978,522	154,988,864	*0.00012*
1	*CD55*	14	205,561,439	205,600,934	*0.00022*	17	*WBP2*	11	71,353,374	71,363,096	*0.00014*
4	*LSM6*	17	147,316,284	147,330,663	*0.00024*	1	*ISG20L2*	19	154,959,036	154,964,329	*0.00016*
9	*ODF2*	8	130,258,252	130,303,060	*0.00035*	1	*C1orf66*	19	154,964,901	154,973,365	*0.00016*
1	*MRPL24*	17	154,973,717	154,977,547	*0.00037*	17	*UNC13D*	13	71,334,901	71,352,393	*0.00022*
7	*FLJ21075*	48	47,801,413	47,825,969	*0.00047*	19	*LSDP5*	24	4,473,543	4,486,208	*0.00025*
2	*ANKRD44*	61	197,567,494	197,771,007	*0.00054*	19	*KIAA1881*	23	4,453,191	4,468,716	*0.00031*
17	*H3F3B*	17	71,284,109	71,287,455	*0.00055*	16	*SSTR5*	30	1,068,869	1,069,964	*0.00031*
1	*C1orf66*	19	154,964,901	154,973,365	*0.00059*	17	*UNK*	21	71,292,275	71,333,481	*0.00035*
17	*TMEM106A*	2	38,719,419	38,727,115	*0.00059*	19	*MAP1S*	25	17,691,302	17,706,324	*0.00038*
1	*ISG20L2*	19	154,959,036	154,964,329	*0.00059*	13	*PARP4*	37	23,893,068	23,984,948	*0.00047*
17	*NBR1*	3	38,576,023	38,719,233	*0.00061*	20	*NKX2-2*	10	21,439,651	21,442,664	*0.00049*
1	*HDGF*	17	154,978,522	154,988,864	*0.00066*	19	*SEMA6B*	24	4,493,599	4,509,503	*0.00068*
17	*LOC100134934*	23	71,486,895	71,508,262	*0.00077*	15	*TLE3*	34	68,127,596	68,177,310	*0.00081*
9	*CERCAM*	8	130,222,579	130,239,451	*0.00090*	16	*C1QTNF8*	28	1,078,226	1,086,245	*0.00089*
						4	*AFP*	13	74,520,796	74,540,356	*0.00090*
						17	*ACOX1*	26	71,449,186	71,487,039	*0.00094*

Figure I Genome-wide association study results of WMH volume (a) and Fazekas score (b) using genotyped data on 542,050 SNPs in LBC1936. QQ and Manhattan plots are shown.

a



b



References

(1) Houlihan LM, Davies G, Tenesa A, Harris SE, Luciano M, Gow AJ et al. Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *Am J Hum Genet* 2010;86:626-31.

(2) Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.

(3) Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010;11:843-54.

(4) Bailey EL, McBride MW, Crawford W, McClure JD, Graham D, Dominiczak AF et al. Differential gene expression in multiple neurological, inflammatory and connective tissue pathways in a spontaneous model of human small vessel stroke. *Neuropathol Appl Neurobiol* 2014;40:855-72.