nature genetics

Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease

Mike A Nalls[1,46], Nathan Pankratz[2,46], Christina M Lill[3,4], Chuong B Do[5], Dena G Hernandez[1,6], Mohamad Saad[7–9], Anita L DeStefano[10–12], Eleanna Kara[13], Jose Bras[13], Manu Sharma[14,15], Claudia Schulte[15], Margaux F Keller[1], Sampath Arepalli[1], Christopher Letson[1], Connor Edsall[1], Hreinn Stefansson[16], Xinmin Liu[17], Hannah Pliner[1], Joseph H Lee[18], Rong Cheng[18], International Parkinson's Disease Genomics Consortium (IPDGC)[19], Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI)[19], 23andMe[19], GenePD[19], NeuroGenetics Research Consortium (NGRC)[19], Hussman Institute of Human Genomics (HIHG)[19], The Ashkenazi Jewish Dataset Investigator[19], Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE)[19], North American Brain Expression Consortium (NABEC)[19], United Kingdom Brain Expression Consortium (UKBEC)[19], Greek Parkinson's Disease Consortium[19], Alzheimer Genetic Analysis Group[19], M Arfan Ikram[20–22], John P A Ioannidis[23], Georgios M Hadjigeorgiou[24], Joshua C Bis[25], Maria Martinez[8,9], Joel S Perlmutter[26–28], Alison Goate[26,28–30], Karen Marder[18,31–33], Brian Fiske[34], Margaret Sutherland[35], Georgia Xiromerisiou[24,36], Richard H Myers[10], Lorraine N Clark[17,18], Kari Stefansson[16], John A Hardy[6], Peter Heutink[37], Honglei Chen[38], Nicholas W Wood[13], Henry Houlden[13], Haydeh Payami[39], Alexis Brice[40–42], William K Scott[43], Thomas Gasser[15], Lars Bertram[3,44], Nicholas Eriksson[5], Tatiana Foroud[45] & Andrew B Singleton[1]

We conducted a meta-analysis of Parkinson's disease genome-wide association studies using a common set of 7,893,274 variants across 13,708 cases and 95,282 controls. Twenty-six loci were identified as having genome-wide significant association; these and 6 additional previously reported loci were then tested in an independent set of 5,353 cases and 5,551 controls. Of the 32 tested SNPs, 24 replicated, including 6 newly identified loci. Conditional analyses within loci showed that four loci, including *GBA*, *GAK-DGKQ*, *SNCA* and the HLA region, contain a secondary independent risk variant. In total, we identified and replicated 28 independent risk variants for Parkinson's disease across 24 loci. Although the effect of each individual locus was small, risk profile analysis showed substantial cumulative risk in a comparison of the highest and lowest quintiles of genetic risk (odds ratio (OR) = 3.31, 95% confidence interval (CI) = 2.55–4.30; $P = 2 \times 10^{-16}$). We also show six risk loci associated with proximal gene expression or DNA methylation.

Increasing evidence supports an extensive and complex genetic contribution to Parkinson's disease. Genome-wide association studies (GWAS) have shed light on the genetic basis of this disease, with the identification and replication of risk loci that fit the common disease, common variant hypothesis[1–17]. The loci identified have both confirmed the central role of the genes previously linked to Parkinson's disease and implicated new proteins in the pathogenic cascade[18]. These data have also shown that, thus far, only a small portion of the heritable component for Parkinson's disease has been identified[19]. Experience in other complex diseases and traits demonstrates that ever greater resolution of genetic risk can be achieved through larger sample sizes and that common genetic variation may have a more substantial role in complex traits than previously anticipated[20–22]. With each of these factors in mind, we performed a meta-analysis of all existing European-ancestry Parkinson's disease GWAS data and a replication study in an independent data set.

We performed a meta-analysis of genome-wide SNP data from 13,708 cases with Parkinson's disease and 95,282 controls. This approach required imputation using the August 2010 release of the 1000 Genomes Project European-ancestry haplotype reference set to standardize data for over 11 million variants[23]. Only markers that were successfully imputed in at least 3 data sets and that had a sample size–weighted minor allele frequency across the meta-analysis of 0.1% or higher were included (n = 7,893,274). The genomic inflation factor for each of the data sets (based on λ values standardized to a scale for 1,000 cases and 1,000 controls) ranged from 0.889 to 1.056 (see **Supplementary Table 1** for study-specific details). Fixed-effect meta-analysis of the summary statistics from each set identified 26 loci associated with risk for

A full list of author affiliations appears at the end of the paper.

Received 23 September 2013; accepted 30 June 2014; published online 27 July 2014; doi:10.1038/ng.3043

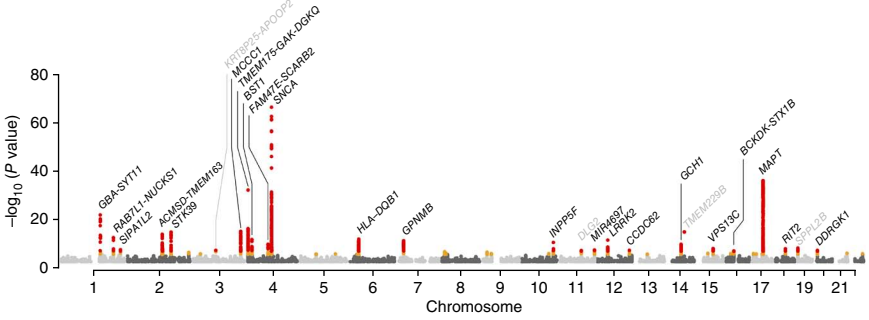

Figure 1 Manhattan plot of discovery-phase meta-analyses. Black font denotes replicated loci from the discovery phase, and gray font denotes loci that did not replicate.

disease in the discovery phase, on the basis of a widely accepted genome-wide *P*-value threshold of 5×10^{-8} (**Fig. 1** and **Table 1**; additional details are provided in **Supplementary Fig. 1** and **Supplementary Table 2**)[24].

To identify which of the putatively associated loci were truly relevant for disease, we attempted to replicate each locus in an independent sample series using a semi-custom genotyping array called NeuroX. This array typed the >240,000 exonic variants available on the Illumina Infinium HumanExome BeadChip and an additional ~24,000 variants proven or hypothesized to be relevant in neurodegenerative disease (M.A.N., J.B., D.G.H., M.F.K., E. Majounie *et al.*, unpublished data). Within the custom content, we included the 26 genome-wide significant candidate loci implicated in Parkinson's disease from the primary meta-analysis. For each independent locus, the most significantly associated SNP and a series of proxy variants were included in the array design. After stringent quality control, high-quality genotype data were available for a sample set of 5,353 cases and 5,551 controls (see the Online Methods for

Table 1 Results of discovery and replication association analyses

SNP information							Discovery phase (13,728 cases and 95,282 controls)		Replication phase (5,353 cases and 5,551 controls)		Joint phase (19,081 cases and 100,833 controls)	
SNP	Chr.	Position (bp)	Nearest gene(s)	Effect allele	Alternate allele	Effect allele frequency	OR	*P*	OR	*P*	OR	*P*
Genome-wide significant, discovery phase												
rs35749011[a]	1	155,135,036	*GBA-SYT11*	A	G	0.017	1.762	6.09×10^{-23}	2.307	7.48×10^{-9}	1.824	1.37×10^{-29}
rs823118	1	205,723,572	*RAB7L1-NUCKS1*	T	C	0.559	1.126	1.36×10^{-13}	1.109	1.43×10^{-4}	1.122	1.66×10^{-16}
rs10797576	1	232,664,611	*SIPA1L2*	T	C	0.14	1.139	1.19×10^{-8}	1.11	3.38×10^{-3}	1.131	4.87×10^{-10}
rs6430538	2	135,539,967	*ACMSD-TMEM163*	T	C	0.43	0.873	5.56×10^{-15}	0.882	9.42×10^{-6}	0.875	9.13×10^{-20}
rs1474055[a]	2	169,110,394	*STK39*	T	C	0.128	1.213	7.12×10^{-16}	1.218	1.07×10^{-6}	1.214	1.15×10^{-20}
rs115185635[a]	3	87,520,857	*KRT8P25-APOOP2*	C	G	0.035	1.789	2.18×10^{-8}	0.931	0.846	1.142	0.022
rs12637471	3	182,762,437	*MCCC1*	A	G	0.193	0.844	3.32×10^{-16}	0.836	3.72×10^{-7}	0.842	2.14×10^{-21}
rs34311866	4	951,947	*TMEM175-GAK-DGKQ*	T	C	0.809	0.784	3.58×10^{-33}	0.791	6.29×10^{-12}	0.786	1.02×10^{-43}
rs11724635	4	15,737,101	*BST1*	A	C	0.553	1.122	8.07×10^{-13}	1.138	2.73×10^{-6}	1.126	9.44×10^{-18}
rs6812193	4	77,198,986	*FAM47E-SCARB2*	T	C	0.364	0.897	7.17×10^{-11}	0.935	0.011	0.907	2.95×10^{-11}
rs356182	4	90,626,111	*SNCA*	A	G	0.633	0.737	3.23×10^{-67}	0.822	1.75×10^{-12}	0.760	4.16×10^{-73}
rs9275326[a]	6	32,666,660	*HLA-DQB1*	T	C	0.094	0.797	5.82×10^{-13}	0.9	0.018	0.826	1.19×10^{-12}
rs199347	7	23,293,746	*GPNMB*	A	G	0.59	1.123	2.37×10^{-12}	1.072	7.66×10^{-3}	1.110	1.18×10^{-12}
rs117896735[a]	10	121,536,327	*INPP5F*	A	G	0.014	1.767	1.21×10^{-11}	1.404	1.10×10^{-3}	1.624	4.34×10^{-13}
rs3793947[a]	11	83,544,472	*DLG2*	A	G	0.443	0.912	2.59×10^{-8}	0.976	0.201	0.929	3.96×10^{-7}
rs329648	11	133,765,367	*MIR4697*	T	C	0.354	1.1	1.65×10^{-8}	1.121	4.38×10^{-5}	1.105	9.83×10^{-12}
rs76904798	12	40,614,434	*LRRK2*	T	C	0.143	1.17	1.33×10^{-12}	1.11	3.69×10^{-3}	1.155	5.24×10^{-14}
rs11060180	12	123,303,586	*CCDC62*	A	G	0.558	1.101	2.14×10^{-8}	1.114	7.26×10^{-5}	1.105	6.02×10^{-12}
rs11158026	14	55,348,869	*GCH1*	T	C	0.335	0.889	7.13×10^{-11}	0.948	0.039	0.904	5.85×10^{-11}
rs1555399[a]	14	67,984,370	*TMEM229B*	A	T	0.468	0.872	5.53×10^{-16}	0.971	0.144	0.897	6.63×10^{-14}
rs2414739	15	61,994,134	*VPS13C*	A	G	0.734	1.114	4.13×10^{-9}	1.109	7.96×10^{-4}	1.113	1.23×10^{-11}
rs14235	16	31,121,793	*BCKDK-STX1B*	A	G	0.381	1.094	3.89×10^{-8}	1.133	7.72×10^{-6}	1.103	2.43×10^{-12}
rs17649553	17	43,994,648	*MAPT*	T	C	0.226	0.771	4.86×10^{-37}	0.764	7.03×10^{-15}	0.769	2.37×10^{-48}
rs12456492	18	40,673,380	*RIT2*	A	G	0.693	0.905	5.12×10^{-9}	0.9	2.16×10^{-4}	0.904	7.74×10^{-12}
rs62120679[a]	19	2,363,319	*SPPL2B*	T	C	0.314	1.141	2.53×10^{-9}	0.999	0.518	1.097	5.57×10^{-7}
rs8118008[a]	20	3,168,166	*DDRGK1*	A	G	0.657	1.111	2.32×10^{-8}	1.113	1.18×10^{-4}	1.111	3.04×10^{-11}
Previously reported as significant in genome-wide studies												
rs34016896	3	160,992,864	*NMD3*	T	C	0.319	1.08	7.68×10^{-6}	1.028	0.174	1.067	1.08×10^{-5}
rs591323	8	16,697,091	*FGF20*	A	G	0.275	0.921	1.30×10^{-5}	0.902	6.16×10^{-4}	0.916	6.68×10^{-8}
rs60298754	8	89,373,041	*MMP16*	T	C	0.024	1.078	0.181	–	–	1.078	0.181
rs7077361	10	15,561,543	*ITGA8*	T	C	0.874	1.11	3.24×10^{-5}	1.044	0.154	1.092	4.16×10^{-5}
rs11868035	17	17,715,101	*SREBF1-RAI1*	A	G	0.298	0.937	2.17×10^{-4}	0.947	0.036	0.939	5.98×10^{-5}
rs2823357	21	16,914,905	*USP25*	A	G	0.37	1.036	0.032	1.018	0.267	1.031	0.027

Note, only replication-phase *P* values are one-sided. Nearest gene or previously published proximal gene names are included. Chr., chromosome; OR, odds ratio.
[a]Replication genotyping for these SNPs failed assay design or quality control, and a suitable proxy variant was selected (rs35749011, proxy rs71628662; rs1474055, proxy rs1955337; rs115185635, proxy rs62267708; rs117896735, proxy rs118117788; rs3793947, proxy rs12283611; rs1555399, proxy rs1077989; rs62120679, proxy rs10402629; rs8118008, proxy rs55785911).

complete details). Association analysis showed replication of 22 of the 26 loci tested, on the basis of a nominal one-sided *P*-value threshold of <0.05 and consistent direction of association that incorporated the premise of prior knowledge for most loci based on previous meta-analysis of GWAS data (**Table 1**); of these loci, 6 were new (*SIPA1L2*, *INPP5F*, *MIR4697*, *GCH1*, *VPS13C* and *DDRGK1*). In addition, we examined association at six loci previously reported to be associated with risk for Parkinson's disease that did not show association at $P < 5 \times 10^{-8}$ in the discovery phase[1,2,4,25]. Although these loci have been reported in samples derived from some of the cohorts included in the discovery phase of this meta-analysis, individuals in the replication samples were distinct from those used to nominate these loci. We found evidence for association, on the basis of a nominal one-sided *P*-value threshold of <0.05 in the replication data, at two of these loci in our replication-phase analyses (*FGF20* and *SREBF1-RAI1*; **Table 1**). We do note that some loci did not replicate; these loci included regions of high effect heterogeneity and low effect size (OR of ~1.1), for which our replication series might have been slightly underpowered. For example, at an OR of 1.1 and an allele frequency of 5%, our replication series were only at a power of ~35% to reach our designated target *P* value, whereas, under the assumption of no effect heterogeneity across the replication samples and no winner's curse phenomenon, we were at ~80% power to reach our target α value for replication if the allele frequency was increased to 40% and the OR remained 1.1. We recognize that study heterogeneity contributed to some of this non-replication (as evidenced by the I2 metrics in **Supplementary Table 2**), particularly in the discovery phase of analyses. In the discovery phase, the associations at rs115185635 and rs1555399 were driven almost completely by data from the IPDGC-UK cohort and were highly heterogeneous across cohorts (I^2 estimates at 91.0 and 97.2, respectively). In addition, rs115185635 was a very difficult variant to impute, likely owing to its frequency, as it passed quality control in only eight of our participating studies. We do not believe there is any major issue with the UK data, on the basis of both previously published studies and consistent effects at more established loci, evidenced by the I^2 metrics listed in **Supplementary Tables 1–3** and **Supplementary Figures 2–4** describe study-specific effect estimates in addition to giving genomic inflation factors. Our strategy of a distinct replication phase was instituted primarily to confirm suspect newly associated loci and to exclude any type of

Table 2 Results of conditional association analyses

Significant conditional SNP, signifying secondary locus		rs114138760	rs79217002	rs34884217	rs1596117*	rs7681154*	rs13201101*	rs10886515	rs117022814
Most significant SNP from discovery phase, used as covariate		rs35749011	rs12637471	rs34311866	rs6812193	rs356182	rs9275326	rs117896735	rs62120679
Nearest gene(s)		*GBA-SYT11*	*MCCC1*	*TMEM175-GAK-DGKQ*	*FAM47E-SCARB2*	*SNCA*	*HLA-DQB1*	*INPP5F*	*SPPL2B*
r^2 between SNPs based on 1000 Genomes Project European-ancestry samples		0.000	0.003	0.012	0.028	0.209	0.002	0.000	0.006
Conditional SNP information	Chr.	1	3	4	4	4	6	10	19
	Position (bp)	154,898,185	183,011,072	944,210	77,151,490	90,763,703	32,343,604	121,343,589	2,209,647
	Effect allele	C	A	A	T	A	T	T	T
	Alternate allele	G	g	C	C	C	C	C	C
	Effect allele frequency	0.012	0.9907	0.9126	0.2005	0.5021	0.0529	0.7145	0.0262
Summary statistics from conditional analyses	OR	1.574	0.669	1.247	1.115	0.841	1.192	1.100	1.341
	P	3.80×10^{-7}	9.31×10^{-6}	1.10×10^{-6}	2.80×10^{-7}	7.09×10^{-19}	3.84×10^{-6}	9.19×10^{-7}	1.95×10^{-6}
Summary statistics from discovery phase	OR	1.497	0.688	1.344	1.094	0.997	1.179	1.105	1.319
	P	2.18×10^{-6}	1.69×10^{-5}	1.56×10^{-12}	6.05×10^{-6}	0.854	4.95×10^{-6}	2.59×10^{-8}	2.00×10^{-6}
Summary statistics from replication phase	OR	1.586	1.076	1.105	1.036	0.934	1.217	1.023	1.094
	P	5.72×10^{-4}	0.714	0.017	0.189	8.02×10^{-3}	8.33×10^{-3}	0.234	0.174
Summary statistics from combined discovery and replication phases	OR	1.519	0.789	1.232	1.083	0.981	1.185	1.084	1.255
	P	9.73×10^{-9}	1.08×10^{-3}	2.51×10^{-11}	9.45×10^{-6}	0.171	2.50×10^{-7}	2.26×10^{-7}	4.82×10^{-6}

Replication genotyping for these SNPs failed assay design or quality control, and a suitable proxy variant was selected (rs1596117, proxy rs4859430; rs7681154, proxy rs3910105; rs13201101, proxy rs8192591; on the basis of discovery series comparison, the minor allele for rs3910105 tags the major allele of rs7681154, and risk is therefore consistent across the proxy and discovery SNP). Note, only replication-phase *P* values are one-sided. Nearest gene or previously published proximal gene names are included.

systematic issue that might lead to false positives at individual loci.

We tested whether multiple independent risk alleles existed at any of the 26 genome-wide significant loci identified in the discovery phase. For each locus, we tested all variants within 1 Mb of the index SNP with the most extreme *P* value. To identify risk alleles independent of the primary effect, the index SNP was included as a covariate in the model (with 0, 1 or 2 copies of the minor allele). Additional independent risk variants were identified at eight of the loci (*P* values of 9.31×10^{-6} to 7.09×10^{-19}) and were also included on the replication array. Four of these variants showed significant association upon conditional analysis of the replication-phase data (**Table 2**; see **Supplementary Table 3** for additional details).

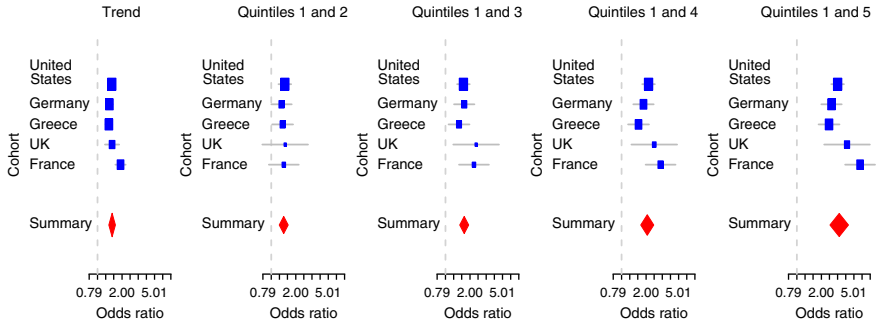

Figure 2 Forest plots describing cohort-level and summary effects from risk profile analyses. Horizontal lines denote 95% confidence intervals.

Risk profiles were generated for each subset of replication samples separately using all SNPs with replication-phase *P* values less than the marginal one-sided *P* value of 0.05 (discovery, candidate and conditional phases; **Tables 1** and **2**). These 28 SNPs were used to compute genetic risk profile scores (for additional risk-profiling methods, please see the Online Methods). Similar to previous studies, we showed marginal predictive power for genetic risk profile scores, with areas under the receiver operator curves of 0.616 without age and sex included as covariates and 0.633 with age and sex included (**Fig. 2**, **Supplementary Fig. 5** and **Supplementary Table 4**)[3]. As expected, individuals with a genetic risk profile score greater than 1 s.d. from the population mean, indicative of a roughly 34% increase in genetic risk score above the mean for controls, had a significantly higher risk of Parkinson's disease (from meta-analysis; OR = 1.51, 95% CI = 1.38–1.66; $P = 2 \times 10^{-16}$). In an analysis of outliers, we compared the fifth quintile of genetic risk scores to the first quintile of genetic risk as a reference; the OR was 3.31 (95% CI = 2.55–4.30; $P = 2 \times 10^{-16}$). These OR estimates are larger in comparison to those in earlier publications and might be due to the finer-scale imputation used in the discovery phase of this project, as well as to the inclusion of additional loci and, to some degree, differing distributions of cumulative genetic risk scores across populations in the analysis[3,4]. Cohort-level summary statistics were significantly heterogeneous for both trend-based analyses ($I^2 = 0.74$, heterogeneity *P* value = 0.003) and comparisons of the highest and lowest risk quintiles ($I^2 = 0.70$, heterogeneity *P* value = 0.01). Therefore, a random-effects model was used to account for the heterogeneity in effect.

For each of the 28 SNPs included in the risk profile analyses, we attempted to infer functional consequences in frontal cortex and cerebellar tissue samples from neurologically normal individuals that were assayed for both genome-wide methylation and expression levels[26]. These analyses may shed light on potential disease mechanisms for follow-up in future studies. We tested *cis* associations (analyzing any methylation or expression probes within 1 Mb of each SNP) in each of the data sets. After quality control, 25 SNPs of interest from our meta-analysis passed quality control in the mRNA expression data sets, and all 28 SNPs of interest passed quality control in the CpG methylation data sets. We tested multiple probes for each SNP in each set of analyses. A total of 336 unique SNP-probe pairs were tested in the frontal cortex mRNA expression data set, 865 pairs were tested in the frontal cortex CpG methylation data set, 333 pairs were tested in the cerebellar mRNA expression data set, and 1,097 pairs were tested in the cerebellar CpG methylation data set. Associations were tested using linear regression adjusting for appropriate covariates, and resulting *P* values were adjusted on the basis of false discovery rate (FDR) correction (see Online Methods for details).

After correcting for multiple tests, we found 30 significant associations between SNPs of interest and either CpG methylation or mRNA expression (**Supplementary Table 5**) across 6 loci. Of particular interest were associations at rs199347 on chromosome 7 and rs823118 on chromosome 1, as both SNPs were significantly associated with both methylation and expression changes in each brain region. The risk allele (A) at rs199347 on chromosome 7 was associated with higher expression of two probes tagging *NUPL2* as well as with decreased methylation of *GPNMB* in both brain regions. These data suggest that risk at the locus containing rs199347 might be due to increased transcription of *NUPL2*, further bolstered by decreased methylation. On chromosome 1, the risk allele (T) at rs823118 was associated with lower expression of *NUCKS1* and higher expression of *RAB7L1*, as well as with increased DNA methylation detected by two probes close to *PM20D1* in both brain tissues. These data suggest a complicated risk locus at the *NUCKS1-RAB7L1-PM20D1* region, where the same risk allele is associated with both increased expression of *RAB7L1* and increased regulation of the nearby genes *NUCKS1* and *PM20D1*. The possibility of multiple functionally active risk variants at this locus seems likely and is evident in the results of our conditional phase of analyses (**Table 2**). The complicated nature of this locus may be suggestive of some type of interaction or epistatic effect as well, and it is likely that future functional and deep sequencing studies will be required to understand the basis of association in this region.

In total, we have here identified 28 independent risk loci for Parkinson's disease: 22 found in the discovery phase and confirmed by replication, 2 previously reported loci confirmed in the replication phase and 4 loci identified by a second risk allele exerting an effect independently of the primary risk allele.

URLs. mach2qtl v1.11, http://www.sph.umich.edu/csg/abecasis/MaCH/download/; Minimac, http://genome.sph.umich.edu/wiki/Minimac; 1000 Genomes Project haplotypes, http://www.sph.umich.edu/csg/abecasis/MaCH/download/. Summary statistics of this study have been made available at http://www.pdgene.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

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

ACKNOWLEDGMENTS

We would like to thank all of the subjects who donated their time and biological samples to be a part of this study. For funding details and additional acknowledgments, please see the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Overall study design: M.A.N., N.P., J.B., A.L.D., B.F., M. Sutherland, J.A.H., N.W.W., T.G., W.K.S., L.B., N.E., T.F. and A.B.S. Design and/or management of the individual studies: M.A.N., C.B.D., J.B., C.S., X.L., J.H.L., R.C., G.M.H., J.S.P., A.G., K.M., A.L.D., R.H.M., L.N.C., J.A.H., P.H., H.C., M. Saad, M. Sharma, M. Sutherland, M.A.I., J.C.B., N.W.W., H.H., H. Payami, H.S., K.S., A.B., W.K.S., T.G., N.E., T.F. and A.B.S. Genotyping: D.G.H., E.K., S.A., C.L., C.E. and H. Pilner. Phenotyping: T.F., G.M.H., J.S.P., K.M., G.X., H.C., N.W.W., H.H., H.S., K.S., A.B., T.F. and W.K.S. Statistical methods and data analysis: M.A.N., N.P., C.M.L., D.G.H., E.K., M. Saad, M. Sharma, C.S., J.P.A.I., M.F.K., M.M., A.L.D., W.K.S., L.B., N.E., T.F. and A.B.S. Writing group: M.A.N., N.P., C.M.L., T.F. and A.B.S. Critical review of the manuscript: M.A.N., N.P., C.M.L., C.B.D., D.G.H., E.K., J.B., C.S., M.F.K., G.M.H., M.M., A.G., B.F., M. Saad, M. Sharma, M. Sutherland, G.X., R.H.M., L.N.C., J.A.H., P.H., H.C., N.W.W., H.H., H. Payami, H. Pilner, H.S., K.S., A.B., W.K.S., T.G., L.B., N.E., T.F., A.B.S. and J.S.P.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Lill, C.M. *et al.* Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet.* **8**, e1002548 (2012).
2. Do, C.B. *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* **7**, e1002141 (2011).
3. International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
4. International Parkinson's Disease Genomics Consortium (IPDGC) & Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet.* **7**, e1002142 (2011).
5. Edwards, T.L. *et al.* Genome-wide association study confirms SNPs in *SNCA* and the *MAPT* region as common risk factors for Parkinson disease. *Ann. Hum. Genet.* **74**, 97–109 (2010).
6. Pankratz, N. *et al.* Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.* **124**, 593–605 (2009).
7. Pankratz, N. *et al.* Meta-analysis of Parkinson's disease: identification of a novel locus, *RIT2*. *Ann. Neurol.* **71**, 370–384 (2012).
8. Simón-Sánchez, J. *et al.* Genome-wide association study confirms extant PD risk loci among the Dutch. *Eur. J. Hum. Genet.* **19**, 655–661 (2011).
9. Hamza, T.H. *et al.* Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* **42**, 781–785 (2010).
10. Liu, X. *et al.* Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Med. Genet.* **12**, 104 (2011).
11. Hernandez, D.G. *et al.* Genome wide assessment of young onset Parkinson's disease from Finland. *PLoS ONE* **7**, e41859 (2012).
12. Pihlstrøm, L. *et al.* Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. *Neurobiol. Aging* **34**, 1708.e7–13 (2013).
13. Sharma, M. *et al.* Large-scale replication and heterogeneity in Parkinson disease genetic loci. *Neurology* **79**, 659–667 (2012).
14. Saad, M. *et al.* Genome-wide association study confirms *BST1* and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum. Mol. Genet.* **20**, 615–627 (2011).
15. Satake, W. *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* **41**, 1303–1307 (2009).
16. Elbaz, A. *et al.* Independent and joint effects of the *MAPT* and *SNCA* genes in Parkinson disease. *Ann. Neurol.* **69**, 778–792 (2011).
17. Psaty, B.M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
18. MacLeod, D.A. *et al.* RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* **77**, 425–439 (2013).
19. Keller, M.F. *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).
20. Wei, Z. *et al.* Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am. J. Hum. Genet.* **92**, 1008–1012 (2013).
21. Willems, S.M., Mihaescu, R., Sijbrands, E.J.G., van Duijn, C.M. & Janssens, A.C.J.W. A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research. *Curr. Diab. Rep.* **11**, 511–518 (2011).
22. Morrison, A.C. *et al.* Whole-genome sequence–based analysis of high-density lipoprotein cholesterol. *Nat. Genet.* **45**, 899–901 (2013).
23. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
24. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
25. van der Walt, J.M. *et al.* Fibroblast growth factor 20 polymorphisms and haplotypes strongly influence risk of Parkinson disease. *Am. J. Hum. Genet.* **74**, 1121–1127 (2004).
26. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).

[1]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, USA. [2]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA. [3]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [4]Department of Neurology, Focus Program Translational Neuroscience, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany. [5]23andMe, Inc., Mountain View, California, USA. [6]Reta Lila Weston Institute, University College London Institute of Neurology, Queen Square, London, UK. [7]Department of Biostatistics, University of Washington, Seattle, Washington, USA. [8]INSERM, UMR 1043, Centre de Physiopathologie de Toulouse-Purpan, Toulouse, France. [9]Paul Sabatier University, Toulouse, France. [10]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA. [11]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [12]National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, Massachusetts, USA. [13]Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK. [14]Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany. [15]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. [16]deCODE Genetics, Reykjavík, Iceland. [17]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, New York, USA. [18]The Taub Institute for Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York, USA. [19]A full list of members and affiliations appears in the **Supplementary Note**. [20]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [21]Department of Radiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [22]Department of Neurology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [23]Stanford Prevention Research Center, Stanford University, Stanford, California, USA. [24]Neuroscience Unit, Department of Neurology, Faculty of Medicine, University of Thessaly, Larissa, Greece. [25]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [26]Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, Missouri, USA. [27]Department of Radiology, Washington University School of Medicine, St. Louis, Missouri, USA. [28]Department of Neurology, Washington University School of Medicine, St. Louis, Missouri, USA. [29]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [30]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [31]Gertrude H. Sergievsky Center, Columbia University Medical Center, New York, New York, USA. [32]Department of Neurology, Columbia University Medical Center, New York, New York, USA. [33]Department of Psychiatry, Columbia University Medical Center, New York, New York, USA. [34]The Michael J. Fox Foundation for Parkinson's Research, New York, New York, USA. [35]Neuroscience Center, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, USA. [36]Department of Neurology, Papageorgiou Hospital, Thessaloniki, Greece. [37]Genome Biology for Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany. [38]Epidemiology Branch, National Institute of Environmental Health Sciences, US National Institutes of Health, Research Triangle, North Carolina, USA. [39]New York State Department of Health Wadsworth Center, Albany, New York, USA. [40]Sorbonne Université, UPMC Université Paris 06, UM 75, INSERM U1127, Institut du Cerveau et de la Moelle, Paris, France. [41]CNRS, UMR 7225, Paris, France. [42]Pitié-Salpêtrière Hospital, Department of Genetics and Cytogenetics, Paris, France. [43]Department of Human Genetics, University of Miami School of Medicine, Miami, Florida, USA. [44]School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London, UK. [45]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [46]These authors contributed equally to this work.
Correspondence should be addressed to A.B.S. (singleta@mail.nih.gov).

ONLINE METHODS

Discovery methods. All studies willing to participate with genome-wide genotyping data on Parkinson's diseases cases and controls were included in this effort. For specific details on these studies, please refer to individual publications (IPDGC, PD GWAS Consortium, 23andMe, CHARGE, PDGene and Ashkenazi studies)[2–6,10,17,27–30]. For this analysis, the 23andMe cohort was split into two subsets 'v2' and 'v3'. The v2 designation refers to a subset of samples genotyped on the Illumina HumanHap550+ BeadChip, and the v3 designation refers to a subset of samples genotyped on the Illumina Human OmniExpress+ BeadChip. Aside from the use of different genotyping arrays, sample handling, ascertainment, quality control and analytic methods were identical across the 23andMe subsets. For 3 of the 15 studies contributing to the discovery phase of analyses, population controls were used to some degree, totaling 8,156 samples (**Supplementary Table 1**). On the basis of disease prevalence of 2 in 1,000, we could estimate a misclassification of approximately 16 samples across data sets. This misclassification would likely have a small impact because of our large sample size, meta-analysis design and interest in relatively common alleles.

In general, standard quality control was applied by each study for sample inclusion, including requirements of a case age at disease onset of over 18 years, no known mutations in genes associated with mendelian forms of Parkinson's disease (*SNCA*, *PARK2*, *DJ-1*, *PINK1* and *LRRK2*), minimum sample call rate of >95%, European ancestry confirmed through principal-components or multidimensional scaling analyses and no relation to other samples in the meta-analysis (checked, when possible, by the use of database of Genotypes and Phenotypes (dbGaP) available data) at the cousin level or closer (except in the case of the Framingham Heart Study). Studies deriving samples from the same geographic region (or globally in the case of the 23andMe data set) cross-checked for relatedness when data access was permitted by using identity-by-descent filtering to remove related samples both within and across data sets contributing to the meta-analysis. If samples overlapped with studies involved in the meta-analysis, these samples were excluded from the series with less dense genotyping. All participants donated DNA samples and provided informed consent for participation in genetics studies. The study was approved by the relevant ethical committees (full details in the **Supplementary Note**). Before imputation, SNPs were filtered using study-specific criteria, including requirements for a minimum call rate of >95%, a minor allele frequency of >1%, a Hardy-Weinberg equilibrium P values $>1 \times 10^{-4}$ in controls and non-random missingness by phenotype or haplotype at a P values of $>1 \times 10^{-4}$. SNPs ambiguous to strand (A/T and G/C) were also removed. Imputation to a standard reference panel from the 1000 Genomes Project (August 2010 release; European ancestry only) was then carried out using Minimac with default settings[31].

Imputed dosages were then analyzed using logistic regression for case-control studies or Cox regression for cohort studies (CHARGE Consortium cohorts with incident cases) adjusting for the first two eigenvectors from principal-components analysis, age at disease onset (cases) or exam (controls), and sex. Eigenvectors from principal-components analysis for use as covariates were generated on a study-specific level, with each data set applying its own adjustment separately. This adjustment was also repeated in the replication phase by generating unique eigenvectors for each ancestry-stratified data set for use as covariates. The Framingham Heart Study used generalized estimating equations clustered on pedigrees to account for family relationships.

Meta-analysis was conducted on the basis of the fixed-effect model as implemented in METAL by combining summary statistics across data sets[32]. At the meta-analysis level, summary statistics were filtered for inclusion after meeting a minimum imputation quality score (RSQ from Minimac) of 0.30, having a minor allele frequency of greater than 0.1% across studies, having realistic β coefficients where the absolute value of β was less than 5 and passing initial quality control in at least three of the contributing studies. In addition, we tested novel random-effects approaches from Han and Eskin[33]; however, this method did not identify any new loci, and, as the results across both methodologies were nearly identical, only fixed-effect results are reported here.

Conditional methods. Conditional analyses were undertaken using identical statistical models as in the discovery phase except for the inclusion of allele dosages from the most significantly associated SNP for each locus as an additional covariate. For each locus identified as being genome-wide significant in the discovery phase, we reran cohort level analyses in a subset of 7 data sets with the largest counts of cases (owing to primary data availability at the participant level, only IPDGC-NIA, IPDGC-NL, IPDGC-GE, 23andMe, PROGENI-GenePD, NGRC and HIHG were included), testing all SNPs within 1 Mb of the 26 genome-wide significant SNPs listed in **Table 1**, while adjusting for the SNP with the most extreme P value for each locus. We then performed meta-analysis on these summary statistics in the same manner as in the discovery-phase analyses. The threshold for multiple-test correction across secondary loci for conditional analyses was set to 1×10^{-5} on the basis of Bonferroni correction for the number of SNPs tested across all regions. These methods were also applied to look for tertiary signals at all loci; three tertiary loci were identified, but these signals were not included in the replication array and are therefore not shown (data available upon request to corresponding author).

Replication methods. Replication genotyping was carried out using the Illumina NeuroX genotyping array, with all samples genotyped at the National Institute on Aging Laboratory of Neurogenetics (LNG). In brief, the NeuroX array includes over 24,000 neurodegenerative-focused variants added to the existing >240,000 exonic variants already available on the Illumina Infinium HumanExome BeadChip. Of these neurodegenerative-focused variants, over 9,000 are dedicated to Parkinson's disease and include tagging SNPs, proxies and technical replicates for loci of interest related to the discovery phase of this study. Each genome-wide significant locus identified in the discovery phase of this study and in the conditional analyses was covered on the array by either five additional proxy SNPs or five technical replicates, if no proxy SNPs were available. Loci were defined as any SNP with a genome-wide significant P value that was correlated at $r^2 < 0.50$ with any other significant SNPs within 250 kb of each genomic region of interest. Proxies were selected on the basis of this 250-kb threshold, and linkage disequilibrium was defined using 1000 Genomes Project European-ancestry samples from the same panel from which the discovery series was imputed. Proxies were ranked by discovery-phase P value after meeting the linkage disequilibrium minimum threshold of $r^2 > 0.50$ with the most significantly associated SNP in the locus. Nominated proxies with the smallest discovery-phase P value were given precedence in replication analyses when the most significant SNP from the discovery phase was not available or not successfully assayed on the NeuroX array. For replication, 39 SNPs or their highest ranked proxy in terms of discovery-phase or conditional P value were used. All summary statistics for these replication SNPs will be made available on the PDGene website (PDGene database; see URLs). Genotypes were called using Illumina GenomeStudio software, and all Parkinson's disease–related SNPs analyzed for this study were manually clustered and visually inspected. Standard exome content variants included on the NeuroX array were called using a cluster file from the CHARGE Consortium on the basis of over 60,000 samples, and these variants were used for sample quality control[34]. Over 14,000 samples genotyped on the array at LNG were used in the variant calling process.

From called genotypes, Parkinson's disease cases and neurologically normal controls were extracted and underwent quality control according to standard GWAS protocols, with slight deviations from normal practices to account for the bias in NeuroX array content. Variants with GenTrain scores of >0.70 (indicative of quality genotype clusters) for the standard content on the NeuroX array were extracted first to calculate call rates. Samples with call rates of <95% were excluded, as were samples whose genetically determined sex did not match that from clinical data and samples exhibiting excess heterozygosity. After these initial exclusions, SNPs overlapping with HapMap Phase 3 samples were extracted from the previous subset and pruned for linkage disequilibrium (SNPs were excluded if they had $r^2 > 0.50$ within a 50-SNP sliding window), and we concurrently excluded SNPs with minor allele frequencies of <5%, Hardy-Weinberg equilibrium P values of $<1 \times 10^{-5}$ in controls and per-SNP missingness rates of >5%. At this stage, pairwise identity-by-descent filtering was used to remove samples that were cryptically related, and principal-components analysis was used to identify samples that were to be excluded owing to their genetic ancestry not being consistent with primarily European ancestry, on the basis of comparisons with HapMap Phase 3 reference populations. After these exclusions, the samples passing quality control were separated into distinct data sets on the basis of country and center of origin. All samples in the replication series were ascertained as follows. Case

doi:10.1038/ng.3043

ascertainment was based on UK brain bank criteria from a clinical visit or on the use of medication for Parkinson's disease or medical records of Parkinson's disease diagnosis by a clinician. Recruited controls included individuals free of known neurological disease as determined by clinical assessment and/or by self-report. The final replication set consisted of 5,353 cases and 5,551 controls, with all relevant phenotypic data for this analysis stratified across US (2,407 cases and 2,782 controls), French (553 cases and 474 controls), German (1,044 cases and 871 controls), Greek (944 cases and 877 controls) and UK (405 cases and 547 controls) participants.

Within each subset of samples passing quality control, principal-components analysis was used to generate eigenvectors for use as covariates to account for population substructure within each cohort, on the basis of common, high-quality SNPs that had also been pruned for linkage disequilibrium as described above. Within each subset of samples, logistic regression adjusting for the first two eigenvectors from principal-components analysis, age at disease onset (cases) or exam (controls), and sex was used to examine the association of each nominated SNP with Parkinson's disease. After subset summary statistics were generated, fixed-effect meta-analyses were used to generate aggregate summary statistics and quantify heterogeneity across subsets for all 39 replication SNPs on the array. Of note, regions within 500 kb of these SNPs were not included in the sample quality control procedure, and these 39 SNPs were manually clustered to evaluate the quality of genotyping.

Risk profiling methods. Risk profiles were generated incorporating three groups of SNPs. The first group included genome-wide significant index SNPs (or their proxies) from the discovery phase that replicated in the independent replication phase ($n = 22$). The second group comprised conditional SNPs that were validated in the replication phase ($n = 4$). The third group consisted of previously reported SNPs that did not quite reach genome-wide significance in the discovery phase but provided evidence of association in the replication phase ($n = 2$). Please see **Supplementary Table 6** for SNP frequencies. Risk profiles were generated using weights based on effect estimates from the discovery phase of this study, using methodologies described in detail elsewhere[3,4,11,35]. In brief, genetic risk scores were scaled on a per-SNP basis using effect estimates from the discovery phase and then applied to the genotype data generated for the samples in the replication phase to create the data set for the analysis of risk profiles. Within each subset of the replication sample series, overall trend tests for Parkinson's disease risk were evaluated using logistic regression, with the risk profile score predicting affected status adjusted for the first two eigenvectors from principal-components analysis as well as for age and sex. Each subset was also divided into quintiles on the basis of risk profile scores, and similar logistic regression models were used to estimate the risk associated with each of the four higher risk quintiles compared to the lowest risk quintile. We performed meta-analysis on all risk profile summary statistics across subsets, using random effects to account for effect heterogeneity. To evaluate the clinical predictability of Parkinson's disease, all risk profiles were combined into one model across the replication subsets, adjusting for age, sex and cohort/subset membership in receiver-operator-curve analyses.

Expression and methylation quantitative trait locus methods. Overlapping SNPs identified in the discovery phase (**Tables 1** and **2**) that were successfully genotyped or imputed in the combined NABEC and UKBEC data sets were tested for association with proximal expression and methylation levels (GSE36192). The allelic dosages of SNPs of interest were tested for association with all methylation and expression probes within 1 Mb of each SNP using linear regression adjusting for covariates of sex, age at death, the first two component vectors from multidimensional scaling, postmortem interval (PMI), brain bank and the batch in which preparation or hybridization was performed, using mach2qtl v1.11 (see URLs). These analyses were run separately for frontal cortex and cerebellar regions, each with analyses focusing on methylation (292 samples) or expression (399 samples). For these four analyses, significance was based on standard FDR adjustments for multiple testing. For further details on consortia membership, acknowledgments and full methods for the expression and methylation portions of this study, please see the **Supplementary Note**.

Expression and methylation methods. Frozen frontal cortex and cerebellar samples were obtained from >399 self-reported European-ancestry samples without determinable neuropathological evidence of disease[26,36]. Genomic DNA was extracted with phenol-chloroform. Bisulfite-converted DNA was assayed at >27,000 sites on the Illumina Infinium HumanMethylation27 BeadChip. mRNA expression levels were assayed using the Illumina HumanHT-12 v3 Expression BeadChip. In brief, individual probes were excluded from analyses if the P value for detection was >0.01 or there was less than 95% completeness of data per probe, and samples were excluded if <95% of probes were detected. Probes were also removed if an analyzed SNP mapped within the probe or if the probe mapped ambiguously to multiple locations in the genome. Expression data were cubic spline normalized and $\log_2$ transformed before analyses.

Each tissue sample was genotyped using the Illumina HumanHap550 v3, Human610-Quad v1 or Human660W-Quad v1 Infinium BeadChip, and shared SNPs were extracted before quality control and imputation. Standard GWAS quality control was undertaken with inclusion criteria such as a minimum call rate of 95% for both participants and SNPs, a minor allele frequency of >0.01, a Hardy-Weinberg equilibrium P value of $>1 \times 10^{-7}$, no first-degree relatives in the sample collection (identity-by-descent score < 0.125 in PLINK) and European ancestry confirmed by multidimensional scaling analyses.

Data were imputed using Minimac (see URLs) to the most recent data freeze of 1000 Genomes Project haplotypes (see URLs), using default settings. All imputed SNPs were filtered for a minimum imputation quality score of 0.30. After quality control, data were available for >10 million SNPs, with expression data on 399 samples (9,814 probes from the frontal cortex and 9,587 probes from the cerebellum), and methylation data were available on 292 samples (27,465 CpG sites in the frontal cortex tissue samples and 27,419 CpG sites in the cerebellum).

Linear regression models were used to estimate associations between the allele dosages of each SNP and gene expression or methylation levels adjusted for covariates of sex, age at death, the first two component vectors from multidimensional scaling, PMI, brain bank and the batch in which preparation or hybridization was performed, using mach2qtl v1.11. Analyses were carried out separately for each brain region and each array type. Only probes within 1 Mb of each SNP of interest were analyzed to test only *cis* associations. From these analysis results, data were mined for the 28 replicated SNPs of interest included in **Tables 1** and **2**.

27. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.* **26**, 657–686 (2011).
28. Ton, T.G. *et al. Post hoc* Parkinson's disease: identifying an uncommon disease in the Cardiovascular Health Study. *Neuroepidemiology* **35**, 241–249 (2010).
29. Ikram, M.A. *et al.* Genomewide association studies of stroke. *N. Engl. J. Med.* **360**, 1718–1728 (2009).
30. Eriksson, N. *et al.* Genetic variant sassociated with breast size also influence breast cancer risk. *BMC Med. Genet.* **13**, 53 (2012).
31. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
32. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
33. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
34. Grove, M.L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS ONE* **8**, e68095 (2013).
35. Ripatti, S. *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* **376**, 1393–1400 (2010).
36. Hernandez, D.G. *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.* **20**, 1164–1172 (2011).