



Published in final edited form as:

Nat Genet. 2009 December ; 41(12): 1308–1312. doi:10.1038/ng.487.

Genome-Wide Association Study reveals genetic risk underlying Parkinson's disease

Javier Simon-Sanchez^{1,2,#}, Claudia Schulte^{3,#}, Jose M Bras^{1,4,#}, Manu Sharma^{3,#}, J Raphael Gibbs^{1,5}, Daniela Berg³, Coro Paisan-Ruiz⁵, Peter Lichtner⁶, Sonja W Scholz^{1,5}, Dena G Hernandez^{1,5}, Rejko Kruger³, Monica Federoff¹, Christine Klein⁷, Alison Goate⁸, Joel Perlmutter⁸, Michael Bonin⁹, Michael A Nalls¹, Thomas Illig¹⁰, Christian Gieger¹⁰, Henry Houlden⁵, Michael Steffens¹¹, Michael S. Okun¹², Mark Cookson¹, Kelly D Foote¹², Hubert H Fernandez¹², Bryan J. Traynor¹, Stefan Schreiber¹³, Sampath Arepalli¹, Ryan Zonoz¹, Katrina Gwinn¹⁴, Marcel van der Brug¹, Grisel Lopez¹⁵, Stephen J Chanock¹⁶, Arthur Schatzkin¹⁶, Yikyung Park¹⁶, Albert Hollenbeck¹⁷, Jianjun Gao¹⁸, Xuemei Huang¹⁹, Nick W Wood⁵, Delia Lorenz²⁰, Gunther Deuschl²⁰, Honglei Chen¹⁸, Olaf Riess⁹, John A Hardy⁴, Andrew B Singleton^{1,CA}, and Thomas Gasser^{3,CA}

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

^{CA}Please address correspondence to Dr. Thomas Gasser (thomas.gasser@uni-tuebingen.de) and Dr. Andrew Singleton (singletona@mail.nih.gov).

[#]these authors contributed equally

Written informed consent was obtained for all participating subjects. Samples derived from the Laboratory of Neurogenetics, NIA, NIH, USA were approved by the Institutional Review Board of the National Institute on Aging: 2003-081 and 2008-146. The specific approval committees for the German samples were as follows: KORA: Bayerische Arztekammer; Popgen/Kiel: Ethik-Kommission der Medizinischen Fakultät der Christian-Albrechts-Universität zu Kiel; Tübingen: Ethik-Kommission der Medizinischen Fakultät und am Universitätsklinikum Tübingen; Lubeck: Ethikkommission der Medizinischen Fakultät der Universität zu Bochum: Ethik-Kommission der Medizinischen Fakultät der Ruhr-Universität Bochum; München: Ethikkommission der Med. Fakultät der LMU München. Samples collected in the United Kingdom received approval from the Joint Research Ethics Committee of the National Hospital and Institute of Neurology. Samples collected from the Parkinson's, Genes & Environment (PAGE) study received approval by the Institutional review Board of NIEHS: 06-E-N093.

All authors participated in critical revision of the manuscript for intellectual content.

AUTHOR CONTRIBUTIONS

J.S.S. performed genotyping, conducted the statistical analyses and participated in writing the manuscript, C.S. conducted the statistical analyses and participated in writing the manuscript, J.M.B. performed genotyping, conducted the statistical analyses and participated in writing the manuscript, M.S. conducted the statistical analyses and participated in writing the manuscript, J.R.G. conducted the statistical analyses, D.B. contributed samples and collected phenotypic data, C.P.-R. contributed samples, performed genotyping, P.L. performed genotyping, S.W.S. performed genotyping, D.G.H. performed genotyping, R.K. contributed samples and collected phenotypic data, M.F. performed genotyping, C.K. contributed samples and collected phenotypic data, A.G. contributed samples, J.P. contributed samples and collected phenotypic data, M.B. performed genotyping, M.A.N. conducted the statistical analyses, T.I. contributed samples and collected phenotypic data, C.G. contributed samples and collected phenotypic data, H.H. contributed samples and collected phenotypic data, M.S. conducted the statistical analyses, M.S.O. contributed samples and collected phenotypic data, M.C. performed eQTL analysis, K.D.F. contributed samples, H.H.F. contributed samples, B.J.T. performed eQTL analysis, S.S. contributed samples and collected phenotypic data, S.A. performed genotyping, R.Z. performed genotyping, K.G. contributed samples and collected phenotypic data, M. van der B. performed eQTL analysis, G.L. contributed samples, S.J.C. contributed control samples, A.S. contributed samples, Y.P. contributed samples, A.H. contributed samples, J.G. contributed samples, X.H. contributed samples, N.W.W. contributed samples and collected phenotypic data, D.L. contributed samples and collected phenotypic data, G.D. contributed samples and collected phenotypic data, H.C. contributed samples and collected phenotypic data, O.R. obtained funding and supervised genotyping, J.A.H. contributed samples and collected phenotypic data, A.B.S. designed and supervised the study, contributed samples and collected phenotypic data and participated in writing the manuscript, T.G. designed and supervised the study, contributed samples and collected phenotypic data and participated in writing the manuscript

¹Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA ²Department of Clinical Genetics Section of Medical Genomics VU Medical Center De Boelelaan 1085 1081 Amsterdam, the Netherlands ³Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and German Center for Neurodegenerative Diseases, Tübingen, Germany ⁴Center for Neurosciences and Cell Biology, University of Coimbra, Coimbra, Portugal ⁵Department of Molecular Neuroscience and Reta Lila Weston Laboratories, Institute of Neurology, University College London, London, United Kingdom ⁶Institute of Human Genetics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany ⁷Section of Clinical and Molecular Neurogenetics at the Department of Neurology, University of Luebeck, Germany ⁸Department of Psychiatry; Department of Neurology, Washington University School of Medicine, St. Louis, Missouri, USA ⁹Department of Medical Genetics, Institute of Human Genetics, University of Tübingen, Tübingen, Germany ¹⁰Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany ¹¹Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany ¹²Movement Disorders Center, University of Florida, Gainesville, Florida, USA ¹³Institut für Klinische Molekularbiologie, Christian-Albrechts-Universität Kiel, Germany ¹⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Texas, USA ¹⁵Parkinson's disease clinic, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA ¹⁶Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA ¹⁷AARP, Washington DC, USA ¹⁸Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, North Carolina, USA ¹⁹Departments of Neurology, Radiology, Neurosurgery, Pharmacology, Kinesiology & Bioengineering, Pennsylvania State University-Milton S. Hershey Medical Center, Hershey, Pennsylvania, USA ²⁰Klinik für Neurologie, UK-SH, Campus Kiel, Christian-Albrechts-Universität Kiel, Kiel, Germany

Abstract

We performed a genome-wide association study (GWAS) in 1,713 Caucasian patients with Parkinson's disease (PD) and 3,978 controls. After replication in 3,361 cases and 4,573 controls, two strong association signals were observed: in the α -synuclein gene (*SNCA*) (rs2736990, OR=1.23, $p=2.24 \times 10^{-16}$) and at the *MAPT* locus (rs393152, OR=0.77, $p=1.95 \times 10^{-16}$). We exchanged data with colleagues performing a GWAS in Asian PD cases. Association at *SNCA* was replicated in the Asian GWAS1, confirming this as a major risk locus across populations. We were able to replicate the effect of a novel locus detected in the Asian cohort (*PARK16*, rs823128, OR=0.66, $p=7.29 \times 10^{-8}$) and provide evidence supporting the role of common variability around *LRRK2* in modulating risk for PD (rs1491923, OR=1.14, $p=1.55 \times 10^{-5}$). These data demonstrate an unequivocal role for common genetic variability in the etiology of typical PD and suggest population specific genetic heterogeneity in this disease.

Advances in genotyping technology have allowed rapid genome-wide screening of common variants in large populations launching a new era in the investigation of the genetic basis of

complex diseases. So far GWAS have contributed little in the field of PD2–4; likely because previous studies lacked power to detect effects of the size expected in these diseases.

The present study was designed as a two-stage GWAS. Characteristics of the cohorts are shown in Table 1. For stage I, genotyping was performed using Infinium BeadChips (Illumina, Inc.). Following quality control 463,185 SNPs were analyzed in 1,713 PD cases and 3,978 controls. To assess the homogeneity of our cohort, pair-wise Identity by State distances (IBS) were calculated using HapMap data as a reference. These analyses revealed that our samples share common Caucasian ancestry (supplementary figures 1 and 2). We chose not to use genomic control as false positive association, possibly caused by population substructure ($\lambda=1.17$), would be controlled for by our two-stage design. Power calculations showed our sample had 80% power to detect variants conferring an odds ratio (OR) of 1.3 with an allele frequency of 10% (supplementary figure 3).

Each SNP was tested for association using a Cochran-Armitage genotypic trend model. Four SNPs on chromosome 4q22 within the *SNCA* locus exceeded Bonferroni corrected genome-wide significance threshold in stage I (most significant $p=5.69\times10^{-9}$, rs2736990; Figure 1, Table 2). Three SNPs at the *MAPT* locus on chromosome 17q21 also surpassed genome-wide significance in stage I (most significant $p=5.05\times10^{-8}$, rs199533). A logistic regression analysis adjusted for the two first components of the multidimensional scaling values following calculation of pair-wise IBS distance suggests that the significant results obtained are not biased by population substructure (supplementary material).

Replication comprised genotyping of 384 SNPs selected based on the p -value observed in stage I (supplementary table 1). Genotyping was performed in an independent cohort of 3,452 cases and 4,756 controls from the US, Germany and Britain. Taking into account the results obtained from pairwise Identity by State distances calculations and considering that genetic heterogeneity and allelic heterogeneity are not likely to produce type I and type II errors when pooling white North American and white North European populations, we decided to analyze all Stage II samples together. Following quality control filtering, 345 SNPs were analyzed in 3,361 cases and 4,573 controls. Twenty-one SNPs within the *SNCA* and *MAPT* loci surpassed Bonferroni threshold for significance ($p<0.000145$; Table 2; supplementary table 2).

Three hundred and forty five stage II SNPs passed our quality control procedures (Supplementary table 1). Notably, we observed clusters of SNPs showing improved association signals when combining our stage I and stage II datasets (supplementary figure 4). Although some of these SNPs are at loci that contain biologically plausible candidate genes for PD, they do not reach genome-wide significance and thus we have resisted drawing conclusions from these data; however, of particular note is a cluster of 7 SNPs in chromosome 10q24.32, with p -values below 1×10^{-3} (Supplementary figure 5). These and other variants showing a consistent, moderate association across stages warrant independent replication.

To further delineate the signals on *SNCA* and *MAPT*, allelic association of significant SNPs was tested, conditioned on alleles of other significant SNPs at the same locus. No

independent signals were identified, suggesting that a single detectable pathogenic variant exists at each locus (supplementary table 3); however without complete sequence data across these loci we cannot rule out independent effects. We did not find evidence for epistasis between *SNCA* and *MAPT* risk alleles (supplementary material).

Analysis of the linkage disequilibrium (LD) structure across the *SNCA* locus revealed two blocks of LD (Figure 2A). The 3' block contains three of the four significantly associated SNPs, suggesting that the causal variant is located in the 3' region of the *SNCA* gene. This is strengthened by analysis of the haplotype frequencies at this locus (supplementary figure 6) and previous studies⁵. The REP1 microsatellite in *SNCA* was previously associated with PD5 and its pathological effect has been suggested to be mediated by gene expression⁶. Analysis of REP1 genotype data in 1,774 samples from the US cohort revealed the risk allele of REP1 is in LD with the 3' risk alleles identified here ($r^2=0.365$ with rs3857059; supplemental figure 7A), thus the association identified at the REP1 locus and the SNPs identified here may be the result of residual LD between these loci. This is supported by logistic regression analysis conditioned on REP1 genotypes, showing that association at REP1 is not independent from the association identified here (supplementary material). We recently reported a significant association of *SNCA* SNPs with another synucleinopathy, multiple system atrophy ($p=5.5\times 10^{-12}$, MSA)⁷; comparison of these data reveals disparate *SNCA* risk SNPs in MSA and PD, a finding that may shed light on the exact pathogenic substrate and molecular etiology of these disorders (supplementary table 4).

As expected, one large highly inter-correlated block of high LD was observed across the *MAPT* locus (Figure 2B). Available genotype data of the H1/H2 haplotypes in this region showed that the risk alleles of the associated SNPs are in LD with the H1 haplotype ($r^2=0.761$ with rs393152; supplementary figure 7B). It is unclear from the current data whether the *MAPT* risk haplotype identified here corresponds to the subhaplotype associated with corticobasal degeneration (CBD) and progressive supranuclear palsy (PSP)⁸. Because of the LD structure we cannot rule out other genes at this locus as the pathogenically relevant genes; however, *MAPT* is biologically the most plausible candidate.

Following data exchange with colleagues performing a PD GWAS in Japan we chose to study two loci implicated in Asian PD on chromosomes 1q32 and 4p15. In our stage I data, the most significant p -values at these 12 SNPs were 1.3×10^{-4} and 6.5×10^{-3} (1q32 and 4p15.3 respectively). The signal at 1q32 was significant enough to carry through to stage II replication, but this SNP had been excluded because of the low minor allele frequency in controls (0.03). Genotyping of these 12 SNPs was performed in an available subset of our replication cohort comprising 2,816 cases and 3,401 controls. The signal on chromosome 1q32 was replicated in this cohort (rs823128, $p=1.9\times 10^{-4}$; Table 2). While this failed to surpass Bonferroni correction, the p -value across stages was highly significant (rs823128, $p=7.3\times 10^{-8}$) and it is worth noting that the significance improved for all SNPs at this locus when combining stage I and II results (Figure 2C). For these reasons and because the association at this locus was consistently detected in the Asian cohort¹ we are confident this signal represents a true association and this has been designated *PARK16*. Although we failed to replicate the signal on chromosome 4p15, which included only one gene, *BST1*, the

low minor allele frequency of the associated SNPs in Caucasian individuals may have affected our ability to observe association.

Mutations in *SNCA* and *MAPT* have been associated with autosomal dominant forms of parkinsonism^{9,10}. Given this, it is interesting that we observed association proximal to *LRRK2*, which also contains mutations causing autosomal dominant PD^{11,12}. In stage I 23 SNPs upstream of *LRRK2*, and 12 SNPs within *LRRK2* were associated with PD (lowest $p=5.03\times 10^{-6}$ in rs2896905, located in *SLC2A13*, 0.27Mb from *LRRK2*). Of these, 3 SNPs surpassed our threshold for replication and were analyzed in stage II. Only one SNP, 0.17Mb upstream of *LRRK2*, remained associated with PD after stage II (rs1491923, $p=0.01$; supplementary material; Figure 2D). While this did not surpass our threshold for multiple testing, the combined stage I and II p -values revealed a compelling association ($p=1.6\times 10^{-5}$). Interestingly, the other 2 replicated SNPs at this locus were also nominally associated with PD after combining stage I and stage II datasets (p -values of 9.5×10^{-5} and 7.8×10^{-3} ; rs11564612 and rs2896905 respectively). Data from the Asian cohort also revealed a significant association with PD at this locus¹. *SLC2A13*, neighboring *LRRK2*, cannot be ruled out as the gene of effect at this locus, however, *LRRK2* is clearly a more plausible candidate.

Although mutations and copy number variants of *SNCA* are the cause of rare familial forms of PD^{10,13}, association of common variants has been more controversial. This study provides unequivocal evidence that variation in *SNCA* contributes to the etiology of sporadic PD. The clustering of associated SNPs in the 3' UTR suggests that the causal variant might affect post-transcriptional RNA processing or RNA stability, possibly mediated by miRNA binding sites or by alternative splicing.

Association of the H1 haplotype at the *MAPT* locus with PSP and CBD has been described and replicated⁸; however, association studies of variants at *MAPT* in PD have produced conflicting results^{14,15}. Our data provide unequivocal evidence for an association of the *MAPT* locus with PD. This is surprising given the classic separation of synucleinopathies and tauopathies, although a cross-talk between molecular pathways characterized by different aggregating proteins has been repeatedly suggested. While there are additional genes at the *MAPT* locus, the role of *MAPT* in neurodegenerative diseases is well established and this association is biologically plausible.

We provide evidence for an association of PD with variability proximal to *LRRK2* and at a novel locus at 1q32 (*PARK16*). The kinase activity of *Lrrk2* is an attractive therapeutic target; our data suggests that this protein is also relevant to the etiology in sporadic PD patients without mutations. The *PARK16* locus spans 5 transcripts, *SNORA72*, *NUCKS1*, *RAB7L1*, *SLC41A1* and *PM20D1*. It will be important to define the immediate biological consequences of all four risk loci identified here. It is notable that three of the most significant loci identified here contain genes known to be mutated in Mendelian forms of parkinsonism. This supports the notion that rare familial disease is etiologically related to typical sporadic PD, and suggests that genes that contain common risk variants are excellent candidates to contain rare disease causing mutations. One might also predict that deep sequencing of these loci will reveal rare mutations that alter risk for, rather than cause,

disease. It is also interesting that two of the four loci discussed here, are risk factors for other neurodegenerative diseases, including MSA (*SNCA*), PSP (*MAPT*) and CBD (*MAPT*).

The combined population attributable risk associated with the identified loci, considering the genotypic counts of those most associated SNPs in our cohort, is approximately 25% (supplemental methods). Our study was a retrospective case-control study and the frequency of the risk variants detected might not reflect the frequencies of true causal variants; thus these values should be interpreted with caution.

In an attempt to define a biological consequence of risk variants, we mined data produced within our laboratory. In this work genome wide genotyping and expression profiling of >22,000 transcripts was performed in 133 human frontal cortex samples, allowing us to determine SNPs significantly associated with expression level. These data revealed a strong association between genotype at the risk alleles of the *MAPT* locus and expression levels of both *MAPT* and *LRR37A* (figure 3), in addition to less significant association with expression levels of other transcripts at this locus, but did not reveal association between risk SNPs and expression levels of proximal genes at the *SNCA*, *LRRK2* or *PARK16* loci (Supplementary table 5, figure 8). Notably the alleles at the *MAPT* locus associated with increase risk of PD are associated with increased expression of *MAPT* in the human brain.

We show for the first time a clear role for common genetic variability in the risk of developing PD. Further we describe a possible population specific genetic heterogeneity in this disorder, since the association to *MAPT* was absent in the Asian cohort. This observation has potential implications for the analysis of complex traits across populations; such genetic heterogeneity, particularly at minor risk loci, has the potential to mask true associations when analyses are performed across populations. With the discovery of the *PARK16* locus in the Asian population, this highlights the power of comparing GWAS across different populations. A further increase in the number and size of cohorts for GWAS in PD will likely reveal additional common genetic risk loci and these, in turn, will improve understanding and ultimately treatment of this devastating disorder.

METHODS

Methods and associated references appear online. *Note:* Supplementary information is available on the *Nature Genetics* website.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the subjects involved in this study whose contribution made this work possible. This work used samples and clinical data from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>). This work was supported in part by the Intramural Research Programs of the National Institute on Aging, the National Institute of Neurological Disorders and Stroke, the National Institute of Environmental Health Sciences, the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; project numbers Z01 AG000949-02 and Z01-ES101986. The KORA research platform (KORA: Cooperative Research in the Region of Augsburg; www.gsf.de/KORA) was initiated and financed by the

GSF—National Research Centre for Environment and Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria.

The study was additionally funded by the German National Genome Network (NGFNplus #01GS08134; German Ministry for Education and Research) and in addition by the German BMBF NGFN (01GR0468). This work also was supported by NIH NINDS P30NS05710 (Neuroscience Blueprint Grant) and Clinical Sciences Translational Award RR024992 to Washington University in St. Louis and the Greater St. Louis Chapter of the American Parkinson Disease Association. Authors received support from the Medical Research Council, UK.

References

1. Toda T. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature Genetics*. 2009 in press.
2. Fung HC, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol*. 2006; 5:911–6. [PubMed: 17052657]
3. Maraganore DM, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet*. 2005; 77:685–93. [PubMed: 16252231]
4. Pankratz N, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet*. 2009; 124:593–605. [PubMed: 18985386]
5. Maraganore DM, et al. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *Jama*. 2006; 296:661–70. [PubMed: 16896109]
6. Chiba-Falek O, Nussbaum RL. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum Mol Genet*. 2001; 10:3101–9. [PubMed: 11751692]
7. Scholz SW, et al. SNCA variants are associated with increased risk of multiple system atrophy. *Annals of Neurology*. 2009 in press.
8. Pittman AM, et al. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet*. 2005; 42:837–46. [PubMed: 15792962]
9. Hutton M. Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms. *Neurology*. 2001; 56:S21–5. [PubMed: 11402146]
10. Polymeropoulos MH, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*. 1997; 276:2045–7. [PubMed: 9197268]
11. Paisan-Ruiz C, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron*. 2004; 44:595–600. [PubMed: 15541308]
12. Zimprich A, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*. 2004; 44:601–7. [PubMed: 15541309]
13. Singleton AB, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science*. 2003; 302:841. [PubMed: 14593171]
14. Johansson A, Zetterberg H, Hakansson A, Nissbrandt H, Blennow K. TAU haplotype and the Saitohin Q7R gene polymorphism do not influence CSF Tau in Alzheimer's disease and are not associated with frontotemporal dementia or Parkinson's disease. *Neurodegener Dis*. 2005; 2:28–35. [PubMed: 16909000]
15. Skipper L, et al. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet*. 2004; 75:669–77. [PubMed: 15297935]
16. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992; 55:181–4. [PubMed: 1564476]
17. The International HapMap Project. *Nature*. 2003; 426:789–96. [PubMed: 14685227]
18. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]
19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–5. [PubMed: 15297300]

20. Gabriel SB, et al. The structure of haplotype blocks in the human genome. *Science*. 2002; 296:2225–9. [PubMed: 12029063]
21. Baker M, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet*. 1999; 8:711–5. [PubMed: 10072441]
22. Scholz S, et al. SNCA variants are associated with increased risk of multiple system atrophy. *Annals of Neurology*. 2009 in press.
23. van der Brug MP, et al. RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways. *Proc Natl Acad Sci U S A*. 2008; 105:10244–9. [PubMed: 18626009]

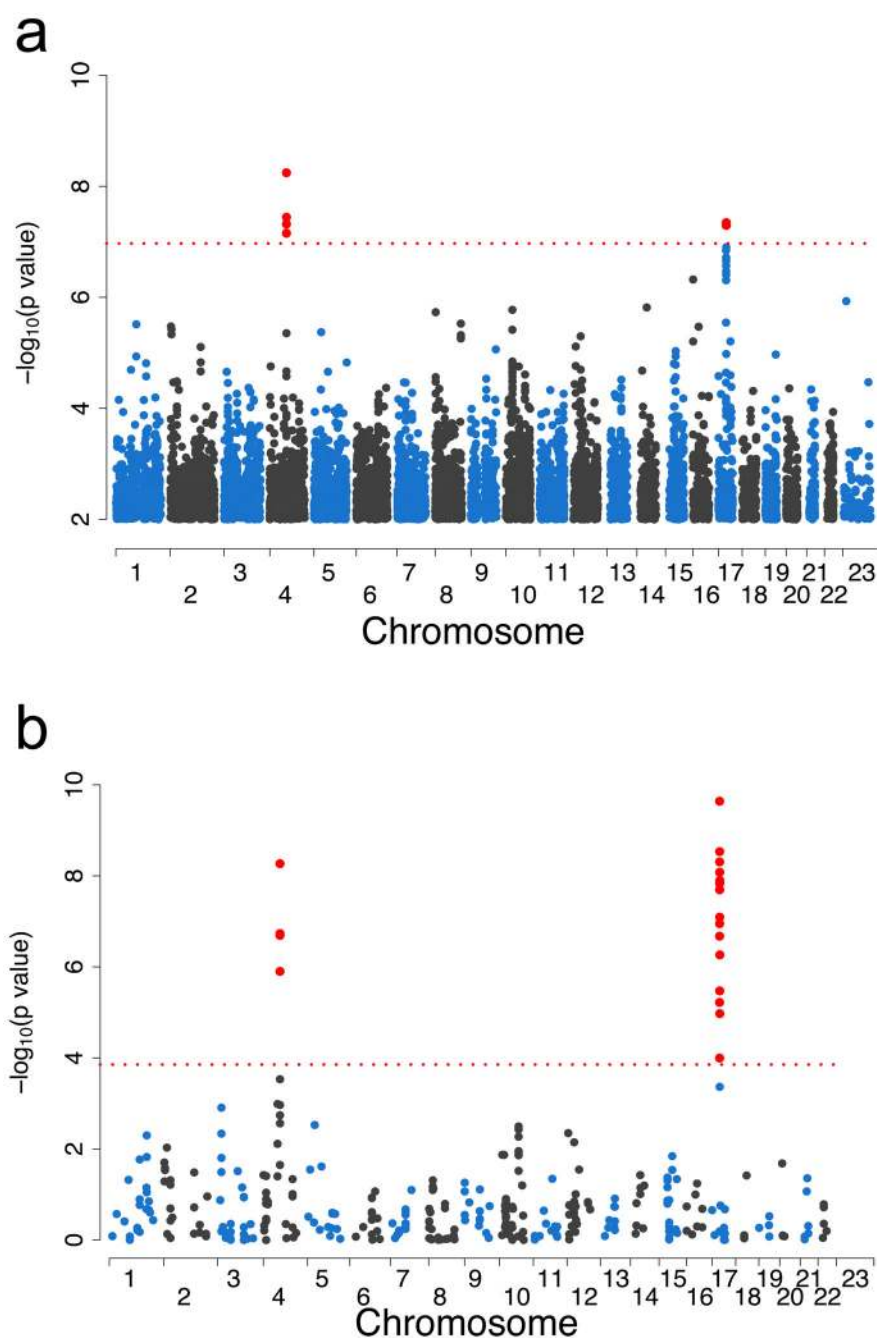
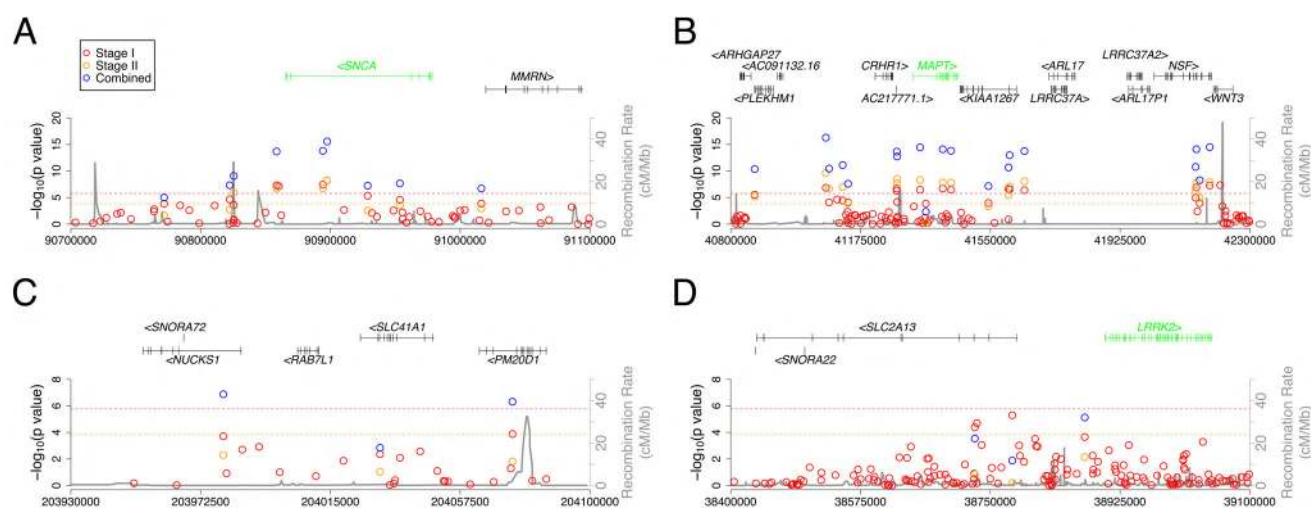
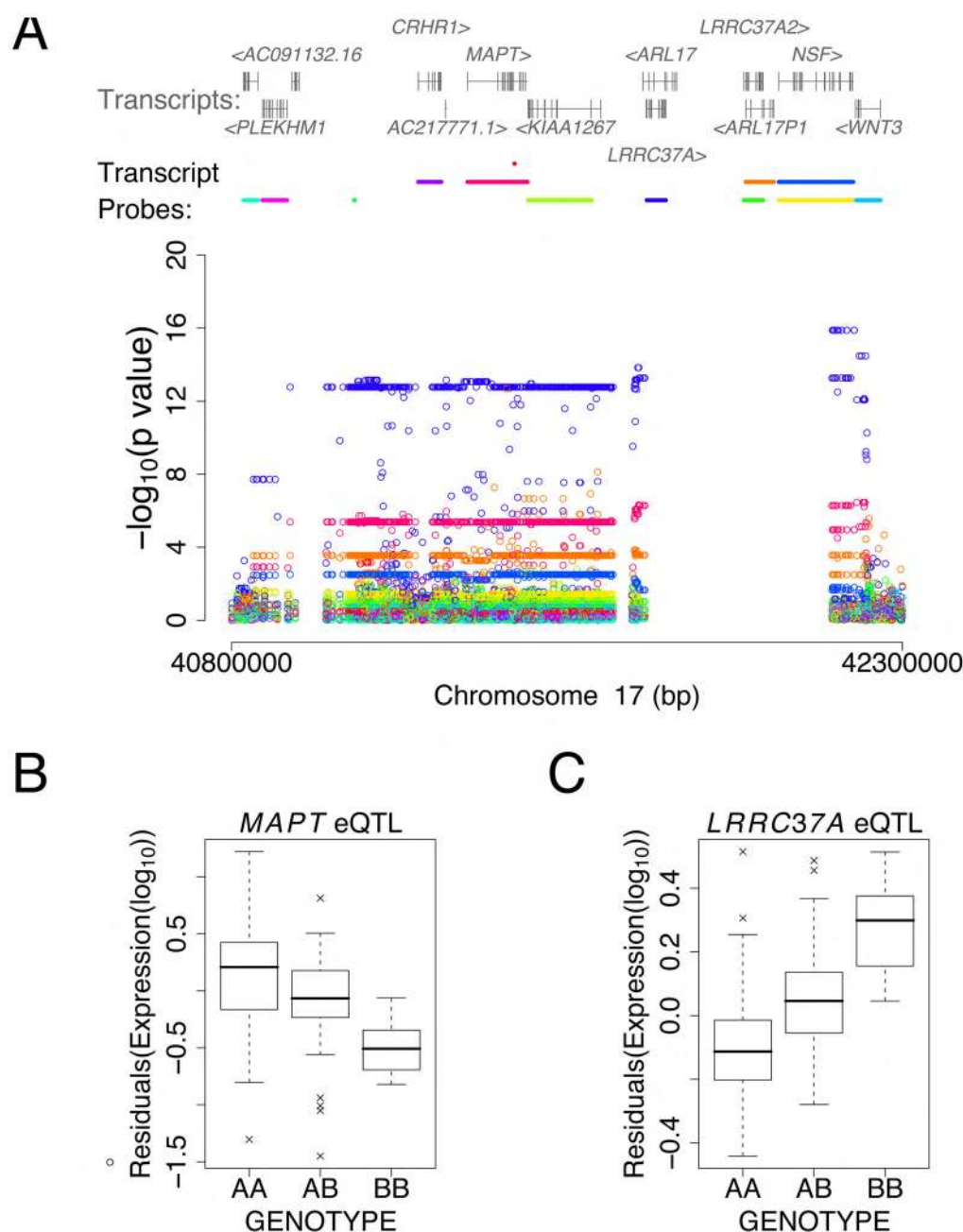


Figure 1.

Graphical representation of p values in stage I and stage II. A) In stage I, p values are log transformed (y-axis) and plotted against chromosomes (x-axis). The red line indicates the Bonferroni threshold. Signals indicated in red are on chromosome 4 and chromosome 17 and surpass Bonferroni threshold for genome wide significance. B) log transformed p-values of Stage II SNPs (y-axis) are plotted against chromosomes (x-axis). Signals indicated in red are on chromosome 4 and chromosome 17 and surpass Bonferroni threshold for multiple testing

**Figure 2.**

Association and recombination rates across *SNCA*, *MAPT*, *PARK16* and *LRRK2*. $-\log_{10} p$ values are shown for stage I and II analyses, annotated transcripts are shown across the top of each plot. Red dotted line indicates threshold for genome wide significance in stage I and orange line indicates threshold for significance of stage II.

**Figure 3.**

Expression quantitative trait loci across the *MAPT* locus measured in 133 human frontal cortex samples; panel A shows association between genotypes and transcript levels across the *MAPT* locus. In this analysis the allelic load at genotyped polymorphisms across the locus is tested for association with transcript levels of each gene across the locus. The results of the analysis are shown as log transformed p values color-coded to match the transcript of interest. Notably genotypes across this locus are associated with *MAPT* (pink) and *LRRC37A* (blue) levels. Boxplots in panels B and C illustrate a dose relationship between allele load at the most significantly associated PD SNP (rs393152) and expression of *MAPT* (B) and

LRRC37A (C), p values 4.1×10^{-6} and 1.7×10^{-13} respectively. Notably, the allele associated with higher risk for PD, A, is associated with high levels of *MAPT* expression and low levels of *LRRC37A* expression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Study characteristics of cases and controls in stage I and II

		cases				controls			
		Sample size	aao mean (s.d.)	male/female ratio	Sample size	aae mean (s.d.)	male/female ratio		
Stage I	USA	988	55.9 (15.1)	1.09	3071	62 (15.6)	0.96		
	Germany	757	56 (11.64)	1.49	976	NA	1.08		
Stage II	USA	1528	62.5 (8.55)	2.44	2044	63 (15.6)	2.45		
	Germany	1100	61 (11.32)	1.37	2168	57 (10.54)	1.4		
	UK	824	59 (12.3)	3.5	544	NA	0.57		

aao = age at onset; aae = age at examination; s.d. = standard deviation; NA = not available

Table 2

The top three SNPs of the 2 loci that surpass Bonferroni threshold for multiple testing in both stages and the top three SNPs in the LRRK2 locus and the additional loci in chromosomes 1 and 4.

locus		Stage I				Stage II				Stage I + II combined					
SNP	Chr.	Position	Alleles (minor/major)	MAF_U	MAF_A	p-value	cOR	MAF_U	MAF_A	p value	cOR	MAF_U	MAF_A	p value	cOR
Genome-wide significant loci															
rs393152	17	41074926	G/A	0.22	0.18	1.42E-07	0.76	0.22	0.18	1.69E-09	0.78	0.18	0.22	1.95E-16	0.77
rs199533	17	42184098	T/C	0.20	0.16	5.05E-08	0.75	0.20	0.17	7.95E-08	0.79	0.17	0.20	1.09E-14	0.78
rs17563986	17	41347100	C/T	0.22	0.18	3.44E-07	0.77	0.21	0.18	3.21E-08	0.79	0.18	0.22	1.67E-14	0.78
rs2736990	4	90897564	C/T	0.46	0.52	5.69E-09	1.27	0.46	0.51	5.52E-09	1.20	0.51	0.46	2.24E-16	1.23
rs3857059	4	90894261	G/A	0.07	0.10	3.60E-08	1.49	0.07	0.10	3.39E-08	1.43	0.10	0.07	3.74E-15	1.48
rs11931074	4	90858538	T/G	0.07	0.10	4.78E-08	1.49	0.08	0.10	3.75E-08	1.42	0.10	0.07	1.62E-14	1.46
Other loci															
rs823128	1	203980001	C/T	0.04	0.03	1.90E-04	0.64	0.04	0.03	1.96E-04	0.68	0.03	0.04	7.29E-08	0.66
rs11240572	1	204074636	T/G	0.04	0.02	1.30E-04	0.63	0.04	0.03	1.81E-03	0.72	0.03	0.04	6.11E-07	0.67
rs823156	1	204031263	C/T	0.18	0.16	4.31E-03	0.85	0.17	0.16	0.06	0.93	0.16	0.18	7.60E-04	0.89
rs12646913	4	15348374	G/A	0.08	0.07	0.041	0.85	0.08	0.07	0.32	0.97	0.07	0.08	0.03	0.92
rs4698412	4	15346446	G/A	0.44	0.43	0.085	0.93	0.45	0.43	0.14	0.95	0.43	0.44	0.03	0.94
rs12502586	4	15335662	A/G	0.10	0.12	6.50E-03	1.20	0.10	0.10	0.89	1.00	0.11	0.10	0.07	1.08
rs11564162	12	38729159	C/T	0.21	0.18	4.00E-05	0.78	0.21	0.19	0.10	0.93	0.19	0.21	9.52E-05	0.87
rs2896905	12	38779683	T/C	0.40	0.35	5.03E-06	0.82	0.38	0.38	0.73	1.01	0.37	0.39	7.81E-03	0.93
rs1491923	12	38877384	C/T	0.31	0.34	2.20E-04	1.20	0.31	0.33	0.01	1.10	0.33	0.31	1.55E-05	1.14

Position from NCBI genome build 36; Chr. = Chromosome; MAF U = minor allele frequency in controls; MAF A = minor allele frequency in cases; p-value = nominal p-value of trend test; cOR = common OR.