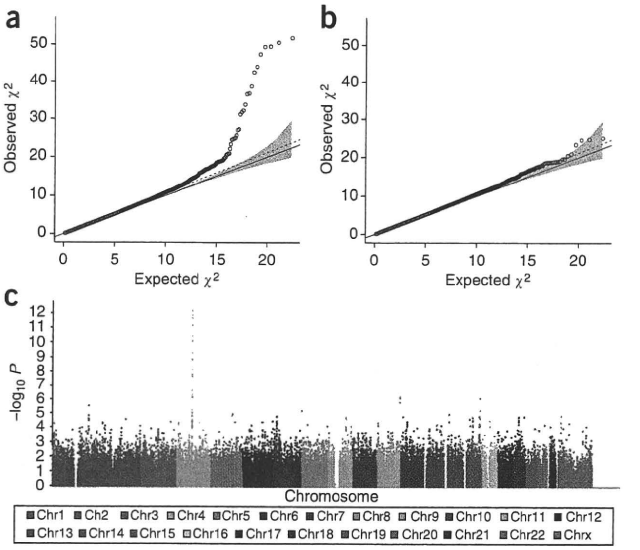

Figure 1 Genome-wide association results from the discovery phase. (**a**) Quantile-quantile plot for test statistics (Cochran-Armitage trend test) for 435,470 SNPs passing quality control. The solid line represents concordance of observed and expected values. Slope of the dashed line represents the genomic inflation factor ($\lambda = 1.055$). The shaded region is the 95% concentration band formed by calculating, for each order statistic, the 2.5th and 97.5th percentiles of the respective distribution under the null hypothesis. (**b**) Quantile-quantile plot for test statistics (Cochran-Armitage trend test) after the removal of the four loci with strong associations in this study (1q32, 4p15, 4q22 and 12q12). (**c**) Manhattan plot presenting the *P* values across the genome. The $-\log_{10} P$ (Cochran-Armitage trend test) from 435,470 SNPs in 988 Parkinson's disease cases and 2,521 controls is plotted according to its physical position on successive chromosomes.

For fast-track replication, we selected the 337 most associated SNPs ($P \leq 0.000533$) from analysis of GWAS data and genotyped them in a sample set of replication 1, which consisted of 612 cases and 14,139 controls from Japan (**Supplementary Note**). Thirty-two SNPs showed association of $P < 0.05$ in replication 1 (**Supplementary Fig. 1**). Combined analyses of the GWAS and replication 1 showed that 12 SNPs in 3 loci (1q32, 4p15 and 4q22) surpassed $P < 5 \times 10^{-7}$. Furthermore, we found association signals ($P = 3.06 \times 10^{-6}$, OR = 1.36) on 12q12, harboring *LRRK2*, which is a causative gene for autosomal dominant parkinsonism (**Table 1**).

In replication 2, we tested 24 SNPs at these four loci for association with PD. An independent sample set (321 cases and 1,614 controls) recruited from Japan was used in replication 2 (**Supplementary Note**). Association evidence was again found at these four loci: 1q32, $P = 2.80 \times 10^{-4}$, OR = 1.37; 4p15, $P = 7.70 \times 10^{-3}$, OR = 1.26; 4q22, $P = 0.02$, OR = 1.22; and 12q12, $P = 6.43 \times 10^{-4}$, OR = 1.57 (**Table 1**). The disease associations on 1q32 and 12q12 exceeded the conservative Bonferroni-corrected threshold for significance ($P = 0.0021$; calculated as 0.05/24). All the SNPs showed allele frequency differences in the same direction in the GWAS, replication 1 and replication 2. Furthermore, combined analysis of the GWAS and two replication stages provided strong evidence of association in the four regions with a significance level of $P = 2.72 \times 10^{-8}$ or less (**Table 1**).

We identified two new susceptibility loci with genome-wide significance on 1q32 and 4p15, which have not been reported to be associated with PD in previous studies[12–14]. On 1q32, seven SNPs (rs16856139, rs823128, rs823122, rs947211, rs823156, rs708730 and rs11240572) reached $P < 5 \times 10^{-7}$ in the overall analysis (**Fig. 2a**). rs947211 showed the strongest association to PD ($P = 1.52 \times 10^{-12}$, OR = 1.30) and is located 8.5 kb upstream of *RAB7L1* and 5.6 kb downstream of *SLC41A1*. Linkage disequilibrium (LD) analysis revealed that SNPs with significant associations to PD lie within several LD blocks containing the following five genes: *SLC45A3*, *NUCKS1*, *RAB7L1*, *SLC41A1* and *PM20D1* (also called *FLJ32569*) (**Table 1** and **Fig. 2a**). Three genes (*NUCKS1*, *RAB7L1* and *SLC41A1*) were contained in the same LD block as rs947211. rs947211 was weakly correlated with the other six SNPs ($r^2 = 0.07$–0.25), and we observed residual association signals when rs947211 and each of the other six SNPs were paired in conditional analyses of our overall data. This result suggests that this locus has multiple independent association signals (**Supplementary Table 1**). These data provide the first evidence, to our knowledge, of an association between 1q32 and PD susceptibility, and we designated this region as *PARK16*.

On 4p15, four SNPs (rs11931532, rs12645693, rs4698412 and rs4538475) reached $P < 5 \times 10^{-7}$ in the combined analysis (**Fig. 2b**). These four SNPs showed strong disease association with almost the same significance levels (ranging from $P = 3.94 \times 10^{-9}$ to $P = 1.78 \times 10^{-8}$, all OR = 1.24); among them, rs4538475 was the most strongly associated. The four SNPs were located from intron 8 to 4.1 kb downstream of *BST1* (bone marrow stromal cell antigen). LD analysis revealed that the four SNPs were correlated with $r^2 > 0.78$ and lie within a 15 kb LD block containing a single gene, *BST1*.

The remaining two intervals (4q22 and 12q12) harbored genes previously found to be causal for autosomal dominant forms of parkinsonism, specifically, *SNCA* and *LRRK2*, respectively. On 4q22, seven SNPs (rs3733449, rs11931074, rs3857059, rs2736990, rs3796661, rs6532194 and rs12233759) throughout the *SNCA* region showed genome-wide significant association in the combined analysis (**Fig. 2c**). The most significantly associated SNPs, rs11931074 ($P = 7.35 \times 10^{-17}$, OR = 1.37) and rs3857059 ($P = 5.68 \times 10^{-16}$, OR = 1.36), are approximately 35.7 kb apart, located 7.2 kb downstream from and in intron 4 of *SNCA*, respectively. The entire *SNCA* gene was divided into two LD blocks at intron 4. Both SNPs were positioned on the 3′ side of the LD block and showed a high degree of LD ($r^2 = 0.98$). Three SNPs (rs2736990, rs3796661 and rs6532194) were moderately correlated with rs11931074 ($r^2 = 0.81$, 0.76 and 0.63, respectively). The remaining two SNPs (rs3733449 and rs12233759) were weakly correlated with rs11931074 ($r^2 = 0.05$ and 0.24, respectively), and residual association signals were marginally observed when rs11931074 and each of these two SNPs were paired in conditional analyses of our overall data (**Supplementary Table 1**). These data confirm *SNCA* as a susceptibility gene for PD.

On 12q12, five SNPs (rs1994090, rs7304279, rs4768212, rs2708453 and rs2046932) surpassed $P < 5 \times 10^{-7}$ in the overall analysis (**Fig. 2d**). The five SNPs showed strong disease association with almost the same significance (ranging from $P = 2.72 \times 10^{-8}$ to $P = 1.09 \times 10^{-7}$, OR = 1.37–1.39); among them, rs1994090 was the most strongly associated to PD. These five SNPs were located from intron 2 of *SLC2A13* to 38.4 kb upstream of *LRRK2*. These SNPs were highly correlated with $r^2 > 0.83$ and were positioned within several LD blocks defined by the method of Gabriel *et al.*[16]. This is the first evidence that common variants proximal to *LRRK2* are associated with PD at genome-wide significance level.

Variants with the highest significance at the four loci detected in this study were common SNPs with risk allele frequencies of 0.50 (rs947211 on 1q32), 0.38 (rs4538475 on 4p15), 0.58 (rs11931074 on 4q22) and 0.08 (rs1994090 on 12q12) (**Table 1**). Population attributable risks for rs947211, rs4538475, rs11931074 and rs1994090 were estimated to be 13%, 8%, 18% and 3%, respectively.

Table 1 Summary of association results for representative SNPs that characterize the association of Parkinson's disease with 1q32 (*PARK16*), 4p15 (*BST1*), 4q22 (*SNCA*) and 12q12 (*LRRK2*)

				GWAS			Replication 1			GWAS + Replication 1		Replication 2			GWAS + Replication 1+2	
Locus	SNP	Pos (Mb)	Allele Minor/ major	MAF Case	MAF Ctrl	P_{trend}	OR (95% CI)	MAF Case	MAF Ctrl	P_{trend}	OR (95% CI)	P_{cmh}	OR (95% CI)	MAF Case	MAF Ctrl	P_{trend}	OR (95% CI)	P_{cmh}	OR (95% CI)
New PD loci																			
1q32	rs16856139	203.91	T/C	0.10	0.14	2.55×10^{-6}	1.50 (1.26–1.77)	0.11	0.13	0.067	1.19 (0.99–1.44)	2.15×10^{-6}	1.35 (1.19–1.54)	0.10	0.13	0.015	1.42 (1.07–1.88)	1.02×10^{-7}	1.36 (1.22–1.53)
(*PARK16*)	rs823128	203.98	G/A	0.10	0.14	2.09×10^{-5}	1.43 (1.21–1.69)	0.11	0.13	0.0056	1.31 (1.08–1.59)	4.67×10^{-7}	1.38 (1.22–1.57)	0.09	0.14	0.0028	1.53 (1.16–2.03)	4.88×10^{-9}	1.41 (1.26–1.58)
	rs823122	203.99	C/T	0.10	0.14	7.98×10^{-5}	1.39 (1.18–1.64)	0.11	0.13	0.013	1.27 (1.05–1.54)	3.87×10^{-6}	1.34 (1.18–1.52)	0.09	0.14	0.0034	1.52 (1.15–2.01)	5.22×10^{-8}	1.37 (1.22–1.54)
	rs947211	204.02	A/G	0.43	0.48	1.15×10^{-4}	1.23 (1.11–1.37)	0.42	0.50	1.35×10^{-6}	1.35 (1.19–1.52)	1.12×10^{-9}	1.28 (1.18–1.38)	0.42	0.50	2.80×10^{-4}	1.37 (1.16–1.63)	1.52×10^{-12}	1.30 (1.21–1.39)
	rs823156	204.03	G/A	0.13	0.17	1.20×10^{-5}	1.40 (1.20–1.62)	0.14	0.17	0.012	1.25 (1.05–1.48)	6.45×10^{-7}	1.33 (1.19–1.49)	0.12	0.17	0.0013	1.52 (1.17–1.95)	3.60×10^{-9}	1.37 (1.23–1.52)
	rs708730	204.04	G/A	0.14	0.18	2.60×10^{-5}	1.37 (1.18–1.59)	0.15	0.17	0.022	1.22 (1.03–1.44)	2.89×10^{-6}	1.30 (1.17–1.46)	0.12	0.18	0.0019	1.48 (1.15–1.89)	2.43×10^{-8}	1.33 (1.21–1.48)
	rs11240572	204.07	A/C	0.13	0.16	1.66×10^{-4}	1.34 (1.15–1.56)	0.13	0.16	0.016	1.24 (1.04–1.48)	9.78×10^{-6}	1.30 (1.16–1.46)	0.12	0.17	0.0024	1.49 (1.15–1.92)	1.08×10^{-7}	1.33 (1.20–1.48)
4p15	rs11931532	15.33	T/C	0.45	0.40	2.75×10^{-4}	1.22 (1.09–1.35)	0.47	0.42	1.86×10^{-4}	1.26 (1.11–1.42)	2.02×10^{-7}	1.23 (1.14–1.34)	0.47	0.41	0.0077	1.26 (1.06–1.49)	5.13×10^{-9}	1.24 (1.15–1.33)
(*BST1*)	rs12645693	15.34	G/A	0.45	0.40	3.06×10^{-4}	1.21 (1.09–1.35)	0.47	0.42	3.00×10^{-4}	1.25 (1.11–1.41)	3.42×10^{-7}	1.23 (1.14–1.33)	0.47	0.41	0.0077	1.26 (1.06–1.49)	8.65×10^{-9}	1.24 (1.15–1.33)
	rs4698412	15.35	A/G	0.38	0.33	5.28×10^{-5}	1.25 (1.12–1.40)	0.40	0.35	4.91×10^{-4}	1.24 (1.10–1.40)	1.03×10^{-7}	1.25 (1.15–1.35)	0.38	0.34	0.055	1.19 (1.00–1.42)	1.78×10^{-8}	1.24 (1.15–1.33)
	rs4538475	15.35	A/G	0.41	0.36	4.05×10^{-5}	1.25 (1.12–1.40)	0.43	0.38	3.48×10^{-4}	1.25 (1.10–1.41)	5.98×10^{-8}	1.25 (1.15–1.35)	0.42	0.37	0.022	1.22 (1.03–1.46)	3.94×10^{-9}	1.24 (1.16–1.34)
Loci located in or near autosomal dominant parkinsonism genes																			
4q22	rs11931074	90.86	G/T	0.32	0.42	6.17×10^{-13}	1.50 (1.34–1.68)	0.36	0.42	2.12×10^{-5}	1.31 (1.16–1.48)	2.19×10^{-16}	1.41 (1.30–1.53)	0.38	0.43	0.034	1.21 (1.01–1.44)	7.35×10^{-17}	1.37 (1.27–1.48)
(*SNCA*)	rs3857059	90.89	A/G	0.32	0.41	1.17×10^{-12}	1.49 (1.34–1.67)	0.36	0.42	6.92×10^{-5}	1.29 (1.14–1.45)	1.54×10^{-15}	1.40 (1.29–1.52)	0.38	0.43	0.041	1.20 (1.01–1.43)	5.68×10^{-16}	1.36 (1.26–1.46)
	rs894278	90.95	G/T	0.43	0.38	7.68×10^{-5}	1.24 (1.11–1.37)	0.39	0.38	0.46	1.05 (0.93–1.18)	4.77×10^{-4}	1.15 (1.07–1.25)	0.42	0.37	0.020	1.22 (1.03–1.45)	3.28×10^{-5}	1.17 (1.09–1.25)
	rs6532194	91.00	C/T	0.31	0.40	6.93×10^{-11}	1.44 (1.29–1.61)	0.36	0.41	0.0014	1.22 (1.08–1.39)	1.77×10^{-12}	1.35 (1.24–1.46)	0.37	0.41	0.040	1.21 (1.01–1.44)	4.15×10^{-13}	1.32 (1.22–1.42)
12q12	rs1994090	38.71	G/T	0.11	0.08	4.45×10^{-5}	1.43 (1.20–1.70)	0.10	0.08	0.018	1.26 (1.04–1.54)	3.06×10^{-6}	1.36 (1.20–1.55)	0.12	0.08	0.0019	1.51 (1.16–1.97)	2.72×10^{-8}	1.39 (1.24–1.56)
(*LRRK2*)	rs7304279	38.75	T/C	0.11	0.08	5.17×10^{-5}	1.42 (1.20–1.69)	0.10	0.08	0.026	1.25 (1.03–1.52)	5.10×10^{-6}	1.35 (1.19–1.54)	0.12	0.09	0.0022	1.50 (1.15–1.95)	5.06×10^{-8}	1.38 (1.23–1.55)
	rs4768212	38.76	C/T	0.11	0.08	3.98×10^{-5}	1.43 (1.20–1.70)	0.10	0.08	0.057	1.21 (0.99–1.48)	1.10×10^{-5}	1.34 (1.18–1.52)	0.12	0.08	0.0020	1.51 (1.16–1.97)	1.09×10^{-7}	1.37 (1.22–1.54)
	rs2708453	38.76	T/G	0.11	0.08	7.46×10^{-5}	1.41 (1.19–1.68)	0.10	0.08	0.063	1.21 (0.99–1.48)	2.04×10^{-5}	1.33 (1.17–1.52)	0.13	0.08	6.43×10^{-4}	1.57 (1.21–2.04)	9.67×10^{-8}	1.38 (1.22–1.55)
	rs2046932	38.87	T/C	0.11	0.08	3.24×10^{-5}	1.44 (1.21–1.71)	0.10	0.08	0.039	1.23 (1.01–1.51)	5.47×10^{-6}	1.35 (1.19–1.54)	0.13	0.09	0.0017	1.52 (1.17–1.97)	4.34×10^{-8}	1.39 (1.23–1.56)

Nucleotide positions refer to NCBI build 36. *P* values obtained in the case-control analysis using the Cochran-Armitage trend test (1 d.f.) are listed (P_{trend}). Combined *P* values (P_{cmh}) and combined ORs of the Cochran-Mantel-Haenzel test statistics are shown. MAF, minor allele frequency.

Next, we exchanged data with colleagues performing a GWAS of PD in individuals of European ancestry[17]. Their study found a strong association at the *MAPT* (microtubule-associated protein tau) region on 17q21. We genotyped our samples for six SNPs at the *MAPT* locus to evaluate these associations in the Asian population; however, the association with *MAPT* was not replicated in our study (**Supplementary Table 2** and **Supplementary Fig. 2**). Conversely, despite strong association signals in our scan of the samples from the Asian population, the association with *BST1* on 4p15 was not detected among individuals of European ancestry[17]. In contrast, the associations we found with *PARK16* and *LRRK2* were replicated among individuals of European ancestry[17]. These data provide evidence that *PARK16* and *LRRK2*, in addition to *SNCA*, are PD risk loci common to Asian- and European-descent populations and indicate that there is population genetic heterogeneity in the *MAPT* region and 4p15 (*BST1*) for PD susceptibility.

The *PARK16* region contains functionally interesting candidate genes for PD etiology. SLC41A1 is a magnesium (Mg^{2+}) transporter[18]. It is of interest that Mg^{2+} deficiency is thought to be an environmental risk factor for the amyotrophic lateral sclerosis

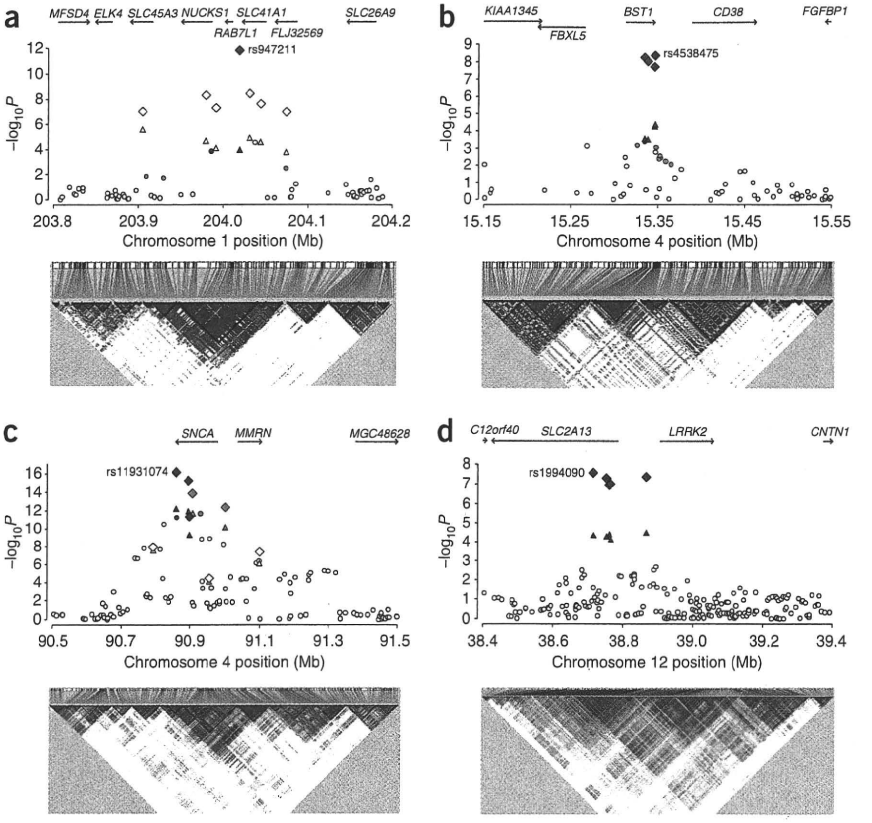

Figure 2 Regional association plots and linkage disequilibrium structure for the four PD risk loci. (**a**) 1q32 (*PARK16*). (**b**) 4p15 (*BST1*). (**c**) 4q22 (*SNCA*). (**d**) 12q12 (*LRRK2*). The $-\log_{10}$ *P* (Cochran-Armitage trend test) for association in the GWAS stage of SNPs across each region are shown as small triangles for SNPs that were selected for replication and as small circles for SNPs not selected. The $-\log_{10}$ combined *P* values (Cochran-Mantel-Haenszel test) for association in overall samples of SNPs selected for replication are shown as large diamonds. In each panel, the SNP with the most significant association in the combined analysis is listed. Proxies are indicated with colors determined from their pairwise r^2 from the JPT and CHB HapMap data (red, $r^2 > 0.8$; orange, $r^2 = 0.5$–0.8; yellow, $r^2 = 0.2$–0.5; white, $r^2 < 0.2$ or no information available). Positions are NCBI build 36 coordinates.

(ALS)-parkinsonism/dementia complex (MIM105500)[19]. Furthermore, RAB7L1 is a small GTP-binding protein that plays an important role in regulation of exo- and endocytotic pathways[20], and NUCKS1 is a nuclear protein containing several consensus phosphorylation sites for casein kinase II and cyclin-dependent kinases of unknown function[21]. We evaluated the relationships between the PD-associated SNPs and the transcript levels of genes in an available genome-wide gene expression database[22]. We found that rs947211 and ten tightly linked HapMap SNPs ($r^2 > 0.9$) were strongly associated with transcript levels of *NUCKS1* (rs947211, $P = 6.0 \times 10^{-15}$; rs823114, $P = 2.7 \times 10^{-34}$). These PD-susceptibility variants are the principal genetic determinants of variation in expression levels of *NUCKS1* (**Supplementary Fig. 3**). These data highlight *NUCKS1* as a promising candidate for association with PD that is worthy of additional follow-up.

The product of *BST1* on 4p15 catalyses formation of cyclic ADP-ribose (cADPR)[23]. cADPR mobilizes calcium (Ca^{2+}) from ryanodine-sensitive intracellullar Ca^{2+} stores in the endoplasmic reticulum[24]. Disruption of Ca^{2+} homeostasis has recently been proposed as a possible cause of selective vulnerability of dopaminergic neurons in PD[25–27]. Associated SNPs in the *BST1* region may modify ADP-ribosylcyclase activity, thus leading to Ca^{2+} dyshomeostasis in dopaminergic neurons.

Two of the four susceptibility loci detected in our scan contained genes linked to autosomal dominant forms of parkinsonism. Gene overdosage is a potential mechanism for the influence of *SNCA* on PD because triplication and duplication of the *SNCA* locus has been seen in families with autosomal dominant parkinsonism[28]. SNPs with prominently low *P* values compared to other SNPs in the region were around the 3′ region of *SNCA*; these SNPs may function as enhancer or silencer elements, improve RNA stability or influence alternative splicing. The associated interval on 12q12 contains *SLC2A13* and the region upstream of *LRRK2*. Given prior evidence, *LRRK2* stands out as the most likely susceptibility gene for PD, although it remains possible that *SLC2A13*, which encodes a H^+-myoinositol cotransporter, may be the PD-related gene in this region[29]. Previous reports have investigated the association of SNPs in *LRRK2* with PD, but the results are a subject of dispute[30,31]. In the present study, it is noteworthy that the PD-associated intervals lie upstream of *LRRK2*. Increased kinase activity of mutant LRRK2 mediates neuronal toxicity[32,33]. PD-associated SNPs may play a role in transcriptional upregulation of *LRRK2*, leading to loss of dopaminergic neurons.

SNCA is a main component of Lewy bodies, a pathological hallmark of typical PD. The clinical features of individuals with *SNCA* duplication or *LRRK2* mutation similar to typical PD. 1.6% of sporadic PD cases among individuals of European ancestry have heterozygous *LRRK2* G2019S mutations[34]. These data support the close involvement of these genes with sporadic PD. Our data clearly show that the genes involved in autosomal dominant parkinsonism play a large part in the complex etiology of typical PD. Genes that cause autosomal dominant parkinsonism through their causative mutations also confer risk of typical PD through their common variants. Although further research is needed, this relationship between rare single-gene disorders and common multifactorial disorders may also be applicable for other disorders beyond PD.

Finally, *MAPT* mutations cause hereditary frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17), a type of autosomal dominant parkinsonism[35], and the *MAPT* H1 haplotype has been reported to be associated with several tauopathies[36–38]. Although the *MAPT* region is divided into two major haplotypes, H1 and H2, in Europeans, the H2 haplotype is absent in East Asians. Therefore, we believe that the differences observed between our study and the findings in populations of European descent reflect population differences in the genetic heterogeneity of PD etiology, although differences in allele frequencies and LD structure and a possible difference in the effect size between the European and East Asian populations may influence the detection power of the two scans.

Further increases in sample sizes for SNP-GWAS efforts and searches for copy number variation and rare variants will reveal additional genetic risk factors and further enhance our understanding of PD.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to the individuals with PD who participated in this study. We also thank T. Takeshima and E. Ohta for PD samples, H. Inoko and K. Tokunaga for control samples, M. Kanagawa and K. Ura for editing, K. Yasuno and R. Ashida for analyses, and K. Yamada for genotyping. We appreciate all the volunteers and participating institutions of BioBank Japan for samples. This work was supported by a grant from Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST); by the Global COE program and KAKENHI (17019044 and 19590990), both from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and by the Grant-in-Aid for 'the Research Committee for the Neurodegenerative Diseases' of the Research on Measures for Intractable Diseases and Research Grant (H19-Genome-Ippan-001), all from the Ministry of Health, Labor and Welfare of Japan.

AUTHOR CONTRIBUTIONS

T. Toda conceived the study. W.S., I.M. and T. Toda designed the study. W.S., Y.N., C.I., M.K. and T.Y. performed genotyping. W.S. and T. Toda wrote the manuscript. W.S., T.K. and T. Tsunoda performed data analysis. W.S., I.M., Y.H., M.W., A.T., H.T., K.N., K.H., F.O., H.K., S.S., M.Y., N.H., M.M. and T. Toda managed Parkinson clinical information and DNA samples. M.K. and Y.N. managed DNA samples belonging to BioBank Japan. T. Toda obtained funding for the study.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. de Rijk, M.C. *et al.* Prevalence of parkinsonism and Parkinson's disease in Europe: the EUROPARKINSON collaborative study. *J. Neurol. Neurosurg. Psychiatry* **62**, 10–15 (1997).
2. Polymeropoulos, M.H. *et al.* Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).
3. Paisán-Ruíz, C. *et al.* Cloning of the gene containing mutations that cause *PARK8*-linked Parkinson's disease. *Neuron* **44**, 595–600 (2004).
4. Zimprich, A. *et al.* Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601–607 (2004).
5. Farrer, M.J. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat. Rev. Genet.* **7**, 306–318 (2006).
6. Thomas, B. & Beal, M.F. Parkinson's disease. *Hum. Mol. Genet.* **16 Spec No. 2**, R183–R194 (2007).
7. Warner, T.T. & Schapira, A.H. Genetic and environmental factors in the cause of Parkinson's disease. *Ann. Neurol.* **53**(Suppl 3), S16–S23 (2003).
8. Pals, P. *et al.* α-Synuclein promoter confers susceptibility to Parkinson's disease. *Ann. Neurol.* **56**, 591–595 (2004).
9. Mizuta, I. *et al.* Multiple candidate gene analysis identifies α-synuclein as a susceptibility gene for sporadic Parkinson's disease. *Hum. Mol. Genet.* **15**, 1151–1158 (2006).
10. Müller, J.C. *et al.* Multiple regions of α-synuclein are associated with Parkinson's disease. *Ann. Neurol.* **57**, 535–541 (2005).
11. Aharon-Peretz, J., Rosenbaum, H. & Gershoni-Baruch, R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. *N. Engl. J. Med.* **351**, 1972–1977 (2004).
12. Maraganore, D.M. *et al.* High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77**, 685–693 (2005).
13. Fung, H.C. *et al.* Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* **5**, 911–916 (2006).
14. Pankratz, N. *et al.* Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.* **124**, 593–605 (2009).
15. Welcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
16. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
17. Simón-Sánchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* advance online publication, doi:10.1038/ng.487 (15 November 2009).
18. Kolisek, M. *et al.* SLC41A1 is a novel mammalian Mg^{2+} carrier. *J. Biol. Chem.* **283**, 16235–16247 (2008).
19. Garruto, R.M. *et al.* Disappearance of high-incidence amyotrophic lateral sclerosis and parkinsonism-dementia on Guam. *Neurology* **35**, 193–198 (1985).
20. Shimizu, F. *et al.* Cloning and chromosome assignment to 1q32 of a human cDNA (*RAB7L1*) encoding a small GTP-binding protein, a member of the RAS superfamily. *Cytogenet. Cell Genet.* **77**, 261–263 (1997).
21. Ostvold, A.C. *et al.* Molecular cloning of a mammalian nuclear phosphoprotein NUCKS, which serves as a substrate for Cdk1 *in vivo*. *Eur. J. Biochem.* **268**, 2430–2440 (2001).
22. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
23. Yamamoto-Katayama, S. *et al.* Crystallographic studies on human BST-1/CD157 with ADP-ribosyl cyclase and NAD glycohydrolase activities. *J. Mol. Biol.* **316**, 711–723 (2002).
24. Lee, H.C. *et al.* Physiological functions of cyclic ADP-ribose and NAADP as calcium messengers. *Annu. Rev. Pharmacol. Toxicol.* **41**, 317–345 (2001).
25. Wilson, C.J. & Callaway, J.C. Coupled oscillator model of the dopaminergic neuron of the substantia nigra. *J. Neurophysiol.* **83**, 3084–3100 (2000).
26. Chan, C.S. *et al.* 'Rejuvenation' protects neurons in mouse models of Parkinson's disease. *Nature* **447**, 1081–1086 (2007).
27. Surmeier, D.J. Calcium, ageing, and neuronal vulnerability in Parkinson's disease. *Lancet Neurol.* **6**, 933–938 (2007).
28. Singleton, A.B. *et al.* α-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
29. Uldry, M. *et al.* Identification of a mammalian H(+)-myo-inositol symporter expressed predominantly in the brain. *EMBO J.* **20**, 4467–4477 (2001).
30. Skipper, L. *et al.* Comprehensive evaluation of common genetic variation within *LRRK2* reveals evidence for association with sporadic Parkinson's disease. *Hum. Mol. Genet.* **14**, 3549–3556 (2005).
31. Biskup, S. *et al.* Common variants of *LRRK2* are not associated with sporadic Parkinson's disease. *Ann. Neurol.* **58**, 905–908 (2005).
32. Smith, W.W. *et al.* Kinase activity of mutant LRRK2 mediates neuronal toxicity. *Nat. Neurosci.* **9**, 1231–1233 (2006).
33. West, A.B. *et al.* Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl. Acad. Sci. USA* **102**, 16842–16847 (2005).
34. Gilks, W.P. *et al.* A common *LRRK2* mutation in idiopathic Parkinson's disease. *Lancet* **365**, 415–416 (2005).
35. Hutton, M. *et al.* Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702–705 (1998).
36. Pittman, A.M. *et al.* Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J. Med. Genet.* **42**, 837–846 (2005).
37. Webb, A. *et al.* Role of the tau gene region chromosome inversion in progressive supranuclear palsy, corticobasal degeneration, and related disorders. *Arch. Neurol.* **65**, 1473–1478 (2008).
38. Healy, D.G. *et al.* Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **75**, 962–965 (2004).

ONLINE METHODS

Study participants. For the GWAS stage, 1,078 cases and 2,628 controls were recruited from Japan through multiple institutions. Two case-control sample sets, which were independent of each other, were used in the two subsequent replication stages (replication 1 and 2): the first sample set (replication 1) consisted of 612 cases and 14,139 controls and the second set (replication 2) consisted of 321 cases and 1,614 controls, all recruited in Japan. For replication 2, we used case samples from two facilities that had not provided any case samples for GWAS and replication 1, in order to eliminate false positive association due to a slight possibility of differences in PD diagnosis among facilities. Genomic DNA was extracted using established methods and quantified using PicoGreen (Invitrogen). Details for all study panels are given in the **Supplementary Note**. Informed consent was obtained from each participant, and approval for the study was obtained from the Ethical Committees of relevant institutions (Osaka University Graduate School of Medicine, National Center Hospital of Neurology and Psychiatry, Juntendo University School of Medicine, Kagawa Prefectural Central Hospital, University of Tsukuba, Tohoku University Graduate School of Medicine and Tottori University Faculty of Medicine).

Genotyping. Genome-wide genotyping was performed using the Illumina Infinium HumanHap550 array. Cases and controls were genotyped at the Division of Clinical Genetics of Osaka University Graduate School of Medicine and at RIKEN Center for Genomic Medicine, respectively. For two subsequent replication studies (replication 1 and 2), 337 SNPs were genotyped for replication 1 and 2 samples using Illumina GoldenGate technology for the VeraCode platform (335 SNPs; cases in replication 1 and cases and controls in replication 2), TaqMan (2 SNPs; cases in replication 1 and cases and controls in replication 2) and the Illumina Infinium HumanHap610 array (337 SNPs; controls in replication 1). For the replication study of the *MAPT* locus, we selected six SNPs (rs417968, rs17690703, rs242557, rs7225002, rs183211 and rs7224296) that showed significant association, including four SNPs showing association with genome-wide significance, in samples of European ancestry. Samples with a call rate > 90% in replications 1 and 2 (877 cases and 15,616 contols) were genotyped. All genotyping was done according to the manufacturer's instructions. To assess consistency across genotyping platforms, we genotyped these SNPs in 95 samples included in the GWAS. After SNP and sample quality control analyses, the mean concordance rates were 99.8% and 99.5% for GoldenGate technology and TaqMan when compared with the HumanHap550 array.

Quality control. In the GWAS stage, case samples with a call rate < 95% and control samples with a call rate < 98% were excluded, according to each criterion of separate institutes which genotyped cases and controls. Remaining samples were reclustered using BeadStudio (Illumina), and genotypes of 1,012 cases and 2,573 controls were then obtained. We excluded samples with ambiguous sex ($n = 18$) by the check-sex function of PLINK 1.01 (ref. 39). We determined identity-by-state (IBS) similarity using PLINK 1.01, estimated the cryptic relatedness for each pair of samples, and excluded one individual from each pair of unexpected duplicates and first- or second-degree cryptic relatives ($n = 55$). To detect population outliers, we assessed 3,512 participants who remained after removal of samples with low call rates, ambiguous sex and familial relationships, together with 201 HapMap subjects without relationships (42 JPT, 45 CHB, 57 CEPH and 57 YRI)[40]. By computing IBS scores for 49,605 SNPs with $r^2 < 0.2$ and using multidimensional scaling, we identified three individuals who seemed to have non-Asian ancestry and excluded those from further analyses. Projection onto the two multidimensional scaling axes is shown in **Supplementary Figure 4**. We excluded SNPs with a call rate < 95% in cases or controls ($n = 7,927$), a minor allele frequency < 5% ($n = 117,908$) in all samples, or a P value of deviation from Hardy-Weinberg equilibrium (P_{HWE}) < 0.001 in the controls ($n = 3,045$). On visual inspection of the cluster plots of SNPs showing apparently strong association, we further removed 69 SNPs with poor clustering. The overall median genotype call rate for quality-controlled SNP was 99.9%. In replication 1, we excluded samples with a call rate < 90%. We also removed SNPs with a call rate <130895% in cases or controls, a $P_{HWE} < 0.001$ in the controls, or poor clustering of SNP plot on visual inspection. Genotypes of 279 SNPs for 559 cases and 14,026 controls were then obtained for further analyses. The overall mean genotype call rate was 99.7% for quality-controlled SNPs. In replication 2, we excluded samples with a call rate < 90% and then obtained genotypes of 318 cases and 1,590 controls. All 24 SNPs showed a call rate > 90% and $P_{HWE} > 0.001$. The overall mean genotype call rate was 99.7% for quality-controlled SNPs. Associated SNPs in each interval had high call rates in each sample set (**Supplementary Table 3**). In the replication study of the *MAPT* locus, all six SNPs showed a call rate > 95% and $P_{HWE} > 0.001$.

Statistical analyses. To calculate the power of our GWAS stage, we used the CaTS program[41]. The GWAS stage had 80% power to detect common alleles that confer a genotype relative risk of 1.3 and 1.43 at a significance of $P < 0.00053$ and $P < 5 \times 10^{-7}$, respectively.

To test for association of each SNP with PD, we used the Cochran-Armitage trend test with 1 d.f. We estimated the odd ratios (OR) and their 95% confidence intervals using logistic regression. Association analysis of the combined samples was conducted using the Cochran-Mantel-Haenezel method. Heterogeneity among sample sets was examined using the Breslow-Day test. There was no heterogeneity among sample sets (rs947211, $P = 0.38$; rs4538475, $P = 0.98$; rs11931074, $P = 0.09$; rs3857059, $P = 0.08$; and rs1994090, $P = 0.48$). SNPs with combined $P < 5 \times 10^{-7}$ were considered to have genome-wide significant evidence for association. SNPs with genome-wide significant evidence for association in the combined analysis of the GWAS and replication 1 and $P < 0.05$ in replication 2 were considered to have confirmed association with PD. To assess whether single or multiple independent association signals existed within each locus, we investigated relationships among multiple SNPs that showed association with PD in the same region ($P < 5.0 \times 10^{-7}$), using logistic regression analysis. We assessed the impact of additional SNPs by a likelihood-ratio test with 1 d.f. A significant residual association signal was defined as $P < 0.05$ in the conditional analysis. We used R 2.8.1 or PLINK 1.01 for general statistical analysis.

The quantile-quantile plot was used to evaluate overall significance of the genome-wide association results and the potential impact of population stratification. Quantile-quantile plots were depicted using the qq.chisq function of the snpMatrix package with a concentration band[15]. The inflation factor λ was calculated by dividing the mean of the lower 90% of the test statistics by the mean of the lower 90% of the expected values from a χ^2 distribution with 1 d.f. Given that the impact of population stratification was found to be minimal, all statistical results are presented without correction for λ. Haploview 4.1 was used to infer the LD structure of the genome in regions containing loci associated with disease risk[42]. The LD patterns were created from the Asian (CHB and JPT) HapMap data (minor allele frequency > 5%, genotyping rate > 95%, and $P_{HWE} > 0.001$), using the methods of Gabriel *et al.*[16].

URLs. PLINK, http://pngu.mgh.harvard.edu/~purcell/plink; R, http://www.r-project.org; SNPmatrix, http://www-gene.cimr.cam.ac.uk/clayton; HapMap, http://www.hapmap.org; Haploview, http://www.broad.mit.edu/mpg/haploview; CaTS, http://www.sph.umich.edu/csg/abecasis/CaTS/; database of expression QTL analysis, http://www.sph.umich.edu/csg/liang/imputation/.

39. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
40. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
41. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
42. Barrett, J.C., Fry, B., Maller, J. & Daly, M. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).

doi:10.1038/ng.485

ORIGINAL CONTRIBUTION

Mutations for Gaucher Disease Confer High Susceptibility to Parkinson Disease

Jun Mitsui, MD; Ikuko Mizuta, MD, PhD; Atsushi Toyoda, PhD; Ryo Ashida, BS; Yuji Takahashi, MD, PhD; Jun Goto, MD, PhD; Yoko Fukuda, PhD; Hidetoshi Date, PhD; Atsushi Iwata, MD, PhD; Mitsutoshi Yamamoto, MD, PhD; Nobutaka Hattori, MD, PhD; Miho Murata, MD, PhD; Tatsushi Toda, MD, PhD; Shoji Tsuji, MD, PhD

Background: Increased frequency of pathogenic variants in *GBA*, the causative gene for Gaucher disease, has been suggested to be associated with Parkinson disease (PD).

Objectives: To conduct comprehensive resequencing of *GBA* to identify all sequence variants and to investigate the association of these variants with PD.

Design: Case-control study.

Setting: Multicenter university-based study.

Participants: Five hundred thirty-four patients with PD, 34 families in which multiple patients with PD are present, and 544 control subjects.

Main Outcome Measures: Disease status and *GBA* variations.

Results: Comprehensive resequencing of *GBA* in 534 patients with PD and 544 controls revealed 27 sequence variants: 11 pathogenic variants associated with Gaucher disease, 11 nonsynonymous variants not associated with Gaucher disease, and 5 synonymous variants. Fifty patients with PD (9.4%) had 1 of the 11 pathogenic variants in the heterozygous state, whereas only 2 controls (0.37%) had such variants (odds ratio, 28.0). Among the pathogenic variants, R120W and L444P/Rec*Nci*I were highly prevalent, and each showed a significant association with PD. Furthermore, other rare pathogenic variants were found in 13 patients with PD but not in the controls, further confirming the role of these rare variants in the susceptibility to PD. Patients with PD carrying pathogenic variants were significantly younger than those not carrying them. In addition, concordance of PD states and pathogenic variants was observed in 8 multiplex families with PD.

Conclusion: Heterozygous pathogenic variants in *GBA* confer a high risk for sporadic PD, even for familial clustering, and are associated with significantly earlier age at onset of disease.

Arch Neurol. 2009;66(5):571-576

Author Affiliations: Departments of Neurology, University of Tokyo Graduate School of Medicine (Drs Mitsui, Takahashi, Goto, Fukuda, Date, Iwata, and Tsuji), Juntendo University School of Medicine (Dr Hattori), and Musashi Hospital, National Center of Neurology and Psychiatry (Dr Murata), Tokyo, and Kagawa Prefectural Central Hospital, Takamatsu (Dr Yamamoto); Division of Clinical Genetics, Department of Medical Genetics, Osaka University Graduate School of Medicine, Suita, Osaka (Drs Mizuta and Toda and Mr Ashida); and Comparative Genomics Laboratory, National Institute of Genetics, Shizuoka (Dr Toyoda), Japan. Dr Murata is now with the National Center of Neurology and Psychiatry, Tokyo.

PARKINSON DISEASE (PD), characterized by tremor, rigidity, bradykinesia, and postural instability, is the second most common neurodegenerative disease after Alzheimer disease, with usual onset in late adulthood, that is, after age 50 years. The prevalence

See also pages 555 and 578

of PD is estimated to be 0.3% in the general population and 1% in individuals older than 60 years.[1] Although *SNCA*, *LRRK2*, *UCHL-1*, *PARK2*, *PINK1*, and *DJ-1* have been identified as the causative genes for familial PD,[2] patients with PD with pathogenic mutations in these genes are rare. Most cases of PD are sporadic and the etiologies poorly understood. A population-based study coupled with genealogic information demonstrated that the estimated risk ratio for PD in siblings of patients with PD was significantly high ($\lambda s=6.7$), which suggests that genetic factors substantially contribute to the development of sporadic PD.[3] To elucidate susceptibility genes for sporadic PD, numerous case-control association studies using the analyses of single nucleotide polymorphisms have been conducted under the common disease–common variants hypothesis; however, only a few consistent findings have been observed.[4] Recently, polymorphisms of *SNCA*, a major component of Lewy bodies, a pathologic hallmark of PD, have been reported to be associated with sporadic PD (odds ratio [OR], 1.4-2.0).[5,6]

Several articles have suggested the association of sporadic PD with heterozygous variants in the glucocerebrosidase gene (*GBA*) (OMIM 606463) encoding the enzyme that is deficient in patients with Gaucher disease, an autosomal recessive lysosomal storage disease. Although *GBA*

Table 1. Demographic Data for Study Participants

Variable	Tier 1		Tier 2	
	Patients With PD (n=61)	Control Subjects (n=47)	Patients With PD (n=473)	Control Subjects (n=497)
Age at sampling, mean (SD), y	66.8 (8.2)	58.4 (11.8)	65.2 (9.9)	43.4 (16.4)
Age at onset of PD, mean (SD), y	58.3 (9.9)	NA	58.2 (10.7)	NA
Sex, male to female ratio	1.44	1.76	1.08	1.12

Abbreviations: NA, not applicable; PD, Parkinson disease.

Table 2. Primers for PCR and Sequence Analysis

Primer	Forward	Reverse
PCR		
Exons 1-5	CCTAAAGTTGTCACCCATAC	AGCAGACCTACCCTACAGTTT
Exons 5-7	GACCTCAAATGATATACCTG	AGTTTGGGAGCCAGTCATTT
Exons 8-11	TGTGTGCAAGGTCCAGGATCAG	ACCACCTAGAGGGGAAAGTG
Sequence		
Exon 1	TAGTGGATCCTCTATCCTTC	AAATTCCAGTGCCAGGATTC
Exon 2	AAAGGCAGCTAAGCCCTGCC	GCTACCAAAGGACTATGAGG
Exon 3	AGTCTCTCCTAGCAGATGTG	TCCATGGTGATCACTGACAC
Exon 4	AAATGGTGTCAGTGATCACC	GCAGAGTGAGATTCTGCCTC
Exon 5	GCAAGTGATAAGCAGAGTCC	CAAGCAGACCTACCCTACAG
Exon 6	AATGGCTGAACCGGATGCAC	AAGTGGAACTAGGTTGAGGG
Exon 7	TCAAGTGATCCACCTGCCTC	AGTTTGGGAGCCAGTCATTT
Exon 8	TGTGTGCAAGGTCCAGGATCAG	GCTTCTGTCAGTCTTTGGTG
Exon 9	ACCCTTACCTACACTCTCTG	GTGATGTAAGCCATCCGATG
Exon 10	GGGTGACTTCTTAGATGAGG	AGCTGAGAGTGTGATCCTGC
Exon 11	GGAAGTGGGCTGAAGACAGC	TTAGTCACAGACAGCGTGT

Abbreviation: PCR, polymerase chain reaction.

variants associated with Gaucher disease are diverse and each carrier frequency is rare, most of the previous studies analyzed only specific variants[7-16] and sample sizes were small.[17-21] Therefore, ORs assessed for the *GBA* variants have been highly variable in the subsequent studies, making the medical implications of *GBA* variants associated with PD inconclusive. We conducted extensive resequencing analysis of *GBA* in patients with PD and in control subjects and found that *GBA* variants that are pathogenic for Gaucher disease confer high susceptibility to sporadic PD and, furthermore, familial clustering of PD.

METHODS

SUBJECTS

We conducted a resequencing of *GBA* in patients with PD and control subjects using a microarray-based, high-throughput resequencing system (first tier). As an independent data set, resequencing of *GBA* was conducted on large-scale samples (second tier) using direct nucleotide sequence analysis. The first tier comprised 61 unrelated patients with PD at the University of Tokyo Hospital and 47 controls provided by the Japan Multiple System Atrophy Research Consortium. The second tier comprised 473 unrelated patients with PD and 497 controls provided by the Japanese Parkinson Disease Susceptibility Gene Consortium (**Table 1**). In addition, 34 families in which multiple patients with PD are present (hereafter referred to as "multiplex families") independent of participants in tiers 1 and 2, having more than 1 patient with PD in the second degree, were provided by the Japanese Parkinson Disease Susceptibility Gene Consortium. The diagnosis of PD was based on diagnostic criteria for PD.[22] This study was approved by the institutional review boards of the participating institutions.

GENOMIC DNA AND AMPLIFICATION OF *GBA*

Genomic DNA was extracted from peripheral blood leukocytes using standard procedures. Three primer pairs were designed to selectively amplify *GBA* but not its pseudogene, as previously described (**Table 2**).[23]

RESEQUENCING OF TIER 1

Resequencing of *GBA* was conducted using newly designed resequencing microarrays TKYPD02 and TKYPD03, both of which were composed of tiled sequences of all 11 exons of *GBA* and the flanking 12 base pairs of the splicing junctions.[24] The analysis was conducted according to the manufacturer's instructions (Affymetrix Inc, Santa Clara, California). All variants were further confirmed at direct nucleotide sequence analysis using a genetic analyzer (ABI PRISM 3100; Applied Biosystems Inc, Foster City, California).

RESEQUENCING OF TIER 2

The polymerase chain reaction products were subjected to direct nucleotide sequence analysis for the coding sequences and the flanking splice sites of *GBA* using DNA analyzers (ABI3730xl; Applied Biosystems Inc). The primers for sequence analysis are given in Table 2.

Table 3. Frequencies of the Pathogenic *GBA* Variants in Patients With PD and Control Subjects

Variants	Patients With PD (n=534)	Control Subjects (n=544)	*P* Value[a]	OR (95% CI)
R120W	15	0	<.001	NA
R131C	1	0	NA	NA
N188S	4	0	.06	NA
R120W-N188R-V191G-S196P-F213I	1	0	NA	NA
G193W	1	0	NA	NA
F213I	1	0	NA	NA
R329C	2	0	.25	NA
L444P	8	0	.004	NA
L444P-A456P-V460V (Rec*Nci*I)	14	2	.002	7.3 (1.7-66.4)
A456P-V460V	1	0	NA	NA
R496C	2	0	.25	NA
Total (%)	**50** (9.4)	**2** (0.37)	6.9×10^{-14}	28.0 (7.3-238.3)

Abbreviations: CI, confidence interval; NA, not applicable; OR, odds ratio; PD, Parkinson disease.
[a] Fisher exact test.

STATISTICAL ANALYSIS

Standard statistical methods were used to test the difference in carrier frequency (Fisher exact test), to compute ORs and corresponding 95% confidence intervals, and to compare mean age at onset of PD (*t* test). For a meta-analysis, a pooled OR was calculated using a fixed-effects model (Mantel-Haenszel method). $P<.05$ was considered statistically significant. Data were analyzed using commercially available statistical software (StatsDirect version 2.6.5; StatsDirect Ltd, Cheshire, England).

RESULTS

Resequencing of tier 1 (61 patients with PD and 47 controls) revealed that 6 patients with PD carried the variants (1 R120W, 1 R329C, 3 Rec*Nci*I, and 1 R496C) that are pathogenic for Gaucher disease, whereas none of these variants were present in the controls. Given this result, we further expanded the comprehensive resequencing analysis to tier 2 (473 patients with PD and 497 controls) and identified 44 patients with PD carrying the variants that have been reported to be pathogenic for Gaucher disease, whereas these variants were present in only 2 controls.

Pathogenic variants were either single-base substitutions (R120W, R131C, N188S, G193W, F213I, R329C, L444P, and R496C) or complex multiple substitutions (R120W-N188R-V191G-S196P-F213I, L444P-A456P-V460V, and A456P-V460V). The precise structures of the complex alleles were confirmed at nucleotide sequence analysis of the subcloned mutant alleles. Among the complex mutant alleles, L444P-A456P-V460V is a Rec*Nci*I allele, a recombination allele that consists of 3 single-base substitutions of the pseudogene origin in exon 10.[25] In summary, we found that 50 of 534 patients with PD (9.4%) had these pathogenic variants in the heterozygous state, whereas only 2 of 544 controls (0.37%) had such variants in the heterozygous state (OR [95% confidence interval] for patients with PD compared with controls, 28.0 [7.3-238.3], which was highly significant ($P=6.9\times10^{-14}$) (**Table 3**). When individual variants were analyzed, the frequency of the R120W, L444P, and Rec*Nci*I carriers was significantly higher in patients with PD than in controls ($P<.001$, .004, and .002, respectively). In addition, we identified 11 nonsynonymous variants and 5 synonymous variants in tiers 1 and 2, and none of these has been shown to be causative for Gaucher disease. When these variants were analyzed individually and in combination, the frequency of patients with PD was not significantly different from that of the controls (**Table 4**).

We analyzed the clinical manifestations in the 50 patients with PD carrying pathogenic variants in *GBA*. The age at disease onset in the patients with PD who were carriers of such variants was significantly younger than in those who were not carriers (**Table 5**). Detailed clinical data were available for 49 of 50 patients with PD carrying pathogenic variants. Forty-one of 49 patients with PD (83.7%) showed good responsiveness to antiparkinsonian drug treatment. Iodine 123–labeled metaiodobenzylguanidine cardiac scintigraphy[26] was carried out in 33 patients with PD, revealing that 29 of 33 patients with PD (87.9%) had reduced cardiac uptake, consistent with a diagnosis of PD. In the 49 patients with PD, 13 (26.5%) manifested overt dementia (clinical dementia rating[27] ≥1) and 17 (34.7%) developed visual hallucinations during the course of the disease (mean [SD] interval between onset of PD and evaluation of dementia or visual hallucinations, 9.1 [4.1] and 7.9 [5.0] years, respectively). *N*-isopropyl-*p*-[^{123}I]-iodoamphetamine single-photon emission computed tomography was performed in 15 patients with PD, of whom 8 had dementia. All 8 patients with dementia exhibited hypoperfusion in the occipital areas. In the 7 patients without dementia, 5 exhibited hypoperfusion in the occipital areas and 2 had normal findings.

Detailed inquiry into the family history of the 50 patients with PD carrying pathogenic variants in *GBA* revealed that 11 patients (22.0%) had parents or siblings with PD. Genomic DNA was available for 3 affected siblings. All 3 affected siblings had the same *GBA* variants (2 R120W and 1 Rec*Nci*I) as did their probands. Given the concordant *GBA* variants in the 3 affected siblings, we analyzed probands of an additional 34 multiplex families independent of those in tiers 1 and 2 with more than 1 patient (parent or sibling) with PD. We found that 5

Table 4. Frequency of Nonpathogenic *GBA* Variants in Patients With PD and Control Subjects

Variant	Patients With PD (n=534)	Control Subjects (n=544)	*P* Value[a]	OR (95% CI)
I(-20)V	77	66[b]	.28	1.2 (0.84-1.8)
L(-15)F	1	0	NA	NA
L67Q	0	1	NA	NA
V121V	0	1	NA	NA
D153N	1	0	NA	NA
R163Q	4	7	.55	0.58 (0.12-2.3)
P299T	1	0	NA	NA
G307S	1	0	NA	NA
T334I	0	1	NA	NA
L336L	1	0	NA	NA
G344G	1	0	NA	NA
F347L	0	1	NA	NA
R359L	1	0	NA	NA
V460V	2	1	.62	2.0 (0.11-120.7)
K466K	11	8	.50	1.4 (0.51-4.1)
I489V	4	3	.72	1.4 (0.23-9.3)
Nonsynonymous	90	79	.32	1.2 (0.85-1.7)
Synonymous	15	10	.32	1.5 (0.64-3.9)

Abbreviations: CI, confidence interval; NA, not applicable; OR, odds ratio; PD, Parkinson disease.
[a] Fisher exact test.
[b] One subject had I(-20)V in the homozygous state.

Table 5. Age at Onset of PD in Carriers and Noncarriers of the Pathogenic *GBA* Variants

Variable	No. of Patients	Age at Onset of PD, Mean (SD)	*P* Value[a]
Carriers of pathogenic *GBA* variants	50	52.5 (7.4)	<.001
Carriers of R120W	15	51.8 (8.2)	.01
Carriers of L444P/Rec*Nci*I	22	52.6 (6.9)	.007
Noncarriers of pathogenic *GBA* variants	484	58.8 (10.7)	NA

Abbreviations: NA, not applicable; PD, Parkinson disease.
[a] *t* Test.

of 34 probands (14.7%) had pathogenic variants in *GBA* (1 each, R120W, N188S, IVS6 + 1g>a, L444P, and Rec*Nci*I), and all 5 affected relatives also concordantly had the same *GBA* variants as did their probands. The splice junction mutation IVS6+1g>a is a novel variant that has not been reported even in patients with Gaucher disease; however, it is likely the pathogenic variant because it would affect splicing of intron 6. In total, 8 multiplex families with patients with PD concordantly carrying the pathogenic variants were identified (**Figure**).

We compared the distributions of pathogenic variants in *GBA* in the 534 Japanese patients with PD (50 alleles) with those of the mutations that have been previously described in the 50 Japanese patients with Gaucher disease (100 alleles).[28] R120W was present in 30% of the pathogenic variants in the patients with PD, whereas it was not described in patients with Gaucher disease. F213I was the second most common mutation in patients with Gaucher disease (14%), but it was present in only 2% of the pathogenic variants in patients with PD. In contrast, the frequency of L444P and Rec*Nci*I was comparable in the 2 groups, and these were the most common variants.

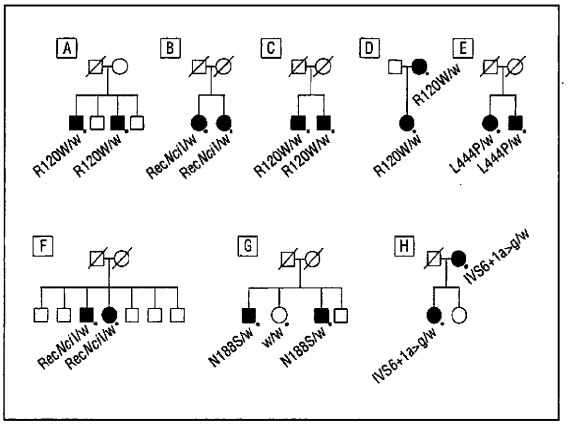

Figure. Pedigree charts for 8 families in which multiple patients with Parkinson disease are present. Squares indicate males; large circles, females; slash, deceased; solid square or large circle, affected individual; and small dot, individual with genomic DNA.

COMMENT

Multiple rare *GBA* variants that are responsible for Gaucher disease confer high risk for PD on the basis of the extensive resequencing of *GBA* of large data sets of Japanese patients with PD and controls. The combined carrier frequency of the pathogenic variants was as high as 9.4% in patients with PD and highly significantly more frequent than in controls (0.37%) with a markedly high OR (95% confidence interval) for patients with PD compared with controls (28.0 [7.3-238.3]). The frequency of nonneuronopathic and neuronopathic Gaucher disease in Japan is estimated to be 1 in 500 000 and 1 in 1 200 000 live births, respectively,[29] which is in accord with the frequency of pathogenic variants in the controls (2 carriers per 544 individuals) in this study.

Among the pathogenic variants identified in the patients with PD, R120W and L444P/Rec*Nci*I were highly prevalent. The identification of multiple rare variants that are pathogenic for Gaucher disease was achieved only by extensive resequencing of large data sets, as clearly demonstrated in the present study. For these pathogenic variants except R120W and L444P/Rec*Nci*I, the frequency of the individual variants was low in patients with PD, and the association with PD should be confirmed in much larger association studies. However, we observed these various rare pathogenic variants in 13 patients with PD, whereas such variants were not observed in the controls. These findings further strengthen the role of these rare *GBA* variants in susceptibility to PD as well.

In contrast to the present findings, previous association studies demonstrated substantially variable ORs. In the studies that demonstrated a significant association of *GBA* variants with PD,[7,11,13-17,20,21] N370S is the variant accounting for most of the significant association. N370S is highly prevalent in the Jewish population, with a carrier frequency of 4% to 6%,[7,8,16,20] and that significant association of N370S with PD has not been demonstrated in other ethnic populations. In contrast to N370S, L444P/Rec*Nci*I has been found regardless of ethnic background. When previous studies that analyzed L444P/Rec*Nci*I[7,9-21] and the present study were subjected to meta-analysis (4181 patients with PD and 9587 controls), a high pooled OR (95% confidence interval) of 6.8 (4.0-11.8) for L444P/Rec*Nci*I was obtained without evidence of significant heterogeneity (Cochran $Q=7.3$; $P=.88$), further confirming the role of *GBA* variants in PD. However, previous studies with small sample sizes failed to detect controls carrying L444P/Rec*Nci*I,[9,11,14,15,17-21] although L444P/Rec*Nci*I was detected in patients with PD. Thus, it is crucially important to determine the frequency of the *GBA* variants in the controls for accurate evaluation of ORs conferred by rare variants, necessitating the analysis of large data sets with at least several hundred patients and controls. Furthermore, there seems to be a bias in the distribution of sequence variants in *GBA* associated with PD compared with that observed in Gaucher disease. In most of the previous studies,[7-16] however, only specific variants considered common in patients with Gaucher disease have been analyzed, which may have led to the underestimation of mutant *GBA* carrier frequency.

Clinically, patients with PD with heterozygous pathogenic variants in *GBA* were significantly younger at disease onset than those without such variants, which confirms findings of previous studies.[7,11,12,16,20] To further determine the exact effects of heterozygous *GBA* variants on PD phenotypes, extensive clinical and epidemiologic analyses should be conducted in large cohorts.

In the present study, we identified 8 multiplex families with patients with PD concordantly having heterozygous pathogenic variants in *GBA*. Given the markedly high ORs caused by heterozygous pathogenic variants in *GBA*, it is conceivable that such variants underlie not only sporadic PD but also familial PD.

The roles of the pathogenic variants in the pathogenesis of PD still needed to be elucidated. Gain of toxic functions of mutant glucocerebrosidase proteins independent of enzyme activities might be involved in the pathogenesis. However, all variants associated with PD are pathogenic variants for Gaucher disease, which raises the possibility that decrease in glucocerebrosidase activities has a role in the pathogenesis of PD. Identification of the splice junction mutation IVS6+1g>a in the present study may further support this notion.

We should emphasize a paradigm shift from the common disease–common variants hypothesis to the common disease–multiple rare variants hypothesis in our search for disease susceptibility genes in sporadic PD, which may be applicable to studies of other diseases. The multiple rare variants can be identified only by extensive resequencing and are difficult to detect in association studies using common single nucleotide polymorphisms. Such multiple rare variants confer strong genetic risks, as demonstrated in the present study, which is also in striking contrast to the low ORs of those identified in genomewide association studies using common single nucleotide polymorphisms. Our results strongly emphasize the importance of conducting a comprehensive resequencing analysis of disease susceptibility genes in detecting even the rarest variants.

In conclusion, we have established *GBA* as a robust and relatively prevalent genetic risk factor for sporadic PD. Further studies of the biological implications of mutant glucocerebrosidase in the pathophysiologic processes of PD are expected to provide new avenues for developing therapeutic measures for PD.

Accepted for Publication: September 3, 2008.
Correspondence: Shoji Tsuji, MD, PhD, Department of Neurology, University of Tokyo Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-8655, Japan (tsuji@m.u-tokyo.ac.jp) or Tatsushi Toda, MD, PhD, Division of Clinical Genetics, Department of Medical Genetics, Osaka University, Graduate School of Medicine, 2-2-B9 Yamadaoka, Suita, Osaka 565-8071, Japan (toda@clgene.med.osaka-u.ac.jp).
Author Contributions: Dr Mitsui had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Mitsui and Mitzuta contributed equally to the study. *Study concept and design:* Mitsui, Mizuta, Takahashi, Goto, Date, Iwata, Toda, and Tsuji. *Acquisition of data:* Mitsui, Mizuta, Toyoda, Ashida, Takahashi, Goto, Yamamoto, Hattori, Murata, Toda, and Tsuji. *Analysis and interpretation of data:* Mitsui, Mizuta, Takahashi, Goto, Fukuda, Toda, and Tsuji. *Drafting of the manuscript:* Mitsui, Mizuta, Toyoda, Takahashi, Goto, Toda, and Tsuji. *Critical revision of the manuscript for important intellectual content:* Ashida, Takahashi, Goto, Fukuda, Date, Iwata, Yamamoto, Hattori, Murata, and Tsuji. *Statistical analysis:* Mitsui, Mizuta, and Fukuda. *Obtained funding:* Tsuji. *Administrative, technical, and material support:* Toyoda, Takahashi, Goto, Date, Yamamoto, Hattori, Murata, and Tsuji. *Study supervision:* Goto, Iwata, Toda, and Tsuji.
Financial Disclosure: None reported.
Funding/Support: This study was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas (Dr Tsuji), Applied Genomics (Dr Tsuji), the 21st Century COE Program "Center for Integrated Brain Medical

Science" (Dr Tsuji), and Scientific Research (A) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Dr Tsuji), by a Grant-in-Aid from the Research Committee of Ataxic Diseases and the Research Committee of CNS Degenerative Diseases of the Research on Measures for Intractable Diseases from the Ministry of Health, Labour, and Welfare of Japan (Dr Toda), and by grants from the Core Research for Evolutional Science and Technology of the Japan Science and Technology Agency and the Takeda Science Foundation (Dr Tsuji).

Additional Contributions: The technical staff of the Sequencing Technology Team at RIKEN Genomic Sciences Center provided assistance; Hiroyuki Tomiyama, MD, PhD, and Taku Hatano, MD, PhD, collected the clinical information for the patients with Parkinson disease; and Yuko Nakabayashi, BS, provided technical help.

REFERENCES

1. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet Neurol.* 2006;5(6):525-535.
2. Farrer MJ. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat Rev Genet.* 2006;7(4):306-318.
3. Sveinbjörnsdottir S, Hicks AA, Jonsson T, et al. Familial aggregation of Parkinson's disease in Iceland. *N Engl J Med.* 2000;343(24):1765-1770.
4. Warner TT, Schapira AH. Genetic and environmental factors in the cause of Parkinson's disease. *Ann Neurol.* 2003;53(suppl 3):S16-S25.
5. Mizuta I, Satake W, Nakabayashi Y, et al. Multiple candidate gene analysis identifies alpha-synuclein as a susceptibility gene for sporadic Parkinson's disease. *Hum Mol Genet.* 2006;15(7):1151-1158.
6. Maraganore DM, de Andrade M, Elbaz A, et al; Genetic Epidemiology of Parkinson's Disease (GEO-PD) Consortium. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *JAMA.* 2006;296(6):661-670.
7. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. *N Engl J Med.* 2004; 351(19):1972-1977.
8. Clark LN, Nicolai A, Afridi S, et al. Pilot association study of the beta-glucocerebrosidase *N370S* allele and Parkinson's disease in subjects of Jewish ethnicity. *Mov Disord.* 2005;20(1):100-103.
9. Sato C, Morgan A, Lang AE, et al. Analysis of the glucocerebrosidase gene in Parkinson's disease. *Mov Disord.* 2005;20(3):367-370.
10. Toft M, Pielsticker L, Ross OA, Aasly JO, Farrer MJ. Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population. *Neurology.* 2006; 66(3):415-417.
11. Tan EK, Tong J, Fook-Chong S, et al. Glucocerebrosidase mutations and risk of Parkinson disease in Chinese patients. *Arch Neurol.* 2007;64(7):1056-1058.
12. Wu YR, Chen CM, Chao CY, et al. Glucocerebrosidase gene mutation is a risk factor for early onset of Parkinson disease among Taiwanese. *J Neurol Neurosurg Psychiatry.* 2007;78(9):977-979.
13. De Marco EV, Annesi G, Tarantino P, et al. Glucocerebrosidase gene mutations are associated with Parkinson's disease in southern Italy. *Mov Disord.* 2008; 23(3):460-463.
14. Spitz M, Rozenberg R, Pereira LdaV, Reis Barbosa E. Association between Parkinson's disease and glucocerebrosidase mutations in Brazil. *Parkinsonism Relat Disord.* 2008;14(1):58-62.
15. Mata IF, Samii A, Schneer SH, et al. Glucocerebrosidase gene mutations: a risk factor for Lewy body disorders. *Arch Neurol.* 2008;65(3):379-382.
16. Gan-Or Z, Giladi N, Rozovski U, et al. Genotype-phenotype correlations between *GBA* mutations and Parkinson disease risk and onset. *Neurology.* 2008;70 (24):2277-2283.
17. Lwin A, Orvisky E, Goker-Alpan O, LaMarca ME, Sidransky E. Glucocerebrosidase mutations in subjects with parkinsonism. *Mol Genet Metab.* 2004;81(1): 70-73.
18. Eblan MJ, Nguyen J, Ziegler SG, et al. Glucocerebrosidase mutations are also found in subjects with early-onset parkinsonism from Venezuela. *Mov Disord.* 2006;21(2):282-283.
19. Ziegler SG, Eblan MJ, Gutti U, et al. Glucocerebrosidase mutations in Chinese subjects from Taiwan with sporadic Parkinson disease. *Mol Genet Metab.* 2007; 91(2):195-200.
20. Clark LN, Ross BM, Wang Y, et al. Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. *Neurology.* 2007;69(12):1270-1277.
21. Bras J, Paisan-Ruiz C, Guerreiro R, et al. Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal [published online ahead of print December 19, 2007] *Neurobiol Aging.* doi:10.1016/j.neurobiolaging .2007.11.016.
22. Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. *Neurology.* 1999; 52(6):1214-1220.
23. Koprivica V, Stone DL, Park JK, et al. Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. *Am J Hum Genet.* 2000; 66(6):1777-1786.
24. Takahashi Y, Seki N, Ishiura H, et al. Development of a high-throughput microarray-based resequencing system for neurological disorders and its application to molecular genetics of amyotrophic lateral sclerosis. *Arch Neurol.* 2008;65(10): 1326-1332.
25. Latham T, Grabowski GA, Theophilus BD, Smith FI. Complex alleles of the acid beta-glucosidase gene in Gaucher disease. *Am J Hum Genet.* 1990;47(1):79-86.
26. Orimo S, Ozawa E, Nakade S, Sugimoto T, Mizusawa H. (123)I-metaiodobenzylguanidine myocardial scintigraphy in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 1999;67(2):189-194.
27. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1982;140:566-572.
28. Eto Y, Ida H. Clinical and molecular characteristics of Japanese Gaucher disease. *Neurochem Res.* 1999;24(2):207-211.
29. Ida H, Rennert OM, Kawame H, Maekawa K, Eto Y. Mutation prevalence among 47 unrelated Japanese patients with Gaucher disease: identification of four novel mutations. *J Inherit Metab Dis.* 1997;20(1):67-73.

ORIGINAL ARTICLE

Multicenter Analysis of Glucocerebrosidase Mutations in Parkinson's Disease

E. Sidransky, M.A. Nalls, J.O. Aasly, J. Aharon-Peretz, G. Annesi, E.R. Barbosa, A. Bar-Shira, D. Berg, J. Bras, A. Brice, C.-M. Chen, L.N. Clark, C. Condroyer, E.V. De Marco, A. Dürr, M.J. Eblan, S. Fahn, M.J. Farrer, H.-C. Fung, Z. Gan-Or, T. Gasser, R. Gershoni-Baruch, N. Giladi, A. Griffith, T. Gurevich, C. Januario, P. Kropp, A.E. Lang, G.-J. Lee-Chen, S. Lesage, K. Marder, I.F. Mata, A. Mirelman, J. Mitsui, I. Mizuta, G. Nicoletti, C. Oliveira, R. Ottman, A. Orr-Urtreger, L.V. Pereira, A. Quattrone, E. Rogaeva, A. Rolfs, H. Rosenbaum, R. Rozenberg, A. Samii, T. Samaddar, C. Schulte, M. Sharma, A. Singleton, M. Spitz, E.-K. Tan, N. Tayebi, T. Toda, A.R. Troiano, S. Tsuji, M. Wittstock, T.G. Wolfsberg, Y.-R. Wu, C.P. Zabetian, Y. Zhao, and S.G. Ziegler

ABSTRACT

BACKGROUND

Recent studies indicate an increased frequency of mutations in the gene encoding glucocerebrosidase (*GBA*), a deficiency of which causes Gaucher's disease, among patients with Parkinson's disease. We aimed to ascertain the frequency of *GBA* mutations in an ethnically diverse group of patients with Parkinson's disease.

METHODS

Sixteen centers participated in our international, collaborative study: five from the Americas, six from Europe, two from Israel, and three from Asia. Each center genotyped a standard DNA panel to permit comparison of the genotyping results across centers. Genotypes and phenotypic data from a total of 5691 patients with Parkinson's disease (780 Ashkenazi Jews) and 4898 controls (387 Ashkenazi Jews) were analyzed, with multivariate logistic-regression models and the Mantel–Haenszel procedure used to estimate odds ratios across centers.

RESULTS

All 16 centers could detect two *GBA* mutations, L444P and N370S. Among Ashkenazi Jewish subjects, either mutation was found in 15% of patients and 3% of controls, and among non–Ashkenazi Jewish subjects, either mutation was found in 3% of patients and less than 1% of controls. *GBA* was fully sequenced for 1883 non–Ashkenazi Jewish patients, and mutations were identified in 7%, showing that limited mutation screening can miss half the mutant alleles. The odds ratio for any *GBA* mutation in patients versus controls was 5.43 across centers. As compared with patients who did not carry a *GBA* mutation, those with a *GBA* mutation presented earlier with the disease, were more likely to have affected relatives, and were more likely to have atypical clinical manifestations.

CONCLUSIONS

Data collected from 16 centers demonstrate that there is a strong association between *GBA* mutations and Parkinson's disease.

The authors' full names, degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Sidransky at the Section on Molecular Neurogenetics, National Human Genome Research Institute, Rm. 1A213, 35 Convent Dr., Bethesda, MD 20892-3708, or at sidranse@mail.nih.gov.

N Engl J Med 2009;361:1651-61.

Several lines of evidence suggest an association between parkinsonism and mutations in the gene encoding the lysosomal enzyme glucocerebrosidase (*GBA*), which is deficient in patients with Gaucher's disease. In this rare mendelian disorder, lysosomal accumulation of glucocerebroside results in a broad spectrum of disease manifestations including hepatosplenomegaly, anemia, thrombocytopenia, bone disease, and, at times, neurologic involvement.[1,2] Multiple independent studies have reported an increased frequency of *GBA* mutations in various cohorts of patients with parkinsonism.[3-21] However, several genomewide association studies have not identified this locus, and the degree of association has remained somewhat unclear, as many studies were not large enough to unequivocally label *GBA* mutations as risk alleles associated with Parkinson's disease.

Recognition of the association between *GBA* mutations and parkinsonism began in the clinic, with the identification of rare patients with Gaucher's disease who also had parkinsonian symptoms. Clinical descriptions appeared as case reports,[22,23] larger series of patients[24,25] and prospective studies.[26] Pedigree analyses revealed that relatives of patients with Gaucher's disease, many of whom were obligate heterozygotes, had an increased incidence of Parkinson's disease.[27,28]

Almost 300 *GBA* mutations have been identified in patients with Gaucher's disease, including missense, nonsense, and frameshift mutations as well as insertions, deletions, and complex alleles.[29] Various expression studies have shown that many of these mutations result in a significant loss of glucocerebrosidase activity.[30] The *GBA* gene, located on 1q21–22, includes 11 exons and has a similar pseudogene 16 kb downstream. The 85-kb region surrounding *GBA* is particularly gene-rich, encompassing seven genes and two pseudogenes.[31] Recombination within and around the *GBA* locus occurs relatively frequently,[32] complicating genotype analyses.

The frequency and distribution of *GBA* mutations vary among populations. Among Ashkenazi Jews, for whom the carrier frequency is between 1 person in 12 and 1 in 16,[1,6] one common mutation, N370S, accounts for approximately 70% of mutant alleles.[1] In other ethnic groups, the carrier frequency is usually less than 1%, and the associated mutations are more diverse. Although N370S is also common in European populations and is exclusively associated with non-neuronopathic Gaucher's disease, it has not been encountered among Asians.[33] A second relatively frequent and panethnic mutation associated with neuronopathic Gaucher's disease is L444P, which can be a point mutation or part of a complex allele encompassing a portion of the pseudogene sequence.[32]

This international collaborative study of *GBA* mutations in patients with Parkinson's disease was undertaken to better ascertain the frequency of *GBA* mutations by pooling data for individual persons from 16 centers, representing 4 continents, including 5691 patients and 4898 controls. Findings based on the data from 13 of the 16 centers (4185 of the case patients and 3597 of the controls) have been published previously.[5-16,19] Our goals were to establish the combined frequency of *GBA* mutations among sites, to explore the range of *GBA* mutations encountered, and to identify clinical features shared among *GBA* mutation carriers.

METHODS

STUDY SUBJECTS AND PROCEDURES

Researchers known to be genotyping cohorts of patients with Parkinson's disease for *GBA* mutations were solicited for this collaboration. Sixteen centers participated, and data were collected and analyzed at the National Human Genome Research Institute (NHGRI), Bethesda, Maryland (Table 1). The subjects were from a total of 16 centers: 4 in North America, 1 in South America, 3 in Asia, 2 in Israel, and 6 in Europe. Ethnicity was self-reported. Ashkenazi Jews provided the country of origin of their grandparents. Written informed consent was obtained from all subjects under the supervision of each local ethics committee. All patients fulfilled the clinical diagnostic criteria for Parkinson's disease of the U.K. Parkinson's Brain Bank.[34]

Because the detection methods and number of mutations that could be identified varied greatly from center to center, 12 standard DNA samples from patients with Gaucher's disease were genotyped by each center, and the results were analyzed at the NHGRI. All centers could reliably detect mutations N370S and L444P, unless the mutant allele included large stretches of the *GBA* pseudogene sequence. Some centers could identify three to eight specific selected point mutations (Table 1). In addition, five centers sequenced all the exons of *GBA*, and a sixth sequenced exons for a sub-

group of subjects. Thus, results evaluating two mutations (N370S and L444P), six to eight mutations, and the entire coding sequence were analyzed separately (see Table 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Two frequent *GBA* variants, E326K and T369M, were evaluated but not included as mutant alleles.

The extent of clinical data collected from each site varied (Table 1). Some study centers provided only the age, family history, sex, ethnic group, and diagnosis of each subject, whereas others reported more complete data, including symptoms at presentation, age at disease onset, specific clinical manifestations, response to medications, and standardized scores on the Hoehn and Yahr staging system, the Unified Parkinson's Disease Rating Scale, or both. Only one proband was studied per family, and patients with diagnoses other than Parkinson's disease were excluded.

Controls were screened for signs or symptoms of parkinsonism, and centers attempted to match controls and patients with regard to age, sex, and ethnic group. Controls with a family history of Parkinson's disease were removed because of their increased risk of having parkinsonism ultimately develop.

STUDY DESIGN

The a priori aim in this study was to conduct a combined analysis of risk associated with *GBA* mutations from various centers. Our analyses used patient-level data from both published and unpublished studies. We assessed associations between Parkinson's disease and *GBA* mutations in not only all available genotyped samples, but also in distinct subgroups of the total population (see the Supplementary Appendix for detailed analyses of subgroups).

STATISTICAL ANALYSIS

Descriptive statistics were calculated for each study center. Available data on family history of Parkinson's disease and seven clinical features of Parkinson's disease (asymmetric onset, bradykinesia, cognitive changes, orthostatic hypotension, postural instability, resting tremor, and rigidity) were analyzed. Clinical features were compared between case patients carrying *GBA* mutations and those without mutations, means were compared with the use of a two-tailed Student's t-test, and frequency differences were compared between patients with and those without *GBA* mutations by means of the chi-square test.

We first estimated the odds ratios for mutations in *GBA* across cohorts. To accomplish this, we conducted fixed-effects Mantel–Haenszel analyses using all available case patients and controls from each study center. These models calculate crude odds ratios and confidence intervals from counts of mutations in case patients and controls in each study and then combine the odds ratios from study-specific two-by-two tables into a summary measure. Three Mantel–Haenszel analyses were performed, using any mutation, N370S, and L444P as the separate independent variables. No multiple-test correction was used in our analyses.

The heterogeneity of effects in the Mantel–Haenszel analyses were evaluated using Woolf's test for heterogeneity.[35] Analyses of the possible interactions contributing to the heterogeneity of odds ratios were limited by the data available for analysis, so several additional Mantel–Haenszel analyses were carried out. First, we excluded data from Japan and Norway, both together and independently, on the basis of the assumption that general genetic homogeneity in these samples could be compounded by the active recruitment of family members of case patients, influencing the independence (or in the Norwegian cohort, the nonindependence) of differences in mutation frequency between case patients and controls. An additional analysis, performed by omitting data from Tel Aviv, Israel — the center with the most robust odds ratio with respect to the standard error of the estimate — confirmed that these data were not inflating the combined odds ratio.

To follow up on the results from the Mantel–Haenszel analyses, multivariate logistic-regression models were used to examine the association between *GBA* mutations and Parkinson's disease in subgroups stratified on the basis of the level of sequencing coverage and Ashkenazi Jewish or non–Ashkenazi Jewish ethnicity. These stratified models combined data for subjects across studies, including all subjects with complete data for covariates of self-reported ethnic group, age at the time of sampling, and sex. For regression modeling and covariate descriptions, data for 702 patients and 570 controls were excluded because age, sex, or ethnic group was unknown. Identical parameters were used in logistic-regression models testing the association of Parkinson's disease with variants E326K and T369M. Chi-square tests of

heterogeneity were used to compare effect-size differences between models stratifying Ashkenazi Jewish and non–Ashkenazi Jewish ethnicity assessing the risks attributable to all *GBA* mutations and N370S and L444P only, across all levels of sequencing coverage. Detailed results and a description of these regression models are given in the Supplementary Appendix.

All data analyses were conducted with the use of R software (version 2.8.0).[36] Source code for plotting of the meta-analysis results is available in the software package "r.meta" (available from http://cran.cnr.berkeley.edu).

RESULTS

Genotyping was performed for a total of 5691 patients with Parkinson's disease: 780 Ashkenazi Jewish patients and 4911 patients with no known Ashkenazi Jewish ancestry. The 4898 controls for whom genotyping was performed included 387 Ashkenazi Jewish persons and 4511 from other ethnic groups. Table 1 lists the frequency of mutations and baseline characteristics of cohorts at each center. Full sequencing was performed on DNA samples obtained from 2060 patients and 1677 controls.

Table 1. Baseline Data for the Study Groups, According to Center.*

Center	No. of Subjects (No. Who Are Ashkenazi Jewish)†	Sex Ratio	Mean Age of Subjects at Time of Collection	Mutations	Mutations Screened	Clinical Data Provided‡
		M:F (unknown)	*yr*	*no. of subjects (%)*		
Brazil					N370S, L444P, G377S	Limited
Patients	65 (0)	42:23	54.1	4 (6)		
Controls	264 (0)	169:95	54.4	0		
New York					Full sequencing	Full
Patients	275 (177)	171:104	65.6	34 (12)		
Controls	140 (65)	73:67	62.8	3 (2)		
France					N370S, L444P, D409H	Full
Patients	297 (0)	185:112	57.8	12 (4)		
Controls	251 (0)	142:109	57.8	1 (<1)		
Haifa, Israel					N370S, L444P, c.84dupG, V394L, IVS2+1G→A, R496H	Partial
Patients	162 (162)	94:65 (3)	68.5	40 (25)		
Controls	NP	NP	NP	NP		
Italy					L444P, N370S	Full
Patients	395 (0)	244:151	66.5	11 (3)		
Controls	483 (0)	180:303	56.9	1 (<1)		
Norway					L444P, N370S	Limited
Patients	311 (0)	186:123 (2)	NP	7 (2)		
Controls	473 (0)	267:206	64.1	8 (2)		
NHGRI					Full sequencing	Partial
Patients	539 (0)	275:166 (98)	73.5	29 (5)		
Controls	209 (1)	100:109	67.6	6 (3)		
Portugal					Full sequencing	Limited
Patients	231 (0)	110:121	65.3	15 (6)		
Controls	482 (0)	Unknown	65.5	6 (1)		
Rostock, Germany					Full sequencing	Partial
Patients	298 (0)	190:108	64.3	18 (6)		
Controls	212 (0)	105:107	74.5	5 (2)		

Table 1. (Continued.)

Center	No. of Subjects (No. Who Are Ashkenazi Jewish)†	Sex Ratio	Mean Age of Subjects at Time of Collection	Mutations	Mutations Screened	Clinical Data Provided‡
		M:F (unknown)	*yr*	*no. of subjects (%)*		
Singapore					L444P, N370S	Partial
Patients	329 (0)	170:158 (1)	70.3	8 (2)		
Controls	201 (0)	99:102	64.2	0		
Taiwan					N370S, L444P, Rec*Nci*I, R120W, some full sequencing (183 patients and 88 controls)	Full
Patients	559 (0)	304:255	69.1	22 (4)		
Controls	377 (0)	198:179	60.5	4 (1)		
Tel Aviv, Israel					c.84dupG, IVS2+1G→A, N370S, V394L, D409H, L444P, R496H, RecTL	Limited
Patients	420 (419)	262:158	68.0	81 (19)		
Controls	321 (321)	159:162	65.3	13 (4)		
Japan					Full sequencing	Limited
Patients	534 (0)	282:252	65.3	50 (9)		
Controls	546 (0)	294:252	44.8	2 (<1)		
Tübingen, Germany					L444P, N370S	Partial
Patients	377 (0)	222:155	64.8	12 (3)		
Controls	325 (0)	192:132 (1)	58.3	0		
Toronto					N370S, K178T, L444P, c.84dupG, R329C, IVS2+1G→A, Rec*Nci*I	Partial
Patients	88 (2)	51:37	55.0	5 (6)		
Controls	96 (0)	27:69	69.6	1 (1)		
Seattle					N370S, L444P, Rec*Nci*I, L444R	Full
Patients	811 (20)	607:204	66.9	24 (3)		
Controls	518 (0)	193:324 (1)	65.2	2 (<1)		

* NHGRI denotes National Human Genome Research Institute, and NP not provided.
† Ashkenazi Jewish ancestry was self-reported.
‡ "Limited" clinical data were age, family history, sex, ethnic group, and diagnosis of each subject; "full" were the limited clinical data plus symptoms at presentation, age at disease onset, specific clinical manifestations, response to medications, and standardized scores on the Hoehn and Yahr staging system, the Unified Parkinson's Disease Rating Scale, or both; and "partial" were the limited clinical data plus some but not all of the full clinical data.

Figure 1 shows the odds ratios for mutations in patients versus controls, standard errors, and confidence limits from the Mantel–Haenszel procedure performed on data from each independent study center; the odds were also combined across the centers. Data from four centers were excluded in this analysis, because individual controls were not provided (Haifa, Israel) or mutations were not found among the controls (Brazil, Singapore, and Tübingen, Germany).

Results from the analysis of any *GBA* mutation as an independent variable (Fig. 1A) show that each center had an overrepresentation of mutations among patients as compared with controls, although the confidence intervals varied considerably. Eight centers had odds ratios greater than 5. The overall combined odds ratio shows how greatly the confidence interval is reduced when data from the individual centers are combined. Individual odds ratios for the *GBA* point mutations L444P and N370S are also shown (Fig. 1B and 1C, respectively). Although the odds ratios for each mutation are overwhelmingly positive, they were higher for L444P.

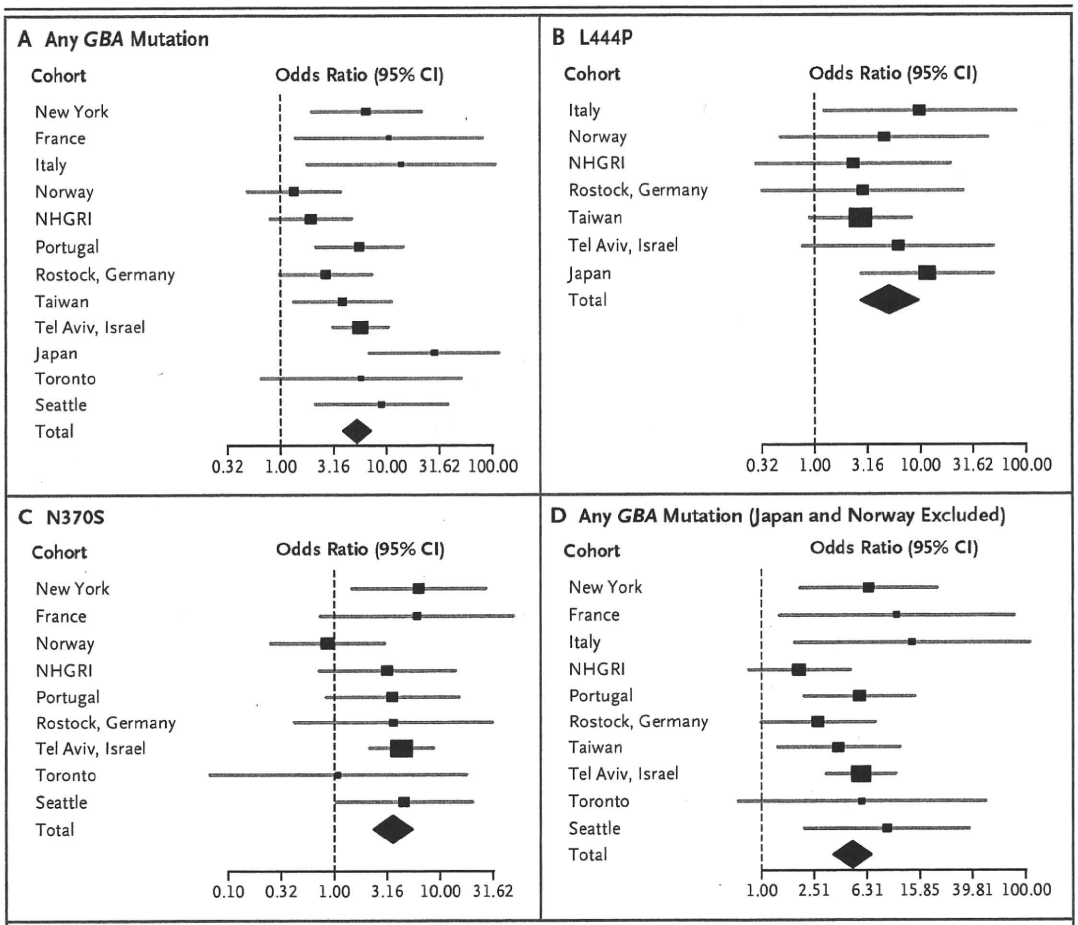

Figure 1. Odds Ratios for *GBA* Mutations among Patients, as Compared with Controls, at Each Study Center and Overall.

Shown are the combined estimates (on a $\log_{10}$ scale) of the risk of possessing any *GBA* mutation (Panel A), the L444P mutation (Panel B), the N370S mutation (Panel C), and any *GBA* mutation when all data from the Japan and Norway study centers are excluded (Panel D). The squares represent point estimates, with the height of the square inversely proportional to the standard error of the estimates. The horizontal lines indicate the 95% confidence intervals of the estimates. The diamonds represent the summary odds ratio, with the width indicating the 95% confidence interval. Data from four centers were excluded, because individual controls were not provided (Haifa, Israel) or mutations were not found among the controls (Brazil, Singapore, and Tübingen, Germany). NHGRI denotes National Human Genome Research Institute.

Combined odds ratios from Mantel–Haenszel analyses were homogeneous for L444P and N370S (P=0.35 and P=0.50, respectively, by Woolf's test for heterogeneity). This finding allows for confidence in reporting the Mantel–Haenszel combined odds ratios for N370S (odds ratio, 3.96; 95% confidence interval [CI], 2.60 to 6.02) and L444P (odds ratio, 6.73; 95% CI, 4.50 to 15.42). However, the Mantel–Haenszel odds ratio for any *GBA* mutation was significantly heterogeneous (P=0.02 by Woolf's test for heterogeneity). After excluding data from Japan and Norway, the odds ratio was slightly attenuated (5.43; 95% CI, 3.89 to 7.57), but the heterogeneity was no longer significant (P=0.41 by Woolf's test for heterogeneity), and further excluding the Tel Aviv data resulted in a homogeneous odds ratio (P=0.34 by Woolf's test for heterogeneity) of 5.30 (95% CI, 3.59 to 7.94). Independent exclusions of data from either Norway or Japan resulted in homogeneous odds ratios (Japan data excluded: odds ratio, 4.91; 95% CI, 3.60 to 6.70; P=0.14 by Woolf's test for heterogeneity; Norway data excluded: odds ratio, 6.35, 95% CI, 4.60 to 8.75; P=0.10 by Woolf's test for heterogeneity) (Fig. 1D). Thus, we can confidently report a Mantel–Haenszel odds ratio of 5.43 for

GBA mutations, in patients versus controls, even when these centers are excluded.

Overall, when screening solely for N370S and L444P, one of these two mutations was found in 15% of Ashkenazi Jewish patients as compared with 3% of Ashkenazi Jewish controls and in 3% of non–Ashkenazi Jewish patients as compared with less than 1% of non–Ashkenazi Jewish controls. However, the frequency of N370S in Ashkenazi Jewish patients differed among the centers. N370S was not seen in any Asian patients or controls.

All Ashkenazi Jewish subjects were screened for the presence or absence of six to eight different *GBA* mutations (Table 1). Including this larger number of screened mutations (rather than just L444P and N370S) increased the odds ratio for any mutation, among patients as compared with controls, to 6.48 (95% CI, 3.78 to 11.09). The distribution of specific *GBA* mutations among Ashkenazi Jewish patients with Parkinson's disease, of whom approximately 20% carried a *GBA* mutation, is shown in Figure 2A. Approximately 20% of the identified mutant alleles were neither L444P nor N370S. Among the non–Ashkenazi Jewish subjects, the entire *GBA* coding region was sequenced in 1883 patients and 1611 controls; the odds ratio for any *GBA* mutation was 6.51 (95% CI, 3.62 to 11.74). The distribution of mutations identified among non–Ashkenazi Jewish patients (Fig. 2B) indicates that as many as 45% of mutant alleles could be missed if focusing solely on N370S and L444P. Moreover, 20 of 43 non–Ashkenazi Jewish patients who had complete sequence data and had L444P carried other pseudogene changes and hence had recombinant alleles.

The full-sequencing data identified two common *GBA* variants, E326K and T369M, which are not pathogenic in patients with Gaucher's disease.[37] Neither mutation had a significant association with Parkinson's disease (see the Supplementary Appendix).

Seventeen patients (13 of whom were Ashkenazi Jews) carried two *GBA* mutations. Thirteen of these patients had the N370S/N370S genotype, two had the N370S/R496H genotype, one had the N370S/V394L genotype, and one had the N370S/L444P genotype.

Age at onset was found to be significantly lower among patients with *GBA* mutations as compared with those without mutations ($P<0.001$), with a mean age of 54.9 years and 58.8 years, respectively. The mean period from diagnosis to evaluation was 7.8 years in both groups.

Information about a family history of parkinsonism was available for 4401 of the patients studied. In all, 18% of patients without *GBA* mutations reported having a first- or second-degree relative with Parkinson's disease, as compared with 24% of patients with a *GBA* mutation ($P=0.006$).

The frequencies of seven clinical findings associated with Parkinson's disease in patients with *GBA* mutations and those without mutations are shown in Figure 3. In general, the symptom profile for the two groups was similar, although mutations were associated with a significantly lower frequency of asymmetric onset ($P<0.001$), bradykinesia ($P<0.001$), resting tremor ($P=0.03$), and rigidity ($P<0.001$). There was no significant difference between the two groups in the frequency of orthostatic hypotension ($P=0.26$) or postural instability ($P=0.12$). Among the 1948 patients for whom the presence or absence of cognitive changes was recorded, changes were reported for 26% of patients with mutations and 19% of those without ($P=0.007$).

DISCUSSION

The results of our analysis indicate that there is an association between *GBA* mutations and Parkinson's disease. The combined data show that this finding is not exclusive to a specific ethnic group or a specific *GBA* mutation. Although mutations in *GBA* are most likely a susceptibility factor rather than a mendelian cause of Parkinson's disease, the high frequency of mutations among ethnically diverse, heterogeneous cohorts of patients with Parkinson's disease makes the mutations in this gene the most common genetic risk factor for Parkinson's disease that have been identified to date.

The major limitation of this study was the unavoidable differences in data-ascertainment methods among the study centers. Moreover, some sites were more successful than others in matching case patients and controls with regard to age and sex. To ensure that the analysis was not driven by a small subgroup of centers, we evaluated the variance in odds ratios across centers. Excluding the center with the most precise estimate (Tel Aviv) and the centers with the most extreme odds ratios (Norway and Japan) resulted in only a slight at-

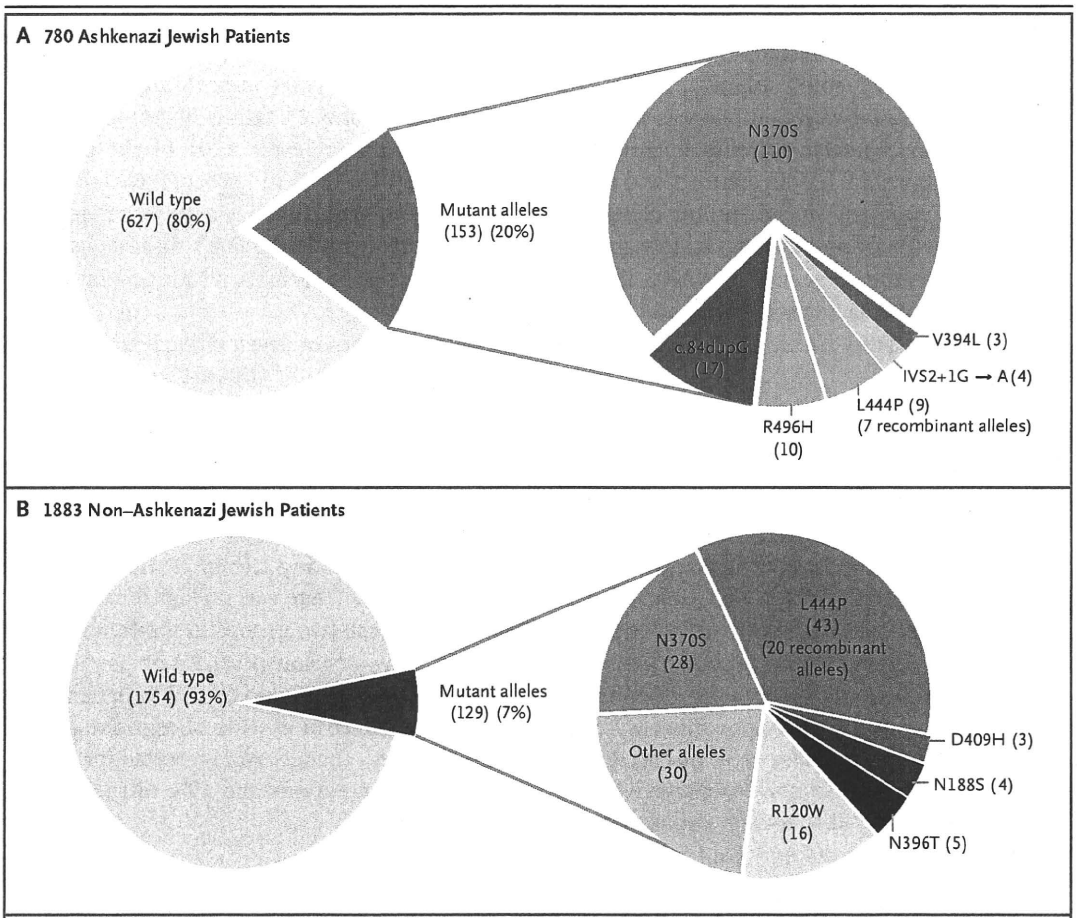

Figure 2. Distributions of Mutations among Study Patients of Ashkenazi Jewish Ancestry and Those of Non–Ashkenazi Jewish Ancestry.

Data are shown for patients with Parkinson's disease, with the numbers (and percentages) of patients in parentheses. A total of 20% of Ashkenazi Jewish patients were found to carry a *GBA* mutation (Panel A). Among non–Ashkenazi Jewish patients for whom full sequencing was performed, a total of 7% carried *GBA* mutations (Panel B). "Other alleles" are D140H, F216Y, N370K, D380N, F213I, G193W, G344S, K198T, K27R, L336P, L371I, Q497R, R131C, R131S, R163Q, R163X, R239C, R262H, R39H, R44C, S271G, T323I, R329C, R496C, and V460M.

tenuation of the combined odds ratio for any *GBA* mutation among patients as compared with controls, which in all analyses remained 5.43 or higher.

This study confirms the need to sequence all exons to accurately ascertain the frequency of *GBA* mutations among both patients and controls. The combined odds ratios may be underestimated, because so many of the centers performed limited mutation analyses. Our data show that among non–Ashkenazi Jewish patients, as many as 45% of mutant alleles can be missed when screening for only the N370S and L444P mutations. Furthermore, analysis of specific mutations may produce a serious bias. Among the 1883 non–Ashkenazi Jewish patients who had sequenced samples, *GBA* mutations were found in 7% (odds ratio vs. controls, 6.51), although only 35% of the samples included in our entire analysis were fully sequenced. An adequate sample size and accurate genotyping of control samples are imperative to avoid underestimating the frequencies of rare variants.

Despite the difficulty in determining the phenotypic profile associated with the *GBA* mutations found in this study, which intentionally included only patients that met diagnostic criteria for Parkinson's disease, some trends are apparent. As compared with patients who did not have a *GBA* mutation, those with a mutation presented on av-

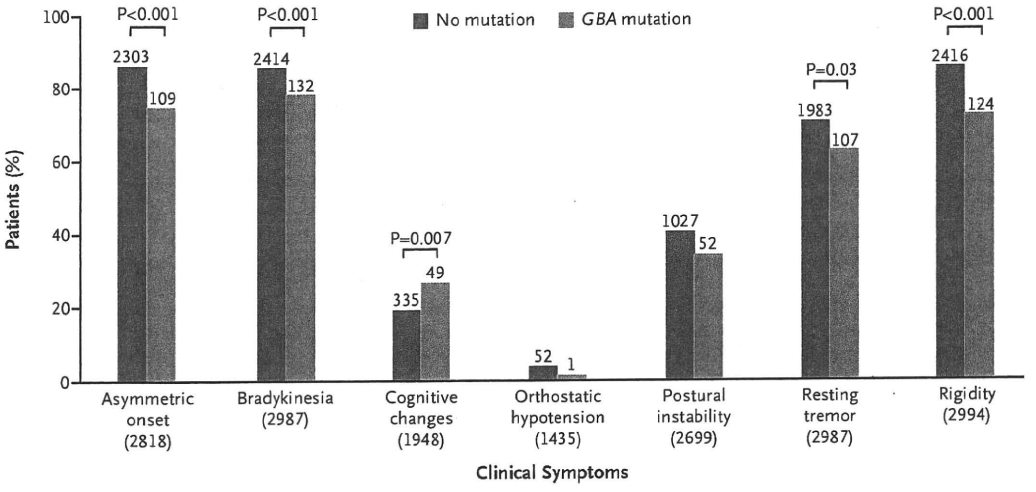

Figure 3. Clinical Manifestations of Parkinson's Disease in Study Patients with a *GBA* Mutation and Those without a Mutation.

The numbers of patients for whom the presence or absence of each symptom was recorded are indicated below the pairs of bars. The number of patients with the symptom is given at the top of each bar.

erage 4 years earlier, were more likely to have a family history of Parkinson's disease, and had a lower incidence of bradykinesia and resting tremor and a higher incidence of cognitive changes, a trend supporting earlier reports.[3,5,6,18,19,24,25] However, because of the exclusion criteria used in the analysis, our findings do not accurately reflect the full spectrum of parkinsonian symptoms associated with *GBA* mutations. An increased frequency of *GBA* mutations has also been described in cohorts with Lewy-body disorders,[38-40] although not in multiple-system atrophy,[41] and further studies are warranted.

Even though our data provide validation of *GBA* as a risk factor associated with Parkinson's disease, the ultimate challenge is to establish the mechanisms contributing to this association. Both a gain-of-function mechanism due to enhanced protein aggregation or lysosomal dysfunction[42] and a loss of function related to fluctuations in levels of ceramide[43] have been postulated. Further research to elucidate the pathophysiology of both Gaucher's disease and Parkinson's disease will facilitate more accurate genetic counseling and lead to the development of new therapeutic strategies.

Supported in part by the Intramural Research Programs of NHGRI and the National Institute on Aging (project Z01AG000957-05), National Institutes of Health; grants from the National Institute of Neurological Disorders and Stroke and the National Center for Research Resources (NS050487, NS060113, and NS40256); grants from the Parkinson's Disease Foundation; a Merit Review Award from the Department of Veterans Affairs, Seattle; a grant from the German National Genome Network, German Ministry for Education and Research (01GS08134); a grant from the French Parkinson's Disease Genetics Study Group; a grant from the Duke–National University of Singapore (NUS) Graduate Medical School, National Medical Research Council, Biomedical Research Council, Singapore Millennium Foundation; a grant from SingHealth, Singapore General Hospital, National Neuroscience Institute; the Tel Aviv Sourasky Medical Center Grant of Excellence; grants from the Wolfson and Kahn Foundations; a grant from the Portuguese Fundação para a Ciência e a Tecnologia; a grant from the National Science Council, Executive Yuan, Taiwan (NSC-95-2314-B-182A-061); and a Grant-in-Aid for Scientific Research on Priority Areas, Applied Genomics, and Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Dr. Berg reports receiving lecture fees from UCB, GlaxoSmithKline, Teva, and Lundbeck, advisory fees from Novartis, UCB, GlaxoSmithKline, and Teva, and grant support from Janssen Pharmaceuticals and Teva Pharma; Dr. Fahn, receiving lecture fees from Boehringer Ingelheim and consulting fees from Anteres Pharma, Boehringer Ingelheim, Elsai, EMD Serono, Genzyme, Impax Pharmaceuticals, Intec Pharma, Kyowa, Novartis, Teva, Valeant, and Vernalis and serving on the Merz Pharma Data and Safety Monitoring Board; Dr. Farrer, holding a provisional patent for RNA interference as a treatment for neurodegenerative disease and a patent for LRRK2 mutations and receiving research support from Lundbeck and Isis Pharmaceuticals and lecture fees from Genzyme and GlaxoSmithKline; Dr. Gasser, receiving research support from Novartis, consulting fees from Cephalon Pharma and Merck Serono, and lecture fees from Boehringer Ingelheim, Merck Serono, Novartis, and Schwarz Pharma and holding a patent for the detection of LRRK2 mutations; Dr. Griffith, receiving consulting fees from Teva and lecture fees from Teva and Novartis; Dr. Lang, receiving advisory fees from Biovail, Boehringer Ingelheim, Cephalon, Ceregene,