Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes

Vladimir Vacic[1,†], Laurie J. Ozelius[2,3,‡], Lorraine N. Clark[4,5,‡], Anat Bar-Shira[9], Mali Gana-Weisz[9], Tanya Gurevich[10,11], Alexander Gusev[1], Merav Kedmi[9,¶], Eimear E. Kenny[1,§], Xinmin Liu[4], Helen Mejia-Santana[6], Anat Mirelman[10], Deborah Raymond[12], Rachel Saunders-Pullman[12,13], Robert J. Desnick[2], Gil Atzmon[14,15,16], Edward R. Burns[14], Harry Ostrer[15,17,18], Hakon Hakonarson[21], Aviv Bergman[19], Nir Barzilai[14,15,16], Ariel Darvasi[22], Inga Peter[2], Saurav Guha[2,23], Todd Lencz[23,24,20,25,26], Nir Giladi[10,11], Karen Marder[5,6,7,8], Itsik Pe'er[1,*], Susan B. Bressman[12,13] and Avi Orr-Urtreger[9,11]

[1]Department of Computer Science, Columbia University, New York, NY, USA, [2]Department of Genetics and Genomic Sciences and [3]Department of Neurology, Mount Sinai School of Medicine, New York, NY, USA, [4]Department of Pathology and Cell Biology, [5]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, [6]Gertrude H. Sergievsky Center, [7]Department of Neurology and [8]Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY, USA, [9]Genetic Institute and [10]Department of Neurology, Movement Disorders Unit and Parkinson Center, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, [11]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, [12]Mirken Department of Neurology, Beth Israel Medical Center, New York, NY, USA, [13]The Saul R. Korey Department of Neurology, [14]Department of Medicine, [15]Department of Genetics, [16]Institute for Aging Research, [17]Department of Pathology, [18]Department of Pediatrics, [19]Department of Systems and Computational Biology and [20]Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine, Bronx, NY, USA, [21]Center for Applied Genomics, The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, PA, USA, [22]Department of Genetics, Institute of Life Sciences, Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel, [23]Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY, USA, [24]Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, NY, USA and [25]Department of Psychiatry and [26]Department of Molecular Medicine, Hofstra University School of Medicine, Hempstead, NY, USA

The recent series of large genome-wide association studies in European and Japanese cohorts established that Parkinson disease (PD) has a substantial genetic component. To further investigate the genetic landscape of PD, we performed a genome-wide scan in the largest to date Ashkenazi Jewish cohort of 1130 Parkinson patients and 2611 pooled controls. Motivated by the reduced disease allele heterogeneity and a high degree of identical-by-descent (IBD) haplotype sharing in this founder population, we conducted a haplotype association study based on mapping of shared IBD segments. We observed significant haplotype association signals at three previously implicated Parkinson loci: *LRRK2* (OR = 12.05, $P = 1.23 \times 10^{-56}$), *MAPT* (OR = 0.62, $P = 1.78 \times 10^{-11}$) and *GBA* (multiple distinct haplotypes, OR > 8.28, $P = 1.13 \times 10^{-11}$ and OR = 2.50, $P = 1.22 \times 10^{-9}$). In addition, we

---

*To whom correspondence should be addressed at: Department of Computer Science, Fu School of Engineering, Columbia University, 505 Computer Science Building, 500 W 120th Street, Mailcode 0401, New York, NY 10027, USA. Tel: +1 2129397135; Fax: +1 2126660140; Email: itsik@cs.columbia.edu

†Present address: New York Genome Center, New York, NY, USA.
‡These authors contributed equally.
¶Present address: Weizmann Institute of Science, Rehovot, Israel.
§Present address: Department of Genetics and Genomic Sciences, Mt. Sinai School of Medicine, New York, NY, USA.

identified a novel association signal on chr2q14.3 coming from a rare haplotype (OR = 22.58, *P* = 1.21 × 10^{-10}) and replicated it in a secondary cohort of 306 Ashkenazi PD cases and 2583 controls. Our results highlight the power of our haplotype association method, particularly useful in studies of founder populations, and reaffirm the benefits of studying complex diseases in Ashkenazi Jewish cohorts.

INTRODUCTION

Parkinson disease (PD; OMIM: #168600) is a common neurodegenerative disorder, affecting ∼0.5–1% of the population in the 65–69 age group and 1–3% of individuals 80 years of age and older ([1]). Causes of PD remain largely unknown, and the majority of patients do not report a family history of PD. Recent estimates of heritability of PD based on genome-wide complex trait analysis are between 0.16 and 0.49 (median 0.23), depending on the cohort ([2]). Causal mutations have been identified in >10 genes, and the recent string of large genome-wide association studies (GWASs) in European and Japanese cohorts found >30 genes and susceptibility loci associated with this disease ([3–7]).

We here investigate the genetic component of PD in Ashkenazi Jews, a well-characterized, homogeneous founder population particularly useful in genetic studies of diseases ([8],[9]) and an important patient population in its own right. Improved control for population stratification ([10],[11]) and reduced allelic heterogeneity facilitate identification of disease-associated variants in Ashkenazi Jews. About a third of Ashkenazi PD cases carry a mutation in either the leucine-rich repeat kinase 2 (*LRRK2*) or β-glucocerebrosidase (*GBA*) genes ([12]). The p.G2019S mutation in the *LRRK2* gene occurs in 9.9–14.8% of Ashkenazi cases and 1.3–2.4% controls, compared with 2.7–3.1% of non-Jewish European cases and 0.4% of respective controls ([13–15]). Mutations in the *GBA* gene occur in 13.7–17.9% of Ashkenazi cases and 4.2–6.3% controls, compared with 5 and <1% of non-Jewish European cases and controls ([12],[16–18]). The most frequent of these mutations is p.N370S (10.9% Ashkenazi cases, 5.9% controls) also present are c.84dupG (84GG, 1.9% cases, 0.2% controls) and p.R496H (1.67% cases, not observed in controls) ([12],[18],[19]). Another advantageous genetic feature of this founder population is the high frequency of identical-by-decent (IBD) segments sharing among apparently unrelated Ashkenazim ([20]), which facilitates detection and analysis of disease-associated haplotypes.

We performed a GWAS in an Ashkenazi Jewish cohort of 1130 Parkinson patients collected by a consortium of institutions from New York and Tel Aviv and 2611 pooled controls. This is to date the largest study of PD in this or any other founder population. We conducted a genome-wide scan using a haplotype association method DASH ([21]), and concurrently, we analyzed the association of SNPs imputed using a custom Ashkenazi reference panel ([22]). Both approaches identify three genome-wide significant loci: *LRRK2*, *MAPT* and *GBA*. Analysis based on IBD segments refines the association signal from SNP-based analysis and implicates a novel genome-wide significant rare haplotype located on chr2q14.3, which we replicated in a secondary cohort of 306 Ashkenazi PD cases and 2583 controls. We further replicated three previously published Parkinson loci, *SNCA*, *MCCC1/LAMP3* and *PARK16*. Phenotypic variance

explained by genome-wide significant and replicated loci is 16.5% using IBD segments and 10.9% using SNP markers, compared with 6–7% in Europeans ([3]). Our findings reaffirm the benefits of using the Ashkenazi Jewish founder population in studies of complex diseases and highlight the power of our haplotype association method, which is particularly useful in genetic studies of founder populations.

RESULTS

We analyzed genome-wide association of shared IBD segments identified by the use of GERMLINE ([23]) and clustered using DASH ([21]). We concurrently analyzed association of imputed SNPs using EMMAX ([24]), a mixed model well suited to studies of populations with high degree of cryptic relatedness ([25]), as is the case with Ashkenazi Jews. Association results summarized as Manhattan and QQ plots are shown in Figure [1], Supplementary Material, Figures S1 and S2, and the top associated haplotypes and SNPs are listed in Tables [1] and [2], respectively.

Association

LRRK2

The strongest SNP association signal was detected at the *LRRK2* locus [rs1442190, OR = 3.72 (2.98, 4.64), *P* = 1.53 × 10^{-27}, Fig. [2]A]. The best associated DASH segment refines this signal with OR = 12.05 (8.35, 17.41) and *P* = 1.23 × 10^{-56}. We found 170 copies of this segment in PD cases and 35 in controls, which accounted for all known p.G2019S PD cases in our cohort (independently genotyped, see Supplementary Material, Table S2). The two pericentromeric association signals on chr12 reported in Tables [1] and [2] are due to long-range LD in this region (chr12:33–40 Mb) ([26]). There was no residual signal on chr12 after conditioning on the independently genotyped p.G2019S carrier status, indicating that in our cohort, p.G2019S is responsible for the association at the *LRRK2* locus.

Chr17q21.31/MAPT

Strong association signal coming from both SNPs [rs17577094, OR = 0.64 (0.56, 0.72), *P* = 4.51 × 10^{-10}] and DASH segments [OR = 0.62 (0.54, 0.72), *P* = 1.78 × 10^{-11}, shared by 423 cases and 1351 controls] was observed within the 900-kb inversion polymorphism on chr17q21.31 ([27]). Association to PD at this locus has been previously reported in a European cohort (rs393152; OR = 0.77; *P* = 1.95 × 10^{-16}) ([7]) and replicated in a subset of our Ashkenazi samples [OR = 0.54 (0.49, 0.74), *P* = 7.16 × 10^{-5}] ([28]). The inverted haplotype (H2) frequency in Europeans varies in a gradient away from the Mediterranean Sea (37.5% in Sardinians to 4.3% in Finns) and is 25.6% in Ashkenazi Jews ([29]). H1 and H2 haplotypes do not recombine and are well tagged by common SNPs, which causes a peculiar LD
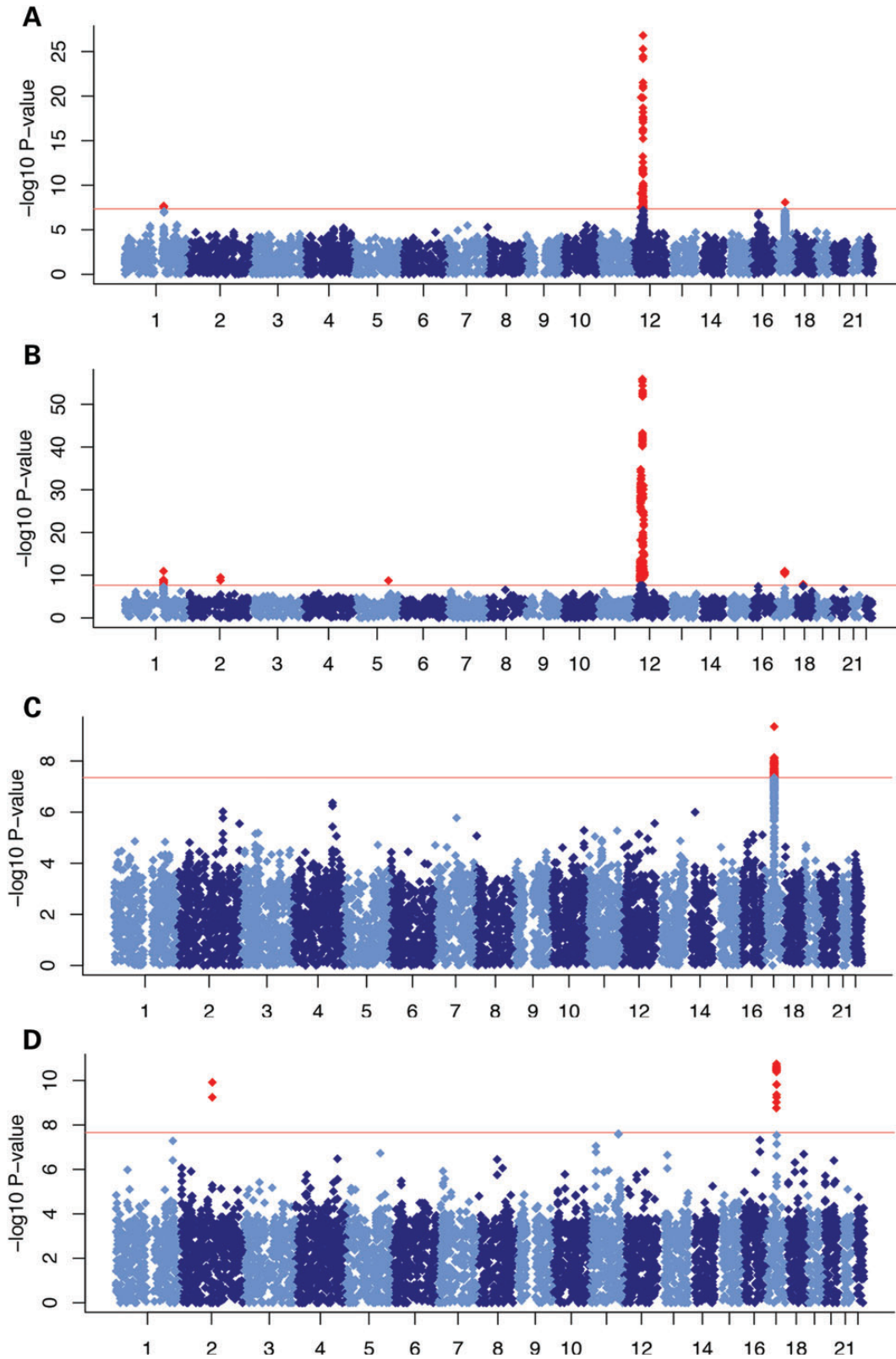
Figure 1. Manhattan plots showing the association of (**A**) imputed SNPs and (**B**) DASH segments. Lower panels show association results conditioned on the *LRRK2* and *GBA* mutation carrier status, for (**C**) imputed SNPs and (**D**) DASH segments. Markers color-coded red surpass the Bonferroni-corrected genome-wide significance thresholds.

Table 1. DASH segment association results with P-values of <10⁻⁷ based on either Fisher's exact (single stratum analysis) or exact conditional test (samples stratified by the *LRRK2* and *GBA* carrier status)

Chromosome	Start	End	Length (kb)	Gene/locus	PD (N)	Controls (N)	PD (%)	Controls (%)	OR (95% CI)	Fisher's exact P	Exact conditional P	P-value (replicated)
12	38 614 682	38 829 030	214	LRRK2	170	35	7.52	0.67	12.05 (8.35, 17.41)	1.23×10^{-56}	0.29	–
12	32 133 518	32 230 612	97	LRRK2	114	28	5.04	0.54	9.85 (6.50, 14.94)	1.73×10^{-35}	0.70	–
1	153 399 460	154 207 572	808	GBA 84GG	21	0	0.93	0	*51.28 (8.28, 2094.86)	1.13×10^{-11}	0.05	–
17	41 436 901	41 475 711	39	MAPT	423	1353	18.72	25.91	0.66 (0.58, 0.74)	1.14×10^{-11}	2.28×10^{-11}	–
2	124 466 279	124 977 213	511	2q14.3	24	3	1.06	0.06	22.58 (6.69, 119.15)	3.23×10^{-10}	1.21×10^{-10}	1.68×10^{-13}
1	153 011 655	153 081 541	70	GBA N370S	96	91	4.25	1.74	2.50 (1.87, 3.35)	1.22×10^{-9}	0.0096	–
5	133 459 416	133 928 873	469	5q31.1	19	1	0.84	0.02	44.25 (5.92, 330.90)	1.79×10^{-9}	1.87×10^{-9}	0.20
18	33 056 406	33 183 462	127	18q12.2	22	4	0.97	0.08	12.82 (4.41, 37.25)	1.31×10^{-8}	4.89×10^{-7}	1
16	28 510 492	28 822 934	312	16p11.2	421	1271	18.63	24.34	0.71 (0.63, 0.81)	4.21×10^{-8}	3.35×10^{-5}	–

Only the single most associated segment is reported per locus. In order to get a finite OR estimate, for the *GBA* 84GG segment, a count of 1 was added to all fields in the contingency table (marked with *).

pattern as seen in Figure 2B and leads to a smaller disparity in the association signals coming from SNPs and DASH segments. Upon conditioning on an inversion marker (rs1800547/C7563692) (30), there is no haplotype association signal remaining at this locus ($P = 0.037$), the SNP association signal for rs17577094 decreases ($P = 1.29 \times 10^{-5}$) and association of SNPs in LD vanishes (Supplementary Material, Fig. S3), suggesting that the inversion event acts as an excellent tag for the causal variant at this locus.

GBA

We captured a genome-wide significant association signal at the *GBA* locus [rs1630500, OR = 1.75 (1.43, 2.15), $P = 2.12 \times 10^{-8}$, Fig. 2C]. Analysis of DASH segments and comparison with the independently genotyped *GBA* mutations revealed several distinct groups of associated haplotypes. An IBD segment of ∼808 kb in length that carries the c.84dupG (84GG) mutation [Haldane-corrected OR = 51.28 (8.28, 2094.86), $P = 1.13 \times 10^{-11}$] occurs in 21 cases and no controls. Following in significance are DASH segments proximal and distal to the location of the GBA gene, one seen in 96 cases and 88 controls [OR = 2.50 (1.87, 3.35), $P = 1.22 \times 10^{-9}$] and the other seen in 193 cases and 256 controls [OR = 1.82 (1.50, 2.20), $P = 1.62 \times 10^{-9}$], both featuring p.N370S mutation carriers. Samples in the proximal segment are a subset of samples in the distal segment. A DASH segment between these two contains a haplotype shared by all p.N370S carriers; however, it contains additional cases and controls and shows a weaker association to PD [OR = 1.65 (1.37, 1.97), $P = 1.22 \times 10^{-7}$]. The shared IBD segment that carries the p.R496H mutation did not reach genome-wide significance [OR = 4.07 (2.00, 8.29), $P = 7.82 \times 10^{-5}$] owing to the small number of p.R496H carriers in our cohort. After conditioning on the *GBA* mutation carrier status (Supplementary Material, Methods), we did not observe any residual association signal at the GBA locus.

Chr2q14.3

A novel genome-wide significant rare DASH segment of ∼500 kb in length, shared by 24 cases and three controls, was observed on chr2q14.3 [OR = 22.58 (6.69, 119.15), $P = 1.21 \times 10^{-10}$, Fig. 2D]. We trained a decision tree classifier to discriminate haplotype carriers from non-carriers (Supplementary Material, Methods). When we applied the four-SNP haplotype model (rs1921815, rs6742628, rs11679636 and rs17010715, highlighted in red in Supplementary Material, Table S3) to a secondary cohort of 306 Ashkenazi PD cases and 2583 controls, we found the haplotype present in 13.1 (95% CI 9.2, 16.9) cases and 0 controls, which corresponds to the replication Fisher's exact $P = 1.68 \times 10^{-13}$ (1.51×10^{-9}, 1.77×10^{-17}). The only annotated gene within 1 Mb on either side of the DASH segment is contactin-associated protein-like 5 *CNTNAP5*, a member of the neurexin family of cell adhesion molecules and receptors, expressed in vertebrate neurons. *CNTNAP5* has been previously implicated in neurodevelopmental phenotypes (31); however, its functional relatedness to PD is not apparent. We sequenced the first 10 exons of *CNTNAP5* that were within the boundaries of the haplotype and observed an SNP in the 5′ UTR of the *CNTNAP5* gene at position chr2:124499 401 (rs77182849), present in all five PD cases that were sequenced and an intronic SNP at chr2:124 920 755,

Table 2. Imputed SNPs associated with PD, with EMMAX *P*-value of $<3.0 \times 10^{-6}$ either unconditioned and conditioned on *LRRK2* and *GBA* mutation carrier status

SNP	Chromosome	Position	Gene/locus	Alleles	MAF (%) PD	MAF (%) controls	OR (95% CI)	*P*-value (unconditional)	*P*-value (conditional)	*P*-value (replicated)
rs1442190	12	39 651 907	LRRK2	T/C	9.16	2.64	3.72 (2.98, 4.64)	1.53×10^{-27}	0.184	–
rs10506095	12	32 691 394	LRRK2	A/G	7.88	2.60	3.20 (2.54, 4.02)	1.36×10^{-20}	0.499	–
rs17577094*	17	41 543 275	MAPT	C/T	18.50	26.25	0.64 (0.56, 0.72)	8.33×10^{-9}	4.51×10^{-10}	–
rs1630500	1	153 121 679	GBA	A/G	7.70	4.54	1.75 (1.43, 2.15)	2.12×10^{-8}	0.036	–
rs10737170	1	154 330 504	GBA	C/A	19.25	13.83	1.49 (1.30, 1.69)	8.62×10^{-8}	0.725	–
rs7185232	16	28 840 969	16p11.2	G/A	20.27	26.45	0.71 (0.63, 0.80)	1.38×10^{-7}	1.98×10^{-5}	0.27
rs823114	1	203 986 155	PARK16	G/A	34.82	41.54	0.75 (0.68, 0.83)	2.76×10^{-6}	0.000683	–
rs1813134*	4	151 450 821	4q31.3	A/G	35.13	29.68	1.28 (1.16, 1.43)	5.69×10^{-6}	4.33×10^{-7}	0.64
rs6739381*	2	174 249 536	2q31.1	T/C	11.02	8.16	1.39 (1.18, 1.64)	3.58×10^{-5}	9.58×10^{-7}	0.012

Only the single strongest associated SNP at each peak has been reported. For the three SNPs that had stronger association in the conditional analysis (marked with *), ORs and 95% CI are based on Mantel–Haenszel pooled odds ratios; otherwise they were single stratum sample ORs.

seen in only two out of the five sequenced samples (Supplementary Material, Table S4). We did not identify any coding variants. According to the 1000 Genomes Project, MAF of rs77182849 is 5% across all ethnic groups and 11% in samples of European ancestry, making it unlikely to be a rare risk allele for PD.

Additional association signals

We observed an SNP hit on chr1q32.1, in the *PARK16* locus [rs823114; OR = 0.75 (0.68, 0.83), $P = 2.76 \times 10^{-6}$], with an effect consistent with the one seen in non-Jewish Europeans (rs823128, OR = 0.66) (7), and recently in two cohorts of Ashkenazi PD patients and controls (28,32). The best DASH segment at this locus captures a rare haplotype that occurs in 11 cases and 2 controls [OR = 12.77 (2.83, 57.64), $P = 7.69 \times 10^{-5}$, Supplementary Material, Fig. S4].

Among the three common loci that were nearing significance—chr2q31.1, chr4q31.3 and chr16p11.2 (Table 2)—that we attempted to replicate in the secondary cohort, we observed moderate replication signal at chr2q31.1 ($P = 0.012$), in the intergenic region between cell division cycle-associated seven gene *CDCA7* and Sp3 transcription factor gene *SP3*.

Two additional genome-wide significant rare DASH segments on chr5q31.1 [OR = 44.25 (5.92, 330.90), $P = 1.79 \times 10^{-9}$] and chr18q12.2 [OR = 12.82 (4.41, 37.25), $P = 1.31 \times 10^{-8}$] did not replicate in the secondary cohort (Supplementary Material, Tables S5 and S6).

Homozygosity mapping, typically informative in studies of founder populations, in our study did not implicate any regions at the genome-wide significance level.

Clinical and demographics variables

Information about sex, age-at-onset (AAO, defined as the age of first symptoms) and whether the case was sporadic or familial (defined as having a first degree relative with PD) is summarized in Supplementary Material, Table S2. Columbia had a higher fraction of male PD cases (67.4% compared with 56.1 and 62.6%, $\chi^2 P = 4.3 \times 10^{-5}$), and Beth Israel had a higher fraction of familial cases (38.6% compared with 17.8 and 16.8%, $\chi^2 P = 2.7 \times 10^{-11}$). There was no significant difference in the AAO between the three centers (logrank $P = 0.55$, Supplementary Material, Fig. S6A). The differences in family history are likely because of differences in sampling strategy between the

three centers: predilection towards early onset PD cases at Columbia, towards familial cases at Beth Israel and consecutive recruitment at Tel Aviv Sourasky.

The average p.G2019S carrier frequency was 14.9% (Supplementary Material, Table S2). Although some heterogeneity was observed among centers (19.8, 13.3 and 14.0% in Beth Israel, Columbia and Tel Aviv Sourasky cohorts, respectively), it was not statistically significant (Fisher's exact $P = 0.085$). Observed frequency of p.G2019S is similar to what was previously published (14.8%), although we should note that there is a sample overlap between this and the previous study. The frequency of GBA mutation carriers was 17.6% in the combined sample, with significant heterogeneity between centers (12.0, 16.7 and 20.2%, $P = 0.021$), and not significantly different from the previously published frequency range of 13.7–17.9% (12,18), although again some samples overlapped between this and the earlier study.

LRRK2 and *GBA* mutation carriers have an earlier AAO than non-carriers (mean $\pm$ SD, 56.4 ± 11.7 and 54.0 ± 11.3%, respectively, compared with 60.6 ± 11.6%, logrank $P = 1.2 \times 10^{-5}$, Supplementary Material, Fig. S6B). Familial cases are enriched in LRRK2 mutation carriers (29.1% of familial cases carry p.G2019S, $P = 7.15 \times 10^{-9}$), comparable to the previously published estimate of 26%, but are not enriched in *GBA* carriers. We did not observe evidence in support of a gender effect for either *LRRK2* ($P = 0.098$) or *GBA* ($P = 0.70$), and there was no difference in AAO between males and females ($P = 0.97$), nor between familial and sporadic cases ($P = 0.27$, Kaplan–Meier curves are shown in Supplementary Material, Fig. S6).

To investigate genetic factors that may influence clinical and demographic variables, we performed case-only genome-wide associations to family history, gender and AAO using Fisher's exact and logrank tests. We did not observe any site that increases gender-specific risk (Supplementary Material, Fig. S7). The only locus strongly associated with familial forms of PD was *LRRK2* (Supplementary Material, Fig. S8). One locus on chr17q24.2 captured by DASH segments in three samples had a genome-wide significant effect on AAO (chr17:62 661 453–63 110 659; mean $\pm$ SD 92.7 ± 1.5, $P < 1.0 \times 10^{-8}$, Supplementary Material, Figs S9C and S10). It contains two genes, proteasome 26S subunit (*PSMD12*) and cytoplasmic phosphatidylinositol transfer protein, cytoplasmic (*PITPNC1*).
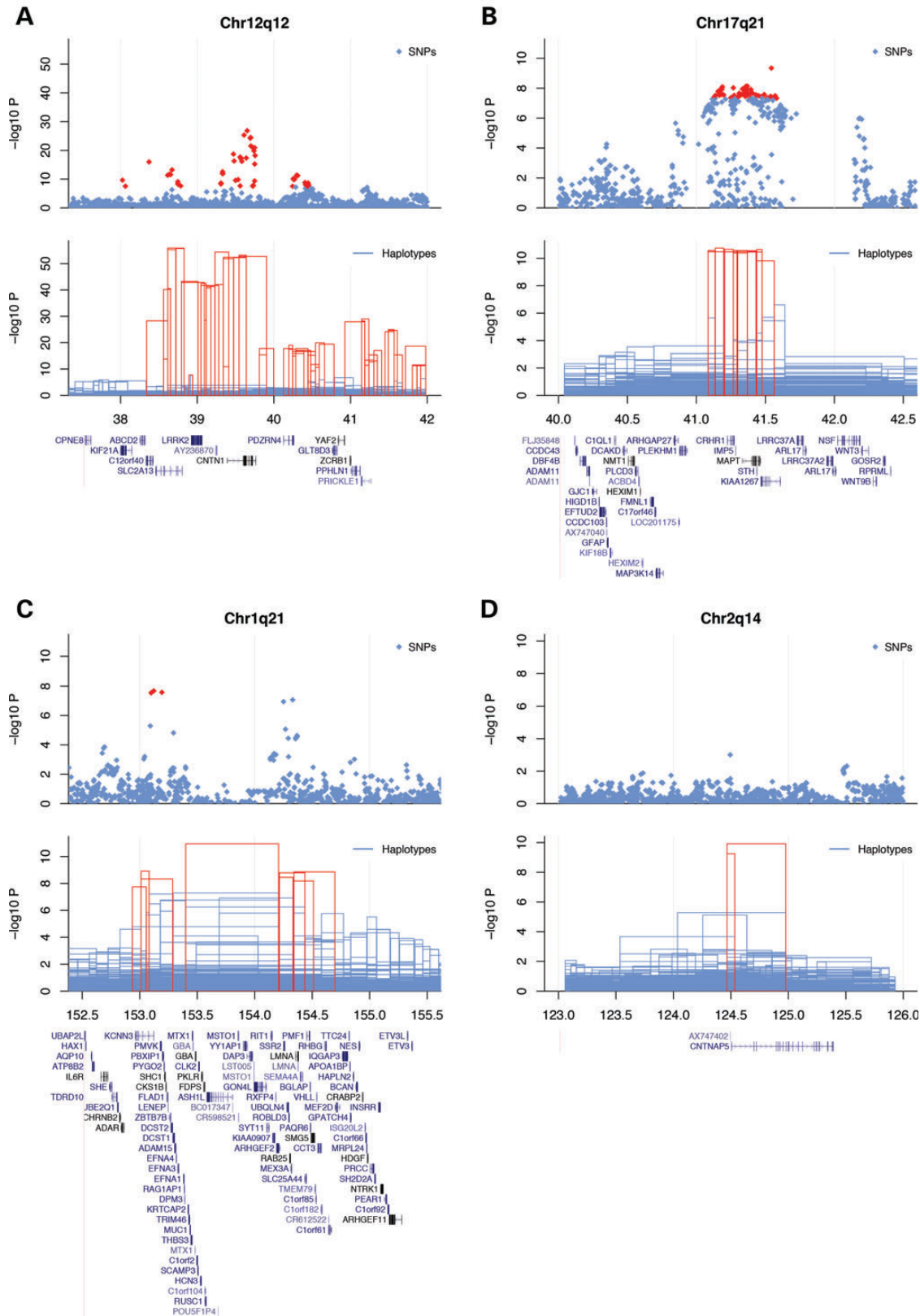
Figure 2. Association of imputed SNPs and DASH segments at the four genome-wide significant loci either also associated in other studies, or replicated as part of this study: (**A**) chr12q12 (with the associated *LRRK2* gene), (**B**) chr17q21 (*MAPT*), (**C**) chr1q21 (*GBA*) and (**D**) chr2q14 (*CNTNAP5*).

Replication of previously published loci

In order to evaluate the overlap between genetic risk factors contributing to PD in Ashkenazi Jews and other populations, we tested all hits with *P*-values of $<10^{-5}$ reported in previously published GWASs (3,5–7,33,34). For the first time in the Ashkenazim, we observed an association signal at the *SNCA* locus [rs356220; OR = 1.22 (1.11, 1.35), $P = 5.6 \times 10^{-5}$], with an odds ratio comparable to that in Europeans (1.23–1.30). At the significance level Bonferroni-corrected by the number of SNPs that we attempted to replicate ($\alpha = 0.05/66 = 0.00076$), we were able to capture association signals at the *MCCC1/LAMP3* [rs9290751; OR = 0.80 (0.71, 0.91), $P = 3.55 \times 10^{-4}$] and *PARK16* loci [rs947211; OR = 0.77 (0.69, 0.87), $P = 6.68 \times 10^{-4}$]. At the nominal significance level, we were able to capture four additional signals (*GAK/DGKQ*, *RIT2/SYT4*, *ITGA8* and *CCDC62/HIP1R*, Supplementary Material, Table S7).

In addition, association of IBD segments was able to capture signals at the *GAK/DGKQ* and *BST1* loci (Supplementary Material, Table S8). At the *BST1* locus—similar to what we observed at the *GBA* locus—there was more than one distinct group of haplotypes carrying the variants that increase the risk of PD: chr4:14 997 475–15 727 723 [OR = 23.2 (3.0, 181.4), $P = 4.98 \times 10^{-5}$, 0.44% in PD, 0.019% in controls] and chr4:15 215 529–15 350 349 [OR = 2.2 (1.3, 3.6), $P = 0.0018$, 1.4% in PD, 0.65% in controls]. However, when we corrected the nominal *P*-values by the number of DASH segments at the *GAK/DGKQ* and *BST1* loci and the number of loci tested genome wide, these two hits were not significant. Nonetheless, carriers of the *BST1* segments, and in particular the high-risk haplotype, are interesting candidates for sequencing follow-ups.

Using both association strategies, we were able to replicate known signals at three loci (*SNCA*, *MCCC1/LAMP3* and *PARK16*), and we have seen nominal significance at five additional loci (*GAK/DGKQ*, *RIT2/SYT4*, *ITGA8*, *CCDC62/HIP1R* and *BST1*). Together with the loci genome-wide significant in this study, we have in all been able to capture association signals at 11 out of 33 currently known PD loci. Top SNPs at the three genome-wide significant loci (*LRRK2*, *GBA* and *MAPT*) taken together explain 8.1% of phenotypic variance (Nagelkerke's pseudo R^2 from logistic regression), and top SNPs at seven loci with replication *P*-value of <0.05 (*PARK16, MCCC1/LAMP3, GAK/DGKQ, SNCA, ITGA8, CCDC62/HIP1R* and *RIT2/SYT4*) explain additional 2.0%, for a total of 10.1%. Genome-wide significant and replicated DASH segments (*LRRK2*, *GBA*, *MAPT* and *CNTNAP5*) explain 15.6% of the phenotypic variance and, together with the segments that have corrected *P*-values of <0.05, explain a total of 16.5% of phenotypic variance. As expected, this is significantly higher than variance explained in outbred European populations, estimated to be $\sim$6–7% (3).

DISCUSSION

We have identified genome-wide significant haplotypes at *LRRK2*, *GBA*, *MAPT* and chr2q14.3 loci, ranging in size from 40 to 800 kb. The first three of these were previously known and were also captured using imputed SNPs as part of our study; however, haplotype associations consistently refined the SNP association signals. In addition, we replicated SNPs at *SNCA*, *MCCC1/LAMP3* and *PARK16* loci. An earlier GWAS conducted on partially overlapping set of samples (268 cases and 178 controls ascertained at Columbia University Medical Center) replicated association signals at the *MAPT* ($P = 7.16 \times 10^{-5}$), *LRRK2* ($P = 1.56 \times 10^{-4}$), *PARK16* ($P = 6.12 \times 10^{-4}$) and *GBA* loci (seven SNPs with nominally significant *P*-values); however, owing to the small sample size, it was underpowered to reach genome-wide significance at any locus (28). Interestingly, we were not able to replicate the exact previously reported SNPs at the *LRRK2* and *GBA* loci, although *LRRK2* and *GBA* were genome-wide significant in our study. This points to the differences in LD structure between the Ashkenazim and non-Jewish Europeans at these two loci.

A point worth making is that we did not observe a significant p.N370S DASH segment that overlaps the *GBA* gene, although two significant p.N370S DASH segments were found proximal and distal to the *GBA* gene. This may indicate a limitation of our method to resolve haplotypes in a complex genetic region, or a possibility that there might be additional mutations contributing to the risk of PD, linked with p.N370S, located within the flanking DASH segments. We observed a secondary distal linked SNP signal [rs10737170; OR = 1.49 (1.30, 1.69); $P = 8.62 \times 10^{-8}$] that corresponds to the distal DASH segment; however, after conditioning on the *GBA* mutation carrier status, there was no residual signal at the *GBA* locus. On the other hand, a recent study reported an association hit in Europeans [OR = 1.67 (1.50, 1.84), $P = 5.7 \times 10^{-9}$] related to the nearby *SYT11* gene (5), close to *GBA* but apparently independent of the mutations in *GBA* in Europeans.

Assembling a sizeable cohort of samples from the Ashkenazi Jewish founder population came at the expense of genotyping platform heterogeneity, which had to be carefully addressed. To provide a uniform, dense set of markers, we created a custom Ashkenazi reference panel (22) and imputed the missing SNPs. Pooling samples and imputation of missing SNPs, as opposed to a more straightforward meta-analysis, allowed consistent quality control across all samples and enabled us to phase SNPs and call shared IBD segments. Some of the control samples used in this study were diagnosed with Crohn disease (CD); nonetheless, CD and PDs are clinically, pathologically and etiologically distinct, and for the variants reported in this study, we did not observe significant differences in allele frequencies between healthy controls and Crohn cases (Supplementary Material, Table S9 and S10). Although variants at *LRRK2* and HLA loci have been associated with both PD and CD, separate analysis of PD and CD in our cohort showed no overlap in terms of risk loci, which together with the increase in power coming from having a larger number of control samples (Supplementary Material, Fig. S5) justifies the use of CD cases as non-Parkinson controls. In addition, we concurrently performed a GWAS of CD in Ashkenazi Jews (22), and this provided an opportunity to verify the association signals. In contrast to CD, which has population prevalence two to four times higher in the Ashkenazi Jews compared with non-Jewish Europeans, and several rare disorders (such as Tay-Sachs and Gaucher diseases) that appear at elevated frequencies among the Ashkenazim, limited surveys of PD prevalence remain inconclusive. One study found no significant difference between Ashkenazi Jews and non-Jewish Europeans (35). A more

recent study indicated that PD prevalence in Israel is elevated compared with Europe and the USA (36), although it did not specifically focus on the Ashkenazim.

Our results highlight the usefulness of our haplotype association method. Until whole-genome sequencing in large cohorts becomes widespread, associated DASH segments can be used to guide targeted region resequencing efforts or to prioritize samples for whole-genome sequencing. Although the use of DASH is not limited to founder populations, pervasive IBD sharing in these populations facilitates haplotype mapping, as demonstrated here. DASH was able to dissect the signal at the *GBA* locus into distinct groups of haplotypes that carry distinct mutations and provide initial evidence that multiple causal variants might also be present at the *BST1* locus. We were also able to identify and replicate a novel risk haplotype on chr2q14.3 occurring at $\sim$1% frequency in PD cases. This haplotype contains a single gene expressed in the brain, *CNTNAP5*, previously not implicated in PD. As rare alleles tend to be population-specific, it remains to be seen whether the risk variant that arose on the backbone of this haplotype, or possibly different variants in the *CNTNAP5* gene, are present in other populations. Taken together, our results reaffirm the complex genetic background of PD and emphasize the need for studying both common and rare variants in PD, with the goal of understanding the molecular processes these variants affect.

MATERIALS AND METHODS

The study was approved by the Institutional Review Boards at all participating institutions, including Beth Israel Medical Center, Columbia University, Tel Aviv Sourasky Medical Center, Mount Sinai School of Medicine, Albert Einstein College of Medicine, The Hebrew University of Jerusalem, Yale University, Children's Hospital of Philadelphia and North Shore University Hospital-Long Island Jewish Medical Center. All patients provided written informed consent (in English or Hebrew) for the collection of samples and subsequent analysis.

Primary cohort

A total of 1248 Ashkenazi PD cases have been ascertained at Tel Aviv Sourasky Medical Center (Tel Aviv, Israel), Center for Movement Disorders at Columbia University and Department of Neurology at Beth Israel Medical Center (New York, USA). Subjects at Tel Aviv Sourasky Medical Center ($n = 597$) were consecutively recruited individuals of Ashkenazi Jewish origin; details of the cohort recruitment, diagnostic criteria and interview procedures are as described in (12,14). The cohort of cases recruited at the Center for Movement Disorders at Columbia University was created by combining participants from two studies of PD: genetic Epidemiology of PD ($n = 168$, enriched for cases with AAO < 50 y/o) and the Ashkenazi Jewish PD study ($n = 184$, selected based on self-reported ancestry). Details of the ascertainment procedures, evaluation criteria and structured interviews are provided in (28,37). Ashkenazi PD cases recruited from the Department of Neurology at Beth Israel ($n = 299$) have been evaluated as previously described in (15). PD cases that carried mutations in the *GBA* gene did not have symptoms of Gaucher Disease. To account for

differences in procedures employed to verify subjects' Ashkenazi Jewish ancestry at different centers, as part of this study ancestry of all subjects was uniformly verified using principle component analysis. Our recent results justify the use of a single population model for Ashkenazi Jews from Israel and the USA (38) and show that after accounting for European admixture, there is no evidence for significant population stratification (39).

The set of 2865 Ashkenazi controls contain both neurologically healthy individuals ($n = 422$) from Tel Aviv Sourasky, Columbia and Beth Israel Medical Centers and pooled controls who have not undergone neurological examination ($n = 2443$) from The Hebrew University of Jerusalem, Albert Einstein College of Medicine, Mt. Sinai School of Medicine, Yale University and Children's Hospital of Philadelphia. Pooled controls consist of apparently healthy individuals (controls in their respective studies, $n = 1474$) as well as individuals diagnosed with CD ($n = 969$). A subgroup of the apparently healthy pooled controls ($n = 646$) was recruited for a study of healthy aging (40). Based on PD population prevalence of $\sim$1% (41), $\sim$24 pooled control samples can be expected to have or will have developed PD. Details of the sample cohorts are summarized in Supplementary Material, Table S1.

Samples were genotyped on Illumina Human 300 k, 500 k, 660 k and Omni1M and Affymetrix GeneChip Human Mapping 500K and Genome-Wide Human SNP 6.0 arrays (Supplementary Material, Table S1). Details about quality control, ancestry filtering, etc., are provided in Supplementary Material, Methods.

Replication cohort

A selected set of SNPs was genotyped in the replicate cohort of 306 Ashkenazi PD cases from Tel Aviv Sourasky Medical Center using ABI TaqMan assays (Applied Biosystems, Foster City, CA, USA). Frequencies of *LRRK2/GBA* mutations in the replicate PD cohort were comparable to the primary sample. Replication controls consisted of 2583 Ashkenazi schizophrenia cases and controls from the North Shore-Long Island Jewish Health System typed on Illumina HumanOmni1-Quad Bead-Chip, imputed using the same procedure and passing the same quality control and ancestry filtering as the discovery cohort.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors thank the Parkinson patients, who have participated in this study. We also thank Ziv Gan-Or and Liron Rozenkrantz for technical assistance.

Conflict of Interest statement. None declared.

FUNDING

REFERENCES

1. Nussbaum, R.L. and Ellis, C.E. (2003) Alzheimer's disease and Parkinson's disease. *N. Engl. J. Med.*, **348**, 1356–1364.
2. Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simon-Sanchez, J., Mittag, F., Buchel, F., Sharma, M., Gibbs, J.R. *et al.* (2012) Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.*, **21**, 4996–5009.
3. Do, C.B., Tung, J.Y., Dorfman, E., Kiefer, A.K., Drabant, E.M., Francke, U., Mountain, J.L., Goldman, S.M., Tanner, C.M., Langston, J.W. *et al.* (2011) Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.*, **7**, e1002141.
4. Lesage, S. and Brice, A. (2009) Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum. Mol. Genet.*, **18**, R48–R59.
5. Nalls, M.A., Plagnol, V., Hernandez, D.G., Sharma, M., Sheerin, U.M., Saad, M., Simon-Sanchez, J., Schulte, C., Lesage, S., Sveinbjornsdottir, S. *et al.* (2011) Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*, **377**, 641–649.
6. Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A. *et al.* (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.*, **41**, 1303–1307.
7. Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G. *et al.* (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.*, **41**, 1308–1312.
8. Motulsky, A.G. (1995) Jewish diseases and origins. *Nat. Genet.*, **9**, 99–101.
9. Ostrer, H. (2001) A genetic profile of contemporary Jewish populations. *Nat. Rev. Genet.*, **2**, 891–898.
10. Mulle, J.G., Dodd, A.F., McGrath, J.A., Wolyniec, P.S., Mitchell, A.A., Shetty, A.C., Sobreira, N.L., Valle, D., Rudd, M.K., Satten, G. *et al.* (2010) Microdeletions of 3q29 confer high risk for schizophrenia. *Am. J. Hum. Genet.*, **87**, 229–236.
11. Shifman, S., Bronstein, M., Sternfeld, M., Pisante-Shalom, A., Lev-Lehman, E., Weizman, A., Reznik, I., Spivak, B., Grisaru, N., Karp, L. *et al.* (2002) A highly significant association between a COMT haplotype and schizophrenia. *Am. J. Hum. Genet.*, **71**, 1296–1302.
12. Gan-Or, Z., Giladi, N., Rozovski, U., Shifrin, C., Rosner, S., Gurevich, T., Bar-Shira, A. and Orr-Urtreger, A. (2008) Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. *Neurology*, **70**, 2277–2283.
13. Clark, L.N., Wang, Y., Karlins, E., Saito, L., Mejia-Santana, H., Harris, J., Louis, E.D., Cote, L.J., Andrews, H., Fahn, S. *et al.* (2006) Frequency of LRRK2 mutations in early- and late-onset Parkinson disease. *Neurology*, **67**, 1786–1791.
14. Orr-Urtreger, A., Shifrin, C., Rozovski, U., Rosner, S., Bercovich, D., Gurevich, T., Yagev-More, H., Bar-Shira, A. and Giladi, N. (2007) The LRRK2 G2019S mutation in Ashkenazi Jews with Parkinson disease: is there a gender effect? *Neurology*, **69**, 1595–1602.
15. Ozelius, L.J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A.L., Klein, C., Henick, B., Hailpern, S.M. *et al.* (2006)
16. Aharon-Peretz, J., Rosenbaum, H. and Gershoni-Baruch, R. (2004) Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. *N. Engl. J. Med.*, **351**, 1972–1977.
17. Neudorfer, O., Giladi, N., Elstein, D., Abrahamov, A., Turezkite, T., Aghai, E., Reches, A., Bembi, B. and Zimran, A. (1996) Occurrence of Parkinson's syndrome in type I Gaucher disease. *Q. J. Med.*, **89**, 691–694.
18. Clark, L.N., Ross, B.M., Wang, Y., Mejia-Santana, H., Harris, J., Louis, E.D., Cote, L.J., Andrews, H., Fahn, S., Waters, C. *et al.* (2007) Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. *Neurology*, **69**, 1270–1277.
19. Horowitz, M., Pasmanik-Chor, M., Borochowitz, Z., Falik-Zaccai, T., Heldmann, K., Carmi, R., Parvari, R., Beit-Or, H., Goldman, B., Peleg, L. *et al.* (1998) Prevalence of glucocerebrosidase mutations in the Israeli Ashkenazi Jewish population. *Hum. Mutat.*, **12**, 240–244.
20. Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., Burns, E. *et al.* (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am. J. Hum. Genet.*, **86**, 850–859.
21. Gusev, A., Kenny, E.E., Lowe, J.K., Salit, J., Saxena, R., Kathiresan, S., Altshuler, D.M., Friedman, J.M., Breslow, J.L. and Pe'er, I. (2011) DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. *Am. J. Hum. Genet.*, **88**, 706–717.
22. Kenny, E.E., Pe'er, I., Karban, A., Ozelius, L., Mitchell, A.A., Ng, S.M., Erazo, M., Ostrer, H., Abraham, C., Abreu, M.T. *et al.* (2012) A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility Loci. *PLoS Genet.*, **8**, e1002559.
23. Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M. and Pe'er, I. (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res.*, **19**, 318–326.
24. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C. and Eskin, E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.*, **42**, 348–354.
25. Kenny, E.E., Kim, M., Gusev, A., Lowe, J.K., Salit, J., Smith, J.G., Kovvali, S., Kang, H.M., Newton-Cheh, C., Daly, M.J. *et al.* (2011) Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. *Hum. Mol. Genet.*, **20**, 827–839.
26. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D. *et al.* (2008) Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.*, **83**, 132–135; author reply 135–139.
27. Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G. *et al.* (2005) A common inversion under selection in Europeans. *Nat. Genet.*, **37**, 129–137.
28. Liu, X., Cheng, R., Verbitsky, M., Kisselev, S., Browne, A., Mejia-Sanatana, H., Louis, E.D., Cote, L.J., Andrews, H., Waters, C. *et al.* (2011) Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Med. Genet.*, **12**, 104.
29. Donnelly, M.P., Paschou, P., Grigorenko, E., Gurwitz, D., Mehdi, S.Q., Kajuna, S.L.B., Barta, C., Kungulilo, S., Karoma, N.J., Lu, R.B. *et al.* (2010) The distribution and most recent common ancestor of the 17q21 inversion in humans. *Am. J. Hum. Genet.*, **86**, 161–171.
30. Osier, M.V., Cheung, K.H., Kidd, J.R., Pakstis, A.J., Miller, P.L. and Kidd, K.K. (2002) ALFRED: an allele frequency database for anthropology. *Am. J. Phys. Anthropol.*, **119**, 77–83.
31. Pagnamenta, A.T., Bacchelli, E., de Jonge, M.V., Mirza, G., Scerri, T.S., Minopoli, F., Chiocchetti, A., Ludwig, K.U., Hoffmann, P., Paracchini, S. *et al.* (2010) Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol. Psychiatry*, **68**, 320–328.
32. Gan-Or, Z., Bar-Shira, A., Dahary, D., Mirelman, A., Kedmi, M., Gurevich, T., Giladi, N. and Orr-Urtreger, A. (2012) Association of sequence alterations in the putative promoter of rab7l1 with a reduced Parkinson disease risk. *Arch. Neurol.*, **69**, 105–110.
33. Pankratz, N., Wilk, J.B., Latourelle, J.C., DeStefano, A.L., Halter, C., Pugh, E.W., Doheny, K.F., Gusella, J.F., Nichols, W.C., Foroud, T. *et al.* (2009) Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.*, **124**, 593–605.

LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *N. Engl. J. Med.*, **354**, 424–425.

34. Spencer, C.C., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C., Band, G., Barker, R.A., Bellenguez, C., Bhatia, K., Blackburn, H. *et al.* (2011) Dissection of the genetics of Parkinson's disease identifies an additional association 5′ of SNCA and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.*, **20**, 345–353.

35. Anca, M., Paleacu, D., Shabtai, H. and Giladi, N. (2002) Cross-sectional study of the prevalence of Parkinson's disease in the Kibbutz movement in Israel. *Neuroepidemiology*, **21**, 50–55.

36. Chillag-Talmora, O., Giladi, N., Linn, S., Gurevich, T., El-Ad, B., Silverman, B., Friedman, N. and Peretz, C. (2011) Use of a refined drug tracer algorithm to estimate prevalence and incidence of Parkinson's disease in a large Israeli population. *J. Parkinson's Dis.*, **1**, 35–47.

37. Marder, K., Levy, G., Louis, E.D., Mejia-Santana, H., Cote, L., Andrews, H., Harris, J., Waters, C., Ford, B., Frucht, S. *et al.* (2003) Familial aggregation of early- and late-onset Parkinson's disease. *Ann. Neurol.*, **54**, 507–513.

38. Gusev, A., Palamara, P.F., Aponte, G., Zhuang, Z., Darvasi, A., Gregersen, P. and Pe'er, I. (2012) The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.*, **29**, 473–486.

39. Guha, S., Rosenfeld, J.A., Malhotra, A.K., Lee, A.T., Gregersen, P.K., Kane, J.M., Pe'er, I., Darvasi, A. and Lencz, T. (2012) Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.*, **13**, R2.

40. Atzmon, G., Schechter, C., Greiner, W., Davidson, D., Rennert, G. and Barzilai, N. (2004) Clinical phenotype of families with longevity. *J. Am. Geriatr. Soc.*, **52**, 274–277.

41. Dorsey, E.R., Constantinescu, R., Thompson, J.P., Biglan, K.M., Holloway, R.G., Kieburtz, K., Marshall, F.J., Ravina, B.M., Schifitto, G., Siderowf, A. *et al.* (2007) Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, **68**, 384–386.