Identification of a Novel Risk Locus for Progressive Supranuclear Palsy by a Pooled Genomewide Scan of 500,288 Single-Nucleotide Polymorphisms

Stacey Melquist,* David W. Craig,* Matthew J. Huentelman, Richard Crook, John V. Pearson, Matt Baker, Victoria L. Zismann, Jennifer Gass, Jennifer Adamson, Szabolcs Szelinger, Jason Corneveaux, Ashley Cannon, Keith D. Coon, Sarah Lincoln, Charles Adler, Paul Tuite, Donald B. Calne, Eileen H. Bigio, Ryan J. Uitti, Zbigniew K. Wszolek, Lawrence I. Golbe, Richard J. Caselli, Neill Graff-Radford, Irene Litvan, Matthew J. Farrer, Dennis W. Dickson, Mike Hutton, and Dietrich A. Stephan

To date, only the H1 *MAPT* haplotype has been consistently associated with risk of developing the neurodegenerative disease progressive supranuclear palsy (PSP). We hypothesized that additional genetic loci may be involved in conferring risk of PSP that could be identified through a pooling-based genomewide association study of >500,000 SNPs. Candidate SNPs with large differences in allelic frequency were identified by ranking all SNPs by their probe-intensity difference between cohorts. The *MAPT* H1 haplotype was strongly detected by this methodology, as was a second major locus on chromosome 11p12-p11 that showed evidence of association at allelic ($P < .001$), genotypic ($P < .001$), and haplotypic ($P < .001$) levels and was narrowed to a single haplotype block containing the DNA damage-binding protein 2 (*DDB2*) and lysosomal acid phosphatase 2 (*ACP2*) genes. Since DNA damage and lysosomal dysfunction have been implicated in aging and neurodegenerative processes, both genes are viable candidates for conferring risk of disease.

Progressive supranuclear palsy (PSP [MIM 601104]) is the second-most-common form of parkinsonism, with a population prevalence rate of 6–6.4 per 100,000.[1,2] Clinical features include vertical-gaze palsy and postural instability.[3,4] PSP is characterized neuropathologically by neuronal and glial inclusions composed of aggregated microtubule associated protein tau (MAPT) in the basal ganglia and brain stem.[5,6] Mutations in the *MAPT* (MIM 157140) gene have been identified in patients with a clinical presentation of PSP.[7–14] A recent report also described linkage to chromosome 1q31.1 in a family with autosomal dominant PSP.[15] However, only the *MAPT* locus has been consistently associated with increased risk for idiopathic PSP.[16–20] The *MAPT* locus exists as two major haplotype groups, termed "H1" and "H2"[16] in European populations, with the H2 haplotype defined by >100 SNPs that are inherited in strong linkage disequilibrium (LD) with each other, reflecting the total absence of H1-H2 recombination.[21] Inheritance of two copies of the H1 haplotype (H1/H1) is a major genetic risk factor for PSP.[16] A large collection of pathologically confirmed PSP samples was used recently to fine map PSP risk on H1 chromosomes in PSP cases and controls. [22,23] PSP risk was associated with an extended subhaplotype, and narrowing the region for PSP risk to a 22-kb region in intron 0 of *MAPT* was accomplished by examining younger patients with, presumably, a larger genetic component to their disease.[22,23] The most likely explanation of the association with the *MAPT* H1 haplotype and PSP is that variants in the H1 (and H2) haplotypes confer risk of (protect against) disease by altering expression at the locus, with the risky H1 haplotypes expressing higher levels of *MAPT*.[22–26]

Calculations of population-attributable risk suggest that only ~68% of the risk of PSP can be accounted for by the *MAPT* H1 haplotype, suggesting there may be additional risk genes involved in PSP. We hypothesized that additional genetic loci involved in conferring risk of PSP could be identified through genomewide association (GWA) methods. The cost of performing an association study that involved individual genotyping of thousands of SNPs for a series this size was prohibitive, so, instead, we used a pooled-DNA approach to identify additional risk factors. Whereas a pooling-based genomewide scan of thousands of SNPs has been proposed in principle, in large part, these studies have not been used for the discovery of genes pre-

From the Departments of Neuroscience (S.M.; R.C.; M.B.; J.G.; J.A.; A.C.; S.L.; M.J.F.; D.W.D.; M.H.) and Neurology (R.J.U.; Z.K.W.; N.G.-R.), Mayo Clinic College of Medicine, Jacksonville, FL; Neurogenomics Division, Translational Genomics Research Institute (TGen), Phoenix (D.W.C.; M.J.H.; J.V.P.; V.L.Z.; S.S.; J.C.; K.D.C.; D.A.S.); Department of Neurology, Mayo Clinic, Scottsdale, AZ (C.A.; R.J.C.); Department of Neurology, University of Minnesota, Minneapolis (P.T.); Pacific Parking's Research Centre, University of British Columbia, Vancouver (D.B.C.); Department of Neuropathy, Northwestern Feinberg School of Medicine, Chicago (E.H.B.); Department of Neurology, University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School, New Brunswick, (L.G.B.): and Department of Neurology, University of Louisville School of Medicine, Louisville (I.L.)

Received September 4, 2006; accepted for publication January 12, 2007; electronically published March 8, 2007.

Address for correspondence and reprints: Dr. Mike Hutton, Department of Neuroscience, Mayo Clinic College of Medicine, 4500 San Pablo Road, Birdsall 2, Jacksonville, FL 32224. E-mail: hutton.michael@mayo.edu

* These two authors contributed equally to this work.

Am. J. Hum. Genet. 2007;80:769–778. 0002-9297/2007/8004-0018$15.00
DOI: 10.1086/513320

Table 1. Predicted Allelic Frequencies for the Top 1,000 SNPs

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

disposing to complex diseases,[27,28] likely because of technical concerns or lack of technology and analysis tools.

The patients used in the initial pooling study, the "original" series, were largely derived from pathologically confirmed subjects collected by the PSP Society and sent to D.W.D. for brain autopsy. As described elsewhere, the patient samples in this brain bank were donated from the United States and Canada.[29] The patient series is similar to the one that we employed in previous studies to fine map the H1 genetic risk,[22] with 288 subjects with a primary pathological diagnosis of PSP used to create the pool of PSP-affected patients. A total of 344 age- and sex-matched cognitively normal control individuals were obtained through the Normal and Pathological Aging protocol at the Mayo Clinic (Scottsdale),[30,31] to create the pool of control individuals. All patient and control individuals were white from the United States and Canada, and institutional review board (IRB)–approved protocols were used in the collection of all samples.

Replicate pools of patients with PSP and control individuals were created as described elsewhere.[32] Samples were genotyped on 20 replicate Affymetrix 500K arrays and 20 Affymetrix 100K, in accordance with the Affymetrix protocols, whereby each of the five replicate pools was genotyped on two replicate arrays. This design therefore yielded probe-intensity data for both platforms on 10 replicate arrays per cohort. Data were analyzed using GenePool software (TGen Bioinformatics Research Unit).[32] In brief, probe-intensity data were directly read from cell-intensity (CEL) files, and relative allele signal (RAS) values were calculated for each quartet. These values yield independent measures of different hybridization events and are consequently treated as individual data points. We

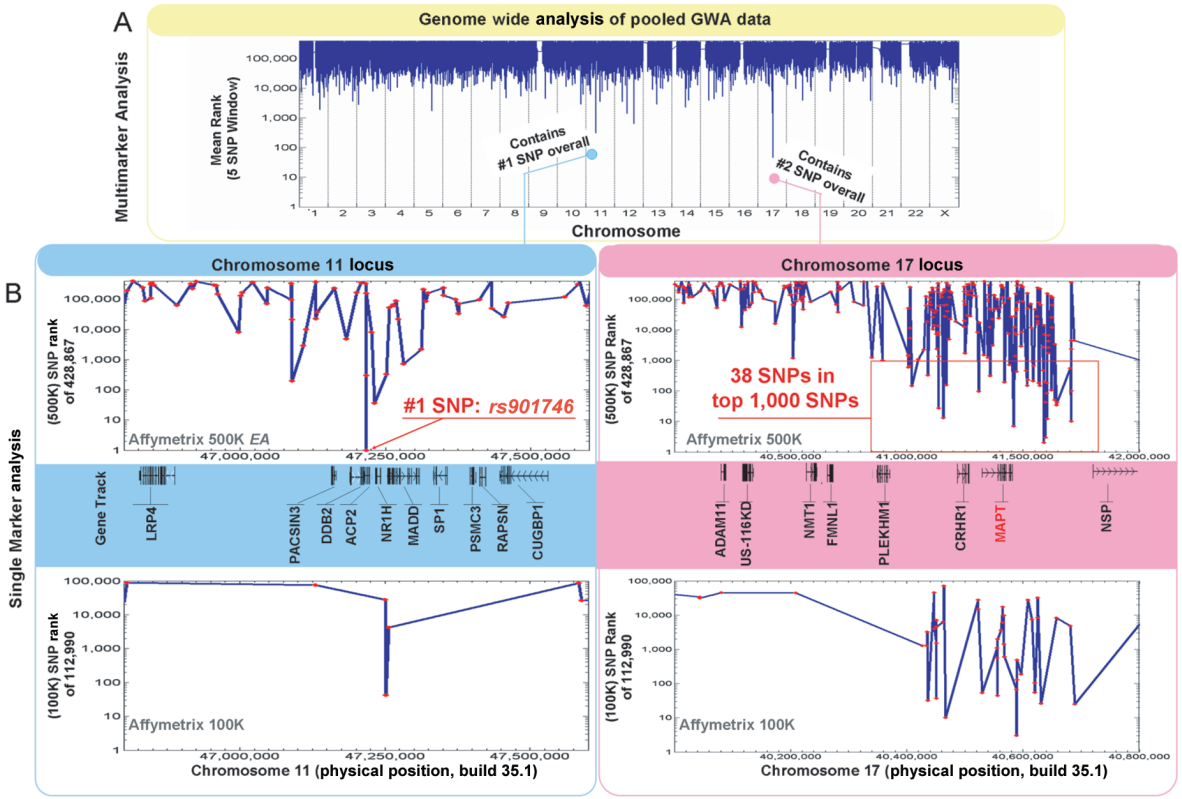

Figure 1. Two loci showing strong support for association by pooled analysis. *A,* Genomewide plot of the mean rank of five consecutive SNPs, calculated to identify clusters of high-ranking SNPs. The single best region was on chromosome 17, neighboring *MAPT,* and the second best region was on chromosome 11p12. Chromosome 11p12 also harbored the SNP that ranked #1 overall by single-marker statistics. *B,* Single-marker rank statistics for SNPs over the *MAPT* (*left*) and *DDB2/ACP2* (*right*) loci. SNPs deemed less reliable or showing high variability among replicates were removed, and the remaining SNPs were ranked in order from 1 (showing the greatest difference between cases and controls) to 428,867 (showing the least difference between cases and controls) with use of a silhouette-test statistic in GenePool software (TGen Bioinformatics Research Unit). Rank scores are plotted versus chromosomal position. Genes within the plotted chromosomal region are shown below. SNPs on the Affymetrix 500K platform are shown above, and SNPs on the Affymetrix 100K platform are shown below. EA = Early Access.

used a silhouette statistic to rank all genotyped SNPs,[33] because it avoids introducing unnecessary variance by averaging probe-intensity data from probes with different hybridization properties. Silhouette scores range from 1, where significant separation between data points has been achieved and cluster assignment can be made with confidence, to −1, where differences in allelic frequencies are less reliable. Poorly performing SNPs were identified by Affymetrix as unreliable in the transition to Mendel3 libraries or exhibited high variance between replicate arrays and were removed from the analysis; 428,867 SNPs remained. SNPs were ranked on the basis of silhouette score, whereby the SNP with the highest score was ranked 1 and the SNP with the lowest score was ranked 428,867, with use of Affymetrix's Mendel3 libraries for the Affymetrix 500K arrays and *Hin*dIII and *Xba*I libraries for the Affymetrix 100K arrays, then were sorted by chromosome and physical position. With this ranking, it is assumed that SNPs approaching a rank of 1 will have larger differences in allelic frequency. With each sample ranked by silhouette score, we calculated a sliding-window statistic of the mean rank for consecutively neighboring SNPs across a fixed window size. Window sizes from 2 to 31 were used.

Since the *MAPT* H1 haplotype is associated with disease with a haplotypic odds ratio (OR) of ~3–4,[16,22,23,34] it served as an internal positive control for the study. For analysis, we used the 500K data to identify chromosomal regions of interest (i.e., those with small mean-rank scores). The 100K data were then used to confirm that a region identified in the 500K analysis contained SNPs with large allelic frequency differences. The SNP with the single best statistical rank on the 500K chip was *rs901746* on chromosome 11p12, and the second-best SNP was *rs17662235,* near *MAPT.* The top 1,000 SNPs, based on individual statistical rank, are given in table 1. Multimarker statistics also identified both chromosome 11p12 and chromosome 17q21 (*MAPT*) regions with sliding windows of multiple sizes. Although we recognize that this type of statistic is biased because of genomewide LD, it allowed us to identify clusters of high-ranking SNPs that neighbor one another, which reduced the possibility of technical errors influencing the results. Shown in figure 1*A,* the *MAPT* locus, labeled as having the #2 SNP overall, showed the greatest evidence of differences between case and control pools with use of the sliding-window analysis, largely because

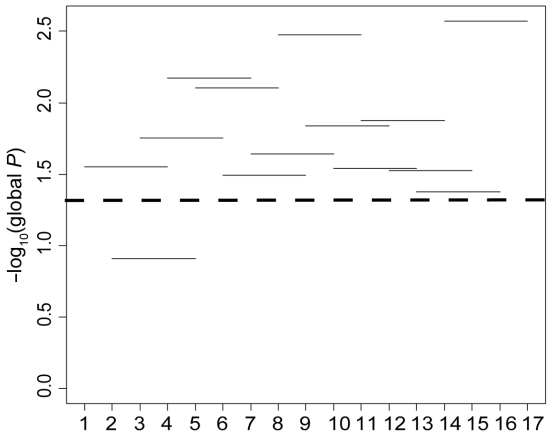

Figure 2. Haplotype sliding window–analysis results. The haplotype score–based method of Schaid et al.[40] was used to investigate evidence of association of haplotypes with case-control status. Only haplotypes with an estimated overall frequency of ⩾5% were considered for these analyses. Reported *P* values are based on asymptotic assumptions but were verified by simulating *P* values derived from 1,000 permutations of case and control labels and were found to be consistent. Global *P* values for each 4-marker haplotype are denoted as lines at the $-\log_{10}P$. Only young pathologically confirmed PSP cases (death at age <76 years) were used for the analysis. All individuals in the control group were used in all analyses, since no single SNP showed significantly different allelic frequency distribution in controls when stratified by age. Global $P = .01$ is denoted by a dashed line. SNP numbers are as noted in table 3.

of 38 SNPs within the top 1,000 SNPs overall and a total of 75 SNPs in the region with a rank score of <10,000 (fig. 1*B* and table 1). Examination of the individual SNPs with high rank scores over this locus showed SNPs that were derived from a region covering the full extent of the *MAPT* H1 haplotype, spanning nearly 1 Mb (fig. 1*B*).[36] All of the 75 SNPs with genotype-frequency data in the database resembled *MAPT* H2 variants (which differentiate between H1 and H2 *MAPT* haplotypes) rather than H1 variants (which differentiate between H1 subhaplotypes); this is because, in white populations, the SNP minor-allele frequency was ~0.2, whereas the minor allele of the SNP was absent or rare in Asian populations and African populations.[36,37] In addition, two of the SNPs with low rank

Table 2. Association Analysis of *rs901746* in Original and Replication Series

		No. (%) of							
		Alleles		Genotypes			GG versus AG and AA		
Population	*n*	A	G	AA	AG	GG	OR	95% CI	*P*
Control combined	**735**	**1,011 (75)**	**335 (25)**	**377 (56)**	**257 (38)**	**39 (6)**	…	…	…
Control original	344	438 (78)	126 (22)	166 (58)	106 (37)	10 (4)	…	…	…
Control replication	391	573 (73)	209 (27)	211 (54)	151 (39)	29 (7)	…	…	…
PSP combined	**501**	**661 (68)**	**317 (32)**	**231 (47)**	**199(41)**	**59 (12)**	**2.2**	**1.5–3.4**	**.0001**
PSP original	288	374 (68)	178 (32)	131 (47)	112 (41)	33 (12)	4.0	1.9–8.3	.0001
PSP replication	213	287 (67)	139 (33)	100 (47)	87 (41)	26 (12)	1.7	1.0–3.0	.05

Table 3. Single-Marker Analysis of Tag SNPs in the Combined Series and in Both Young and Old Patient Populations

tagID[a] (SNP), and Allele	No. (%) of Alleles in Controls (n = 532)	All Cases (n = 448)		Young[b] Cases (n = 162)		Old[b] Cases (n = 182)	
		No. (%) of Alleles	P	No. (%) of Alleles	P	No. (%) of Alleles	P
1 (*rs11039130*):			**.003**		**.02**		.25
C	600 (69)	614 (75)		224 (76)		245 (72)	
T	274 (31)	202 (25)		72 (24)		95 (28)	
2 (*rs4647709*):			.5		.57		.88
C	806 (91)	787 (90)		292 (90)		331 (91)	
T	78 (9)	85 (10)		32 (10)		31 (9)	
3 (*rs2291120*):			**.0004**		**.003**		**.006**
T	781 (92)	859 (87)		280 (86)		317 (87)	
C	67 (8)	115 (13)		44 (14)		47 (13)	
4 (*rs10742797*):			.81		.72		.98
A	591 (81)	572 (81)		212 (82)		243 (80)	
T	143 (19)	134 (19)		48 (18)		59 (20)	
5 (*rs1685404*):			.72		.97		.96
G	598 (68)	560 (67)		213 (68)		237 (68)	
C	282 (32)	274 (33)		101 (32)		111 (32)	
6 (*rs7395581*):			.03		**.02**		.07
A	378 (71)	437 (65)		157 (63)		180 (65)	
G	152 (29)	233 (35)		93 (37)		96 (35)	
7 (*rs11039138*):			**.01**		**.02**		.37
G	470 (56)	442 (62)		168 (64)		173 (59)	
A	372 (44)	268 (38)		94 (36)		121 (41)	
8 (*rs2957873*):			.22		.52		.28
A	728 (83)	679 (81)		257 (81)		281 (80)	
G	150 (17)	163 (19)		59 (19)		69 (20)	
9 (*rs4647736*):			**.03**		**.04**		.12
C	807 (91)	736 (88)		273 (88)		305 (89)	
T	75 (9)	98 (12)		39 (13)		39 (11)	
10 (*rs2013867*):			**.004**		**.006**		.02
T	657 (74)	549 (66)		206 (66)		236 (67)	
C	229 (26)	279 (34)		106 (34)		114 (33)	
11 (*rs901746*):			**<.0001**		**.003**		**.004**
A	659 (76)	570 (67)		204 (65)		242 (68)	
G	213 (24)	282 (33)		110 (35)		116 (32)	
12 (*rs1050244*):			.53		.89		.44
C	851 (97)	823 (96)		307 (97)		343 (96)	
T	29 (3)	33 (4)		11 (3)		15 (4)	
13 (*rs11039143*):			.87		.64		.52
T	830 (98)	782 (98)		293 (98)		319 (98)	
G	18 (2)	16 (2)		5 (2)		5 (2)	
14 (*rs7118396*):			.16		.27		.13
C	741 (86)	688 (84)		253 (84)		287 (83)	
T	117 (14)	132 (16)		49 (16)		59 (17)	
15 (*rs12577530*):			**.0009**		**.005**		**.04**
G	784 (88)	701 (82)		260 (82)		296 (84)	
C	106 (12)	149 (18)		58 (18)		58 (16)	
16 (*rs7114704*):			.01		.3		**.0006**
C	813 (93)	803 (96)		288 (95)		343 (98)	
T	61 (7)	35 (4)		16 (5)		7 (2)	
17 (*rs10501320*):			**<.0001**		**.003**		.16
G	609 (70)	641 (78)		230 (79)		255 (74)	
C	265 (30)	179 (22)		62 (21)		91 (26)	

NOTE.—Significant P values are shown in bold.

[a] Tag SNPs were chosen on the basis of the tagging algorithm in Haploview v3.32 software,[39] with the "Pairwise Tagging Only" option selected and the r^2 threshold set at .8.

[b] Subjects aged <76 years were classified as "young"; subjects aged ⩾76 were classified as "old."

Table 4. SNP Discovery Results from Sequencing *DDB2* and *ACP2* in 18 Subjects with PSP

	Genotype at *DDB2* 5′→3′			Genotype at *ACP2* 3′→5′					
	Intron 8	Intron 9	3′ UTR	Intron 6		Exon 5	Intron 3		Exon 1
Sample	*rs326222*	***rs901746***	*rs1050244*	*rs11039146*	*rs2242261*	*rs10838677*[a]	ss68362654[b]	*rs4752973*	*rs2167079*[c]
1	GG	GG	CT	CT	AA	AG	AG	AG	AA
2	GG	GG	CC	CC	AA	GG	GG	AA	AA
3	GG	GG	CC	CC	CC	GG	GG	GG	AA
4	GG	GG	CC	CC	AC	GG	GG	AG	AA
5	GG	GG	CC	CC	AC	GG	GG	AG	AA
6	GG	GG	CC	CC	CC	GG	GG	GG	AA
7	GG	GG	CC	CC	AA	GG	GG	AA	AA
8	GG	GG	CT	CT	AA	AG	AG	AG	AA
9	AG	AG	CC	CC	AA	GG	GG	AA	AG
10	GG	GG	CC	CC	AC	GG	GG	AG	AA
11	GG	GG	CC	CC	AC	GG	GG	AG	AA
12	GG	GG	CC	CC	AC	GG	GG	AG	AA
13	AG	AG	CC	CC	AA	GG	GG	AA	AG
14	AG	AG	CC	CC	AA	GG	GG	AA	GG
15	AG	AG	CC	CC	AA	GG	GG	AA	AG
16	AG	AG	CC	CC	AA	GG	GG	AA	AG
17	AA	AA	CC	CC	AA	GG	GG	AA	GG
18	AA	AA	CC	CC	AA	GG	GG	AA	GG

[a] Encodes synonymous change L165L.
[b] No rs number; submitted to dbSNP.
[c] Encodes nonsynonymous change R29Q.

scores (*rs12150111* and *rs807072*) were identified definitively as *MAPT* H2 variants from prior *MAPT* genomic sequencing efforts.[22]

The chromosome 11p12 region that showed the highest rank SNP by single-marker statistics and multimarker sliding-window analysis was a novel locus and therefore was examined in greater detail (fig. 1*B*). The top overall ranked SNP, *rs901746,* a SNP in intron 9 of the DNA damage-binding protein 2 (*DDB2* [MIM 600811]) gene, was chosen for follow-up in the individual samples comprising the pooled DNA. A significant increase of 10% in the G allele frequency was seen in cases versus controls ($P = .0002$) (table 2). The SNP was then genotyped in a second U.S. series to confirm the association. This "replication" sample ($n = 161$) was made up of both pathologically confirmed ($n = 97$) and clinically defined PSP case individuals ($n = 64$), as described in Rademakers et al.[22] A total of 165 age- and sex-matched cognitively normal control individuals were obtained from the Normal and Pathological Aging Protocol at the Mayo Clinic (Scottsdale).[30,31] In addition, for the *rs901746* and *rs2167079* analysis, additional pathologically confirmed cases ($n = 41$) and clinically defined PSP case indivuals ($n = 22$) were genotyped, and 252 age- and sex-matched cognitively normal control individuals collected at Mayo Clinic Jacksonville were used as a second source of controls.[22] All case and control individuals in this set were white from the United States and Canada, and IRB-approved protocols were used in the collection of all samples.

When allele frequencies at *rs901746* were examined in the replication sample set, a 6% increase in the frequency of the G allele in subjects with PSP was observed; however, because of the smaller sample size, this allele frequency difference is borderline significant ($P = .05$). When genotype distributions were examined in both PSP case-control series, the frequencies were very similar, with an increase from 4% to 12% in the GG genotype in the original population and an increase from 7% to 12% in the replication set. The allelic frequency difference in both series is explained by an apparent doubling of the GG frequency in subjects with PSP compared with controls, suggesting that risk at this locus acts in a recessive manner. We explicitly tested dominant, recessive, and additive models at this locus, and the model that best fit the data was a recessive one ($P < .0001$). The OR for harboring an *rs901746* GG genotype versus all other genotypes in the original series was 3.7 (95% CI 1.2–3.9) and was 1.7 (95% CI 1.0–3.0) for the replication series. When these individuals in both of these series were combined and analyzed, the combined

Table 5. Association Analysis of *rs2167079* in the Combined Series

	No. (%) of Alleles		
SNP and Allele	All Controls ($n = 735$)	All Patients ($n = 501$)	*P*
rs901746:			**<.0001**
A	1,011 (75)	661 (68)	
G	335 (25)	317 (32)	
rs2167079:			**.002**
G	918 (73)	598 (67)	
A	332 (27)	292 (33)	

NOTE.—Results include the additional cases and controls used in the replication series.

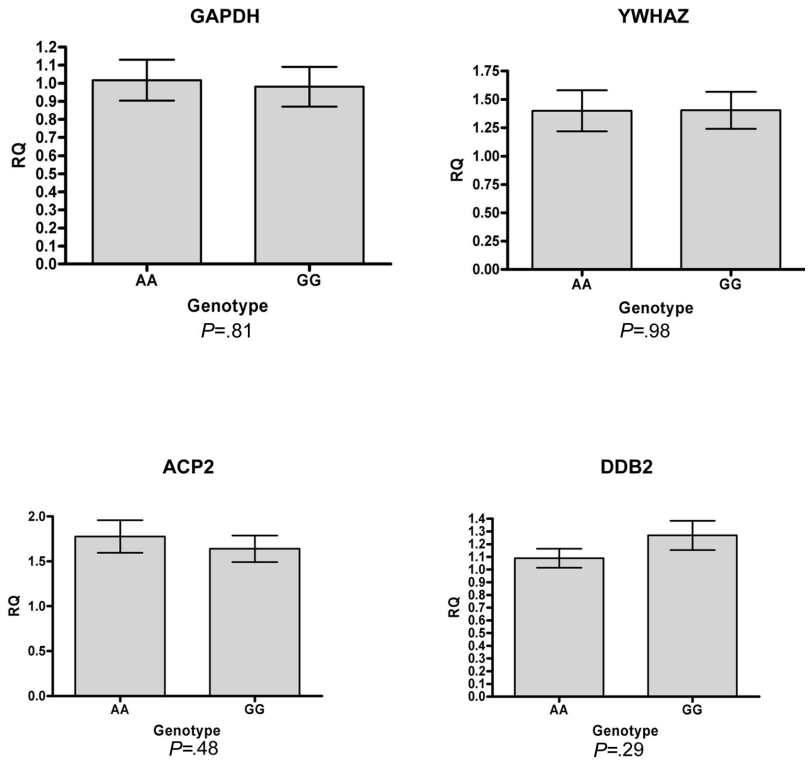

Figure 3. Relative mRNA expression with TATA-binding protein as an endogenous control. Plotted are relative levels of glyceraldehyde-3-phosphate dehydrogenase (GAPDH); tyrosine 3-monooxygenase/tryptophan 5-mono-oxygenase activation protein, zeta polypeptide (YWHAZ); ACP2 (assay Hs00155636_m1 [Applied Biosystems]); and DDB2 (assay Hs00172068_m1 [Applied Biosystems]) for 20 carriers of the *rs901746* AA (neutral) genotype and 20 carriers of the *rs901746* GG (risky) genotype. SE is denoted by the error bars. None of the comparisons between AA and GG carriers reach the level of statistical significance (*P* values noted below each graph). Similar results are seen when GAPDH or YWHAZ was used as the endogenous control (data not shown). RQ = relative quantity.

OR for the GG genotype compared with all other genotypes in the series was 2.2 (95% CI 1.4–3.4). To confirm that the *rs901746* association observed is not a control frequency artifact, we examined allele frequencies for *rs901746* in 250 cognitively normal controls recently published in a Parkinson disease (PD [MIM 168600]) GWA study.[38] We found that the frequency of the *rs901746* G allele in this independent control series was 0.27, consistent with our observed control frequencies (0.22 and 0.27).

The genomic context near *rs901746* was examined by downloading the CEPH-from-Utah SNP genotypes for 100 kb around *rs901746* from the HapMap genome browser and by examining the LD patterns and haplotype-block structure of the region with use of the Haploview software.[39] *rs901746* lies in the middle of a haplotype block encompassing at least two genes—the *DDB2* gene and the lysosomal acid phosphatase 2 (*ACP2* [MIM 171650]) gene—and can extend into the 3′ of another gene, nuclear receptor subfamily 1, group H, member 3 (*NR1H3* [MIM 602423]), depending on the type of haplotype-block definition used.[39] Variation in this 100-kb region could be fully described by 16 additional tag SNPs. These tag SNPs were genotyped in all PSP series, and both single and multimarker analysis was performed on the combined series (table 3). Single-marker analysis showed that nine tag SNPs showed significant allelic association. Of these, five tag SNP associations were highly significant (P values $\leq .003$), with *rs10501320* showing the greatest association after *rs901746* ($P < .0001$).

We examined the *DDB2/ACP2* tag SNP data set in different age groups in our combined series of cases and controls to see whether looking at younger cases might help further refine the associated region, as it had for the *MAPT* locus, where younger cases show a stronger association with the H1/H1 genotype.[22] Pathologically confirmed cases were divided into "young" and "old" groups on the basis of median age at death (75 years), and single-marker allelic association statistics were calculated using 2 × 2 contingency tables and were examined using χ^2 tests. On the whole, all of the SNPs that show significant association in the combined PSP case set also show significant association ($P < .05$) with the younger case subset, whereas there is less-significant association observed for the older cases.

To refine the disease-associated region, we performed haplotype-inference analysis in the young cases versus all controls, using a sliding-window approach.[40] As described elsewhere, this type of approach was key in refining the

associated region on the *MAPT* H1 haplotype.[22] However, as figure 2 displays, when data from all the tag SNPs were included in the analysis, there was no obvious resolution of the associated region when the young cases were considered separately. This may reflect the fact that the contribution to the overall signal of the association at this locus was not as great with the younger cases as had been seen with the *MAPT* locus; therefore, the sample size and/or the number of informative SNPs was inadequate to detect a smaller associated region.

Since a haplotype-inference approach was unsuccessful in narrowing the associated region, we decided to identify additional novel SNPs that may represent functional variant(s) accounting for increased risk of disease by sequencing a series of 18 subjects with PSP who had the various genotypes at *rs901746,* the majority of whom carried the risky GG genotype (n = 11 GG, n = 5 AG, and n = 2 AA) (table 4). Primers were designed to fully sequence coding exons of both *DDB2* and *ACP2.* Only one SNP, found 63 bp downstream of exon 3 in *ACP2,* was not already in the dbSNP database; however, this SNP appeared to be in near-complete LD with nearby *rs10838677* in exon 5 of *ACP2,* encoding a silent change (L165L). Interestingly, a number of SNPs identified through sequencing appeared to be in near-complete LD with *rs901746,* including *rs2167079,* a coding SNP in *ACP2* in which the minor allele changes the amino acid at position 29 from an arginine to a glutamine (R29Q). This converts the protein sequence to the mouse amino acid residue at the equivalent position. Interestingly, this position in ACP2 is predicted to encode the signal peptidase cleavage site,[41] suggesting that carriers of the minor allele encoding glutamine at position 29 may have altered cleavage of the signal peptide compared with those encoding arginine at that position. Since this SNP could affect function of the protein, we genotyped it through the combined series. Results from this analysis are shown in table 5. Overall, the LD between *rs901746* and *rs2167079* was high (cases r^2 = 0.97; controls r^2 = 0.94). As expected, significant allelic association was observed with *rs2167079* (P = .002); however, this was not any greater than the association observed with *rs901746,* suggesting that *rs2167079* is unlikely to fully explain the association at the *DDB2/ACP2* locus. We tested the R29Q variant for dominant, recessive, and additive models, and the additive model best fit the data (P < .0001).

In an alternative method for determining the gene responsible for disease risk at this locus, expression analysis was performed on the *DDB2* and *ACP2* genes. Analysis was performed, using real-time Taqman expression assays (Applied Biosystems), on mRNA extracted from the cerebella of 20 *rs901746* AA and 20 *rs901746* GG genotype carriers, to determine whether risk variants at the *DDB2/ACP2* locus have a direct effect on gene expression. Unfortunately, although DDB2 transcript levels are slightly increased in cases with a GG genotype, no significant differences were observed between the cases with AA and GG genotypes for either *DDB2* or *ACP2* mRNA levels (for *DDB2,* P = .29; for *ACP2,* P = .48) (fig. 3).

GWA studies are appealing because of their lack of bias, in that they represent a model-free approach for identification of new and novel genes that are involved in a disease process that may never be identified using other methodologies. However, even now, individually genotyping hundreds of individuals to perform a "traditional" GWA is not feasible for many rarer diseases, including PSP, because of the lack of available funding. Therefore, this type of pooled genomewide approach potentially represents a fast and economical initial solution to this problem. Pooling methods lack the analytical flexibility inherent in a traditional genomewide study because it is not possible to reanalyze the data with use of subgroups of cases or controls or to perform true haplotype-scanning analyses. However, there is still some uncertainty about how best to analyze the large amounts of individual genotype data used in GWA studies. An early GWA study of the PD showed problems in replication of results, potentially because of problems in study design.[42–47]

Although pooling methods clearly have limitations, the analysis procedures we used in the GenePool software (TGen Bioinformatics Research Unit) were developed using individual genotype data from samples that were also pooled, thereby allowing the algorithms to be adjusted until they predicted SNP ranks on the basis of what was known from the individual genotype data.[32] In the present analysis, we had prior knowledge that the *MAPT* H1 haplotype is associated with disease, so it could serve as a positive control for the genomewide analysis.

The identification of a new risk locus for PSP on chromosome 11 from the pooled genomewide approach was confirmed in a second U.S. PSP case-control series, with similar allele and genotype frequencies. Closer examination of this locus by dense SNP genotyping suggests that the association spans the entire haplotype block containing the *DDB2* and *ACP2* genes. Examination of potential functional variants yielded no definitive explanation for the observed association.

Both *ACP2* and *DDB2* are reasonable candidate genes that highlight previously implicated pathways for neurodegenerative disease. There are many lines of evidence suggesting a role for lysosomes and autophagic processes in neurodegeneration. Autophagy has been implicated in the clearance of protein aggregates, a common feature of many neurodegenerative disorders.[48,49] Interestingly, patients with lysosomal-storage disorders often exhibit neurological phenotypes with pathology similar to that seen in PSP.[50–52] Two lines of evidence implicate *ACP2* in neurodegeneration. First, it has been reported that, in brains of subjects with Alzheimer disease (AD [MIM 104300]), microglia surrounding the amyloid plaques stain strongly for ACP2.[53] In addition, cerebrospinal fluid from half of the examined subjects with AD showed evidence of ACP2 activity, whereas patients not affected with AD showed no activity.[53] These results leave open the question of whether

ACP2 in AD is just a secondary marker of neurodegeneration or perhaps plays a more active role in the neurodegenerative process. Second, knockouts and mutations of *Acp2* in mice have neurological phenotypes.[54,55] Neuropathology of *Acp2*$^{-/-}$ tissue showed increased lysosomal staining (as detected by lamp-1 and cathepsin D immunoreactivity), primarily in glial cells. Interestingly, ~7% of these *Acp2*$^{-/-}$ mice presented with generalized seizures after age 8 wk, and it has been suggested that this phenotype may be correlated with the defective lysosomal storage observed in glial cells.[54] The observation that loss of Acp2 causes deficits in glial lysosomal storage in the *Acp2*$^{-/-}$ mice may also be significant, given that, in PSP, there is abundant MAPT-inclusion pathology within glia (astrocytes and olgodendroglia), as well as in neurons.[56]

Mutations in the *DDB2* gene are responsible for xeroderma pigmentosum (XP) complementation group E (XPE [MIM 278740]). Interestingly, some mutations in the nucleotide excision–repair pathway that cause the diseases XP and Cockayne syndrome (MIM 216400) present with neurological phenotypes; however, XPE does not seem to be one of them.[57,58] DDB2 forms a ubiquitin E3-ligase complex, with DNA damage-binding protein 1 (DDB1 [MIM 600045]) and Cullin 4a (CUL4A [MIM 603137]), that binds damaged DNA. Both histone H2A (H2AA [MIM 603137]) and XP complementation group C (XPC [MIM 278720]) proteins have been implicated as substrates for the DDB1/DDB2/CUL4A complex upon activation.[59,60] Ubiquitination of histone H2A may change local chromatin configuration at the damage site, thereby allowing access to other DNA-repair proteins farther down the pathway.[60] The accumulation of damaged DNA in aging brain suggests that DNA-repair capacity is reduced as we age and appears to be selective to genes important in learning and memory. Interestingly, there is evidence of brain-specific alternatively spliced forms of *DDB2* that splice out either exons 4–7 or exons 4 and 6 alone.[61] The proteins encoded by these alternatively spliced transcripts act as dominant negative inhibitors of DNA repair, when tested in an in vitro system.[61] It will be interesting to tease apart which gene or genes at this locus are involved in conferring risk of PSP, but functional studies, rather than genetic ones, will probably be required to address these issues.

Given the size of the association seen at the *DDB2/ACP2* locus, the fact that the described PSP series represents the largest collection of PSP-affected subjects worldwide, and the fact that our U.S. replication series is underpowered to detect changes with an OR <2.0, we may be at the limit of what can be consistently detected and confirmed using the case-control populations available. Six additional weaker loci were identified in the genomewide screen that still need to be analyzed in detail, and it will be interesting to examine this potential power issue in closer detail. This genomewide analysis has identified a novel second locus implicated in PSP risk, accelerating research and the hope of identifying effective therapeutics for this devastating disease.

Acknowledgments

We are grateful to the patients and their families for their cooperation in this project. We also thank the Society for Progressive Supranuclear Palsy brain bank for providing samples. We thank John Gonzalez and Monica Castanedes-Casey in the Neuropathology Laboratory at Mayo Clinic Jacksonville for technical assistance. This research was funded by National Institutes of Aging grant PO1 AG 17216 (to M.H., D.W.D., and M.J.F.); the Mayo Foundation, USA; The Society for Progressive Supranuclear Palsy; and a Robert and Clarice Smith Fellowship. S.M. was funded by the National Research Service Award postdoctoral fellowship AG24030. Funding was also provided by a foundation grant from the Stardust Foundation (to D.W.C.) and the National Institutes of Health Neuroscience Blueprint grant 1U24NS043571 (to D.A.S.).

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/SNP/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PSP, *MAPT, DDB2,* PD, *ACP2, NR1H3,* AD, XPE, Cockayne syndrome, DDB1, CUL4A, H2AA, and XPC)
TGen Bioinformatics Research Unit, http://bioinformatics.tgen.org/ (for the GenePool source code)

References

1. Steele JC, Richardson JC, Olszewski J (1964) Progressive supranuclear palsy: a heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Arch Neurol 10:333–359
2. Litvan I (2004) Update on progressive supranuclear palsy. Curr Neurol Neurosci Rep 4:296–302
3. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, Goetz CG, Golbe LI, Grafman J, Growdon JH, et al (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP International Workshop. Neurology 47:1–9
4. Litvan I, Agid Y, Jankovic J, Goetz C, Brandel JP, Lai EC, Wenning G, D'Olhaberriague L, Verny M, Chaudhuri KR, et al (1996) Accuracy of clinical criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome). Neurology 46:922–930
5. Hauw JJ, Daniel SE, Dickson D, Horoupian DS, Jellinger K, Lantos PL, McKee A, Tabaton M, Litvan I (1994) Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Neurology 44: 2015–2019
6. Litvan I, Hauw JJ, Bartko JJ, Lantos PL, Daniel SE, Horoupian DS, McKee A, Dickson D, Bancher C, Tabaton M, et al (1996) Validity and reliability of the preliminary NINDS neuropathologic criteria for progressive supranuclear palsy and related disorders. J Neuropathol Exp Neurol 55:97–105
7. Rojo A, Pernaute RS, Fontan A, Ruiz PG, Honnorat J, Lynch T, Chin S, Gonzalo I, Rabano A, Martinez A, et al (1999) Clinical genetics of familial progressive supranuclear palsy. Brain 122:1233–1245 (erratum 123:419)
8. Ros R, Thobois S, Streichenberger N, Kopp N, Sanchez MP, Perez M, Hoenicka J, Avila J, Honnorat J, de Yebenes JG (2005)

A new mutation of the τ gene, G303V, in early-onset familial progressive supranuclear palsy. Arch Neurol 62:1444–1450
9. Spillantini MG, Goedert M (2000) *Tau* mutations in familial frontotemporal dementia. Brain 123:857–859
10. Stanford PM, Halliday GM, Brooks WS, Kwok JB, Storey CE, Creasey H, Morris JG, Fulham MJ, Schofield PR (2000) Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the *tau* gene: expansion of the disease phenotype caused by *tau* gene mutations. Brain 123:880–893
11. Wszolek ZK, Tsuboi Y, Uitti RJ, Reed L, Hutton ML, Dickson DW (2001) Progressive supranuclear palsy as a disease phenotype caused by the S305S tau gene mutation. Brain 124: 1666–1670
12. Pastor P, Pastor E, Carnero C, Vela R, Garcia T, Amer G, Tolosa E, Oliva R (2001) Familial atypical progressive supranuclear palsy associated with homozigosity for the delN296 mutation in the tau gene. Ann Neurol 49:263–267
13. Oliva R, Pastor P (2004) Tau gene delN296 mutation, Parkinson's disease, and atypical supranuclear palsy. Ann Neurol 55:448–449
14. Rossi G, Gasparoli E, Pasquali C, Di Fede G, Testa D, Albanese A, Bracco F, Tagliavini F (2004) Progressive supranuclear palsy and Parkinson's disease in a family with a new mutation in the tau gene. Ann Neurol 55:448
15. Ros R, Gomez Garre P, Hirano M, Tai YF, Ampuero I, Vidal L, Rojo A, Fontan A, Vazquez A, Fanjul S, et al (2005) Genetic linkage of autosomal dominant progressive supranuclear palsy to 1q31.1. Ann Neurol 57:634–641
16. Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M (1999) Association of an extended haplotype in the *tau* gene with progressive supranuclear palsy. Hum Mol Genet 8:711–715
17. Conrad C, Andreadis A, Trojanowski JQ, Dickson DW, Kang D, Chen X, Wiederholt W, Hansen L, Masliah E, Thal LJ, et al (1997) Genetic evidence for the involvement of τ in progressive supranuclear palsy. Ann Neurol 41:277–281
18. Higgins JJ, Golbe LI, De Biase A, Jankovic J, Factor SA, Adler RL (2000) An extended 5′-*tau* susceptibility haplotype in progressive supranuclear palsy. Neurology 55:1364–1367
19. Morris HR, Janssen JC, Bandmann O, Daniel SE, Rossor MN, Lees AJ, Wood NW (1999) The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases. J Neurol Neurosurg Psychiatry 66:665–667
20. Oliva R, Tolosa E, Ezquerra M, Molinuevo JL, Valldeoriola F, Burguera J, Calopa M, Villa M, Ballesta F (1998) Significant changes in the tau A0 and A3 alleles in progressive supranuclear palsy and improved genotyping by silver detection. Arch Neurol 55:1122–1124
21. Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, de Rijk P, Del-Favero J, van Broeckhoven C (2005) Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum Mol Genet 14: 1753–1762
22. Rademakers R, Melquist S, Cruts M, Theuns J, Del-Favero J, Poorkaj P, Baker M, Sleegers K, Crook R, De Pooter T, et al (2005) High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. Hum Mol Genet 14:3281–3292
23. Pittman AM, Myers AJ, Abou-Sleiman P, Fung HC, Kaleem M, Marlowe L, Duckworth J, Leung D, Williams D, Kilford L, et al (2005) 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
24. Kwok JB, Teber ET, Loy C, Hallupp M, Nicholson G, Mellick GD, Buchanan DD, Silburn PA, Schofield PR (2004) Tau haplotypes regulate transcription and are associated with Parkinson's disease. Ann Neurol 55:329–334
25. Caffrey TM, Joachim C, Paracchini S, Esiri MM, Wade-Martins R (2006) Haplotype-specific expression of exon 10 at the human *MAPT* locus. Hum Mol Genet 15:3529–3537
26. Myers AJ, Pittman AM, Zhao AS, Rohrer K, Kaleem M, Marlowe L, Lees A, Leung D, McKeith IG, Perry RH, et al (2006) The *MAPT* H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiol Dis (electronically published ahead of print December 14, 2006; accessed February 7, 2007)
27. Bansal A, van den Boom D, Kammerer S, Honisch C, Adam G, Cantor CR, Kleyn P, Braun A (2002) Association testing by DNA pooling: an effective initial screen. Proc Natl Acad Sci USA 99:16871–16874
28. Meaburn E, Butcher LM, Schalkwyk LC, Plomin R (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans. Nucleic Acids Res 34: e28
29. Josephs KA, Dickson DW (2003) Diagnostic accuracy of progressive supranuclear palsy in the Society for Progressive Supranuclear Palsy brain bank. Mov Disord 18:1018–1026
30. Caselli RJ, Osborne D, Reiman EM, Hentz JG, Barbieri CJ, Saunders AM, Hardy J, Graff-Radford NR, Hall GR, Alexander GE (2001) Preclinical cognitive decline in late middle-aged asymptomatic apolipoprotein E-e4/4 homozygotes: a replication study. J Neurol Sci 189:93–98
31. Caselli RJ, Hentz JG, Osborne D, Graff-Radford NR, Barbieri CJ, Alexander GE, Hall GR, Reiman EM, Hardy J, Saunders AM (2002) Apolipoprotein E and intellectual achievement. J Am Geriatr Soc 50:49–54
32. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, et al (2007) Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide–polymorphism association studies. Am J Hum Genet 80: 126–139
33. Lovmar L, Ahlford A, Jonsson M, Syvanen AC (2005) Silhouette scores for assessment of SNP genotype clusters. BMC Genomics 6:35
34. Pittman AM, Myers AJ, Duckworth J, Bryden L, Hanson M, Abou-Sleiman P, Wood NW, Hardy J, Lees A, de Silva R (2004) The structure of the tau haplotype in controls and in progressive supranuclear palsy. Hum Mol Genet 13:1267–1274
35. Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, Puffenberger EG, Pearson JM, Stephan DA (2005) Identification of disease causing loci using an array-based genotyping approach on pooled DNA. BMC Genomics 6:138
36. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al (2005) A common inversion under selection in Europeans. Nat Genet 37:129–137
37. Evans W, Fung HC, Steele J, Eerola J, Tienari P, Pittman A, Silva R, Myers A, Vrieze FW, Singleton A, et al (2004) The tau H2 haplotype is almost exclusively Caucasian in origin. Neurosci Lett 369:183–185
38. Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez

D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, et al (2006) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 5:911–916
39. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
40. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70:425–434
41. Pohlmann R, Krentler C, Schmidt B, Schroder W, Lorkowski G, Culley J, Mersmann G, Geier C, Waheed A, Gottschalk S, et al (1988) Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment. EMBO J 7:2343–2350
42. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PV, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
43. Clarimon J, Scholz S, Fung H-C, Hardy J, Eerola J, Hellstrom O, Chen C-M, Wu Y-R, Tienari PJ, Singleton A (2006) Conflicting results regarding the semaphorin gene (*SEMA5A*) and the risk for Parkinson disease. Am J Hum Genet 78:1082–1084 (author reply 78:1092–1094)
44. Farrer MJ, Haugarvoll K, Ross OA, Stone JT, Milkovic NM, Cobb SA, Whittle AJ, Lincoln SJ, Hulihan MM, Heckman MG, et al (2006) Genomewide association, Parkinson disease, and *PARK10*. Am J Hum Genet 78:1084–1088 (author reply 78: 1092–1094)
45. Goris A, Williams-Gray CH, Foltynie T, Compston DAS, Barker RA, Sawcer SJ (2006) No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am J Hum Genet 78:1088–1090 (author reply 78:1092–1094)
46. Li Y, Rowland C, Schrodi S, Laird W, Tacey K, Ross D, Leong D, Catanese J, Sninsky J, Grupe A (2006) A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. Am J Hum Genet 78:1090–1092 (author reply 1092–1094)
47. Myers DR (2006) Considerations for genomewide association studies in Parkinson disease. Am J Hum Genet 78:1081–1082
48. Webb JL, Ravikumar B, Atkins J, Skepper JN, Rubinsztein DC (2003) α-Synuclein is degraded by both autophagy and the proteasome. J Biol Chem 278:25009–25013
49. Qin ZH, Wang Y, Kegel KB, Kazantsev A, Apostol BL, Thompson LM, Yoder J, Aronin N, DiFiglia M (2003) Autophagy regulates the processing of amino terminal huntingtin fragments. Hum Mol Genet 12:3231–3244
50. Auer IA, Schmidt ML, Lee VM, Curry B, Suzuki K, Shin RW, Pentchev PG, Carstea ED, Trojanowski JQ (1995) Paired helical filament tau (PHFtau) in Niemann-Pick type C disease is similar to PHFtau in Alzheimer's disease. Acta Neuropathol (Berl) 90:547–551
51. Love S, Bridges LR, Case CP (1995) Neurofibrillary tangles in Niemann-Pick disease type C. Brain 118:119–129
52. Suzuki K, Parker CC, Pentchev PG, Katz D, Ghetti B, D'Agostino AN, Carstea ED (1995) Neurofibrillary tangles in Niemann-Pick disease type C. Acta Neuropathol (Berl) 89:227–238
53. Omar R, Pappolla M, Argani I, Davis K (1993) Acid phosphatase activity in senile plaques and cerebrospinal fluid of patients with Alzheimer's disease. Arch Pathol Lab Med 117: 166–169
54. Saftig P, Hartmann D, Lullmann-Rauch R, Wolff J, Evers M, Koster A, Hetman M, von Figura K, Peters C (1997) Mice deficient in lysosomal acid phosphatase develop lysosomal storage in the kidney and central nervous system. J Biol Chem 272:18628–18635
55. Mannan AU, Roussa E, Kraus C, Rickmann M, Maenner J, Nayernia K, Krieglstein K, Reis A, Engel W (2004) Mutation in the gene encoding lysosomal acid phosphatase (Acp2) causes cerebellum and skin malformation in mouse. Neurogenetics 5:229–238
56. Ishizawa K, Dickson DW (2001) Microglial activation parallels system degeneration in progressive supranuclear palsy and corticobasal degeneration. J Neuropathol Exp Neurol 60:647–657
57. Cleaver JE (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. Nat Rev Cancer 5:564–573
58. Robbins JH, Brumback RA, Moshell AN (1993) Clinically asymptomatic xeroderma pigmentosum neurological disease in an adult: evidence for a neurodegeneration in later life caused by defective DNA repair. Eur Neurol 33:188–190
59. Sugasawa K, Okuda Y, Saijo M, Nishi R, Matsuda N, Chu G, Mori T, Iwai S, Tanaka K, Tanaka K, et al (2005) UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. Cell 121:387–400
60. Kapetanaki MG, Guerrero-Santoro J, Bisi DC, Hsieh CL, Rapic-Otrin V, Levine AS (2006) The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. Proc Natl Acad Sci USA 103:2588–2593
61. Inoki T, Yamagami S, Inoki Y, Tsuru T, Hamamoto T, Kagawa Y, Mori T, Endo H (2004) Human DDB2 splicing variants are dominant negative inhibitors of UV-damaged DNA repair. Biochem Biophys Res Commun 314:1036–1043