Multiple Regions of α-Synuclein Are Associated with Parkinson's Disease

Jakob C. Mueller, PhD,[1] Julia Fuchs, MD,[2] Anne Hofer, MD,[2] Alexander Zimprich, MD,[3]
Peter Lichtner, PhD,[1] Thomas Illig, PhD,[4] Daniela Berg, MD,[5] Ullrich Wüllner, MD,[6]
Thomas Meitinger, PhD,[1] and Thomas Gasser, MD[2]

α-Synuclein is considered to play an important role in the pathogenesis of both the rare familial and the common sporadic forms of Parkinson's disease. Previous reports primarily have tested the association of α-synuclein promoter polymorphisms with idiopathic Parkinson's disease, but results are controversial. We first characterized the linkage disequilibrium structure of the α-synuclein gene region with a dense set of 56 genetic markers and subsequently performed two independent case–control association analyses using tagging markers. We could distinguish two large linkage disequilibrium blocks spanning the α-synuclein gene. Several markers within the 3′-block around exons 5 and 6 showed strong association with Parkinson's disease ($p = 0.00009$). Effects of the associated variants might be mediated by regulatory elements in this highly conserved region or by a frequency shift in a previously described splice variant lacking exon 5. A direct association with promoter polymorphisms could not be replicated in our sample set. A second set of markers in the 5′-block of the gene were also significantly associated with Parkinson's disease, when young patients and female subjects were analyzed separately. These findings indicate locus heterogeneity for the pathogenesis of Parkinson's disease in different genetic or physiological environments, related to sex and age.

Ann Neurol 2005;57:535–541

Three point mutations in the α-synuclein (*SNCA*) gene have been identified so far as causative agents for familial Parkinson's disease (PD) with early-onset and autosomal dominant inheritance.[1–3] These mutations presumably lead to an increased propensity of the α-synuclein protein to aggregate, a process that is thought to be a crucial step in the molecular pathogenesis of the disease.[4] α-Synuclein is a major component of Lewy bodies, which are present in familial and in sporadic PD.[5] However, point mutations of the α-synuclein gene appear to be a rare cause of familial PD and have been excluded in a large number of patients with sporadic and familial PD.[6,7] The presence of α-synuclein–containing aggregates in the absence of coding *SNCA* mutations in sporadic PD suggests that other α-synuclein modifications such as alternative splicing, alterations in gene expression, or additional interacting genes may contribute to sporadic PD.

Evidence for a role of expression levels of wild-type α-synuclein in PD pathogenesis was brought forward by the finding that a triplication of the α-synuclein locus causes autosomal dominant parkinsonism in a large Iowan kindred (Spellman–Muenter and Waters–Miller families).[8] This mutation leads to a doubling in effective gene dosage resulting in a doubling of RNA and protein levels of α-synuclein.[8,9] In addition, several haplotypes in the promotor region of the α-synuclein gene, including the sequence repeat element Rep1, were shown to be associated with an increased risk for sporadic PD, further supporting a putative role of expression levels of α-synuclein.[7,10,11] Results of replication studies, however, have been contradictory.[12,13]

To more thoroughly assess the role of α-synuclein in sporadic PD, we determined the linkage disequilibrium (LD) and haplotype structure of *SNCA* in a whole-gene approach, as Neale and Sham[14] suggested previously. Here, we demonstrate a strong association of

---

From the [1]Institute for Human Genetics, GSF–National Research Centre for Environment and Health, Neuherberg, Germany; [2]Department for Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; [3]Department of Neurology, AKH, University of Vienna, Vienna, Austria; [4]Institute for Epidemiology, GSF–National Research Centre for Environment and Health, Neuherberg, Germany; [5]Department for Medical Genetics, Hertie-Institute for Clinical Brain Research, Tübingen, Germany; and [6]Department of Neurology, University of Bonn, Bonn, Germany.

Received Nov 5, 2004, and in revised form Jan 28, 2005. Accepted for publication Jan 28, 2005.

This article includes supplementary materials available via the Internet at http://www.interscience.wiley.com/jpages/0364-5134/suppmat

Address correspondence to Dr Gasser, Hertie-Institute for Clinical Brain Research, Department for Neurodegenerative Diseases, University of Tübingen, Otfried-Müller-Strasse 27, 72076 Tübingen, Germany. E-mail: thomas.gasser@med.uni-tuebingen.de

Published by Wiley-Liss, Inc., through Wiley Subscription Services

haplotypes within an LD block at the 3′-end of *SNCA* with sporadic PD in two independent data sets. Additional associated loci appeared within subgroups of patients.

Subjects and Methods

Recruitment

PD patients were recruited by participating institutions in Munich, Tübingen, Bonn, and Rostock. Patients were examined by specialists in movement disorders, and diagnoses were established according to the United Kingdom Parkinson's Disease Society Brain Bank criteria.[15] After obtaining informed consent, blood samples were drawn for DNA extraction from 669 German sporadic PD patients. The median age at onset was 55.4 ± 19.1 years. The PD patients were from two independent recruitments referred to as Sample Sets I and II. The first set of 340 PD patients from clinical centers in Munich and Tübingen had a male:female ratio of 1.38, and the second replicating set of 329 PD patients mainly from clinical centers in Bonn and Rostock had a male:female ratio of 1.47. A total of 680 healthy, age- and sex-matched subjects from the KORA (Cooperative Research in the Region of Augsburg) Survey 2000, which studied a large population-based sample,[16] were used as control subjects for the first set, and 322 additional healthy, age- and sex-matched subjects of KORA were used as control for the second set of PD patients. The KORA group consists of H. E. Wichmann (speaker), H. Löwel, C. Meisinger, T. Illig, R. Holle, J. John, and their coworkers, who are responsible for the design and conduct of the KORA studies.

Genotyping

We genotyped single nucleotide polymorphisms (SNPs) covering a genomic region of 178kb, including the α-synuclein gene spanning about 111kb and its flanking 3′- and 5′-regions (Fig). A total of 64 SNPs with an average distance between SNPs of $\approx$2.5kb were identified using public databases (see supplementary table available online for more information). All SNPs showed high genotyping quality and Hardy–Weinberg equilibrium in the control subjects. Genotyping was performed using the matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry method (Sequenom, San Diego, CA). The previously described sequence repeat marker Rep1 was amplified according to Farrer and colleagues.[10] In a pilot study, a subset of 170 case–control pairs of Sample Set I was genotyped for all markers to define the fine-scale LD structure. Tagging and previously identified markers were selected as described later in this article. In a second step, the remaining patients of Sample Set I and those of Sample Set II were genotyped only for these selected SNPs.

Genotyping was performed in the Genome Analysis Center of the GSF.

Statistical Analysis

We followed the D′ method of Gabriel and colleagues[17] for constructing LD blocks as implemented in the program Haploview (J. Barrett, Whitehead Institute for Biomedical Research, Cambridge, MA). Only SNPs with a minor allele frequency greater than 0.1 were considered for the block definition. An optimal set of tagging SNPs that distinguish 95% of all chromosomes (most common haplotypes) within blocks were identified with a nested algorithm implemented in the program HaploBlockFinder.[18]

Logistic regression was used to test single-marker associations based on subjects as units and assuming a dose effect of alleles with the heterozygote effect being intermediate between the homozygote effects. The sequence repeat marker

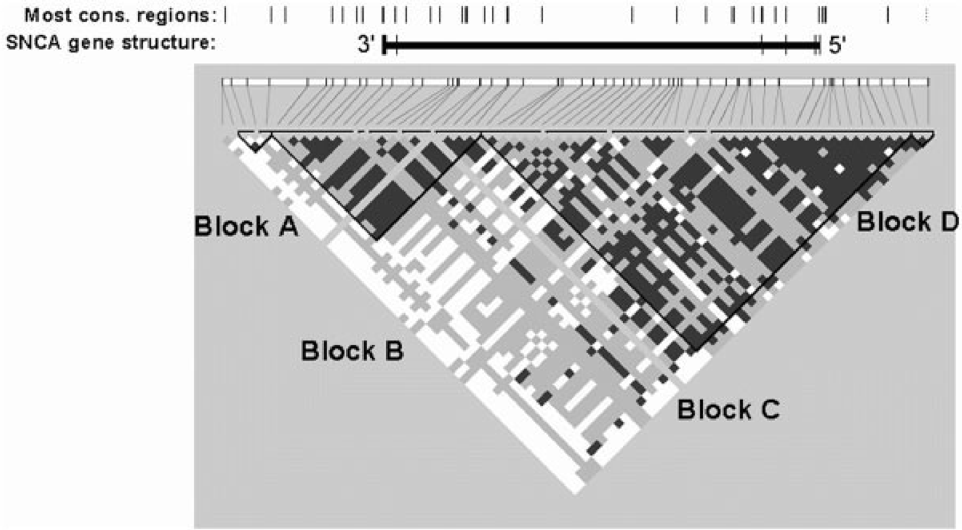

Fig. Linkage disequilibrium (LD) structure, physical gene map, and conservation of the SNCA *region. Blocks are defined according to Gabriel and colleagues[17]: black cells = strong LD; gray cells = intermediate or ambiguous; white cells = evidence for historical recombination. All 65 markers are shown. The gene structure of* SNCA *is shown (top), with vertical bars indicating exons. Most conserved elements predicted by the PhastCons program (implemented at the University of California at Santa Cruz browser[32]) are also shown (top). The prediction is based on the aligned sequences of human, chimpanzee, mouse, rat, dog, chicken, fugu, and zebrafish.*

Rep1 was treated as a two-allele system after pooling the rare alleles. To correct for multiple testing, we evaluated the table-wide significance for each data set by a randomization procedure. The affected status of the individuals was permuted, and the single-marker association analyses were repeated 1,000 times. The table-wide 5% significance level is the level at which 5% of the permuted data sets show at least one smaller p value, regardless of at which marker.

Haplotype phase within the associated LD block was estimated by a Bayesian method, which accounts for the evolutionary relation of estimated haplotypes, and is implemented in software package PHASE.[19] The association of haplotypes with PD was inferred by the haplotype trend regression method that Zaykin and colleagues[20] proposed. Again, individuals were units, and a dose effect of haplotypes was assumed. The maximum likelihood estimates of the odds ratios corresponded to haplotype or allele relative risks in a multiplicative model; that is, the genotype relative risk of one homozygote was the square of the genotype relative risk of the heterozygote, if the other homozygote was taken as reference.

Results

Based on 55 common SNPs, which showed a minor allele frequency of more than 10% in a sample of 170 control subjects (pilot study of Sample Set I), two large blocks and two marginal LD blocks could be defined for the *SNCA* gene (see Fig). Average spacing of SNP markers was ≈3 kb. The largest block "C" comprises the 5′-region and the first 4 exons, whereas the second largest block "B" extends from intron 4 through exons 5 and 6 to the 3′-region of *SNCA*. We selected 14 haplotype tagging SNPs in addition to 5 SNPs that were already significant in the pilot study ($p < 0.01$) for further genotyping. During the study, 14 additional SNPs and the repeat marker Rep1 from the promoter and the 3′-region were genotyped in Sample Set I to explore the boundaries of the two major blocks.

A strong association of several markers in block B with PD was found in Sample Set I, the smallest p value being 0.00009 (Table 1). The trend (multiplicative) model of association showed stronger association signals than the recessive or dominant association models. The p values of six markers in LD block B and one marker in LD block C showed table-wide significance according to a randomization test. The associated region therefore extends from the large intron 4 of the gene to the 3′ noncoding region. The group of associated markers comprises common variants with a minor allele frequency of 35 to 44% in the population. These markers are highly intercorrelated, with r^2 values of 0.70 to 0.99 among block B markers and 0.56 to 0.75 between the block C marker and the block B markers. Based on these associations, estimates of the relative risk are around 1.4 (95% confidence interval, 1.2–1.75).

The association analysis of SNCA markers with PD was replicated in Sample Set II (see Table 1). Again, a strong table-wide association was found in block B with a risk allele frequency of 38% and a relative risk of 1.4. The weaker p values are because of the smaller sample size of the control subjects in this replication set.

The analysis of common block B haplotypes did not indicate stronger association signals than single SNPs. Interestingly, two groups of evolutionarily related haplotypes are obvious: haplotypes 1, 2, 3, and 4 (Group I) and haplotypes 5 and 6 (Group II) (Table 2). Evolutionary intermediate haplotypes are missing. Within each haplotype group, odds ratios for PD risk follow the same trend: In Group I, odds ratios are less than 1, whereas in Group II, odds ratios are greater than 1. The signal strength of single-marker associations is reestablished only when these haplotype groups are tested as a whole for association with PD, indicating a common causal variant (see Table 2). We also analyzed 18 animals from different ape species: chimpanzee, gorilla, and orangutan. All carried the same haplotype, which differed from the risk haplotype 6 by only a single base substitution.

The relative haplotype risk associated with the haplotype groups is about 1.4, and it corresponds to a population-attributable risk of about 29%. With two alternative haplotype groups, however, it is not clear whether a risk allele or the complementary protective allele is causative. The population-attributable "protection" would be about 42%.

Subsequently, we analyzed our pooled Sample Sets I and II (or Sample Set I alone, if genotypes were not available for Sample Set II) about a possible association with a specific sex or age group (see Table 1). Control subjects were subdivided accordingly to maintain age and sex matching. Female and male subjects had similar age distributions, and thus were directly comparable. The region in LD block B (see earlier) showed consistently strong association signals in both female and male patients, as well as in both age groups (young and old PD groups, subdivided at the median age of 57 years). Interestingly, additional significantly associated regions were found for female and young patients in the intron 4 region of block C and 21.5kb 5′-upstream of the *SNCA* gene (LD block D). The two newly associated SNPs rs356188 and rs356186 in block C were highly intercorrelated with each other ($r^2 = 0.88$) and with the two SNPs of block D ($r^2 > 0.73$) and show similar allele frequencies, but exhibit low correlation with the significant markers of block B ($r^2 < 0.03$). This group of four additional markers may represent a second alternative causal variant that is most effective in female and early-onset PD patients. The phenotypic state explained by the genotypic variation was about 1.7% for single (the most associated)

Table 1. Genotypic Association Tests (Logistic Regression) and Allele Frequencies of Selected SNPs in the SNCA Region

				Sample set I (340 cases, 680 controls)			Sample set II (329 cases, 322 controls)			Sample set I (n_I) + sample set II, if available (n_{I+II})							
					Frequency of first allele			Frequency of first allele		Females (n_I=423, n_{I+II}=672)		Males (n_I=593, n_{I+II}=979)		Age < 57 y (n_I=509, n_{I+II}=698)		Age ≥57 y (n_I=508, n_{I+II}=958)	
SNP ID	Alleles	Region	Block ID	p-value	Cases	Controls	p-value	Cases	Controls	p-value	Odds ratio of first allele (95% CI)	p-value	Odds ratio of first allele (95% CI)	p-value	Odds ratio of first allele (95% CI)	p-value	Odds ratio of first allele (95% CI)
rs3857047	T/G	3′- region		0.70329	0.84	0.85				0.5667	0.90 (0.62-1.30)	0.8868	0.98 (0.70-1.36)	0.5421	0.90 (0.64-1.27)	0.9682	1.01 (0.71-1.44)
rs356229	A/G	3′- region		0.00593	0.61	0.67				**0.0005**	0.58 (0.43-0.79)	0.5479	0.93 (0.72-1.19)	0.4990	0.91 (0.70-1.19)	**0.0012**	0.63 (0.47-0.83)
rs3857048	T/C	3′- region	A	0.54033	0.89	0.89				0.8282	0.95 (0.60-1.51)	0.5696	0.89 (0.60-1.33)	0.9104	1.03 (0.66-1.61)	0.4009	0.84 (0.56-1.27)
rs3857049	A/T	3′- region	A	0.00625	0.97	0.98				0.0298	0.32 (0.11-0.90)	0.0796	0.51 (0.24-1.07)	0.0097	0.34 (0.15-0.78)	0.3124	0.61 (0.24-1.56)
rs3906628	T/C	3′- region	A	0.72880	0.11	0.12				0.8921	1.03 (0.66-1.60)	0.5898	0.90 (0.60-1.34)	0.6159	1.12 (0.72-1.75)	0.4076	0.85 (0.57-1.26)
rs356180	C/T	3′- region	B	0.07220	0.66	0.70				0.0034	0.63 (0.47-0.86)	0.9574	1.01 (0.78-1.30)	0.8982	0.98 (0.75-1.28)	0.0143	0.70 (0.52-0.93)
rs356169	C/A	3′- region	B	0.04636	0.37	0.32				0.0063	1.51 (1.12-2.04)	0.7238	1.05 (0.81-1.35)	0.5692	1.08 (0.83-1.40)	0.0238	1.39 (1.04-1.84)
rs2572323	A/G	3′- region	B	0.02333	0.34	0.29				0.0035	1.60 (1.17-2.19)	0.6285	1.07 (0.82-1.40)	0.9205	1.01 (0.77-1.34)	**0.0015**	1.63 (1.20-2.20)
rs356219	G/A	3′- region	B	**0.00009**	0.44	0.35	**0.00467**	0.46	0.38	**0.0001**	1.56 (1.25-1.96)	**0.0008**	1.37 (1.14-1.64)	**0.0002**	1.52 (1.22-1.89)	**0.0007**	1.39 (1.15-1.67)
rs356220	T/C	3′- region	B	**0.00025**	0.44	0.35	**0.00485**	0.46	0.38	**0.0004**	1.52 (1.20-1.89)	**0.0007**	1.39 (1.15-1.67)	**0.0006**	1.47 (1.18-1.82)	**0.0005**	1.41 (1.16-1.69)
rs356165	G/A	3′UTR	B	**0.00015**	0.44	0.35	**0.00555**	0.46	0.38	**0.0006**	1.49 (1.19-1.87)	**0.0004**	1.40 (1.16-1.68)	**0.0006**	1.46 (1.18-1.82)	**0.0006**	1.40 (1.15-1.69)
rs356204	A/G	Intron 4	B	**0.00039**	0.52	0.44				**0.0019**	1.58 (1.18-2.12)	0.0338	1.31 (1.02-1.69)	**0.0022**	1.52 (1.16-1.98)	0.0524	1.30 (1.00-1.70)
rs3822086	T/C	Intron 4	B	0.00249	0.10	0.06	0.01697	0.11	0.07	0.0215	1.61 (1.08-2.44)	**0.0014**	1.72 (1.23-2.38)	**0.0005**	2.00 (1.33-2.94)	0.0292	1.47 (1.04-2.08)
rs356203	G/A	Intron 4	B	**0.00026**	0.44	0.35	**0.00548**	0.46	0.38	**0.0008**	1.47 (1.18-1.85)	**0.0004**	1.39 (1.16-1.69)	**0.0006**	1.47 (1.18-1.82)	**0.0007**	1.39 (1.15-1.67)
rs356168	C/T	Intron 4	B	**0.00031**	0.53	0.44	0.17933	0.52	0.48	0.0031	1.39 (1.12-1.75)	0.0111	1.27 (1.05-1.54)	**0.0019**	1.41 (1.14-1.75)	0.0133	1.27 (1.05-1.52)
rs356199	C/T	Intron 4	C	0.24657	0.31	0.28	0.73808	0.31	0.30	0.0492	1.27 (1.00-1.62)	0.9096	1.01 (0.83-1.23)	0.1718	1.17 (0.93-1.47)	0.6619	1.05 (0.85-1.29)
rs356188	A/G	Intron 4	C	0.15669	0.83	0.81	0.01784	0.84	0.79	**0.0001**	1.82 (1.35-2.50)	0.8056	1.03 (0.81-1.32)	**0.0009**	1.64 (1.22-2.22)	0.5795	1.08 (0.84-1.37)
rs356187	A/G	Intron 4	C	0.21542	0.31	0.28	0.60062	0.32	0.30	0.0357	1.30 (1.02-1.65)	0.7735	1.03 (0.84-1.26)	0.1385	1.19 (0.95-1.49)	0.5171	1.07 (0.87-1.32)
rs356186	C/T	Intron 4	C	0.50907	0.84	0.83	0.09613	0.86	0.83	**0.0009**	1.69 (1.23-2.33)	0.7602	0.96 (0.75-1.23)	0.0352	1.39 (1.02-1.89)	0.6812	1.05 (0.82-1.35)
rs2737033	A/G	Intron 4	C	0.21774	0.70	0.72				0.0061	0.64 (0.46-0.88)	0.5039	1.10 (0.83-1.45)	0.4683	0.90 (0.68-1.19)	0.2970	0.85 (0.63-1.15)
rs2737029	G/A	Intron 4	C	**0.00158**	0.46	0.39	0.09994	0.46	0.42	**0.0010**	1.45 (1.16-1.82)	0.0392	1.22 (1.01-1.47)	**0.0001**	1.56 (1.23-1.92)	0.1361	1.15 (0.95-1.39)
rs2583959	G/C	Intron 4	C	0.25544	0.30	0.27	0.81992	0.31	0.31	0.0337	1.30 (1.02-1.67)	0.9361	0.99 (0.81-1.22)	0.2305	1.15 (0.91-1.46)	0.5369	1.07 (0.87-1.32)
rs3775444	C/T	Intron 4	C	0.08041	0.97	0.98	0.15877	0.99	0.98	0.7813	0.89 (0.39-2.01)	0.8802	0.95 (0.48-1.87)	0.4422	0.74 (0.35-1.59)	0.7459	1.13 (0.55-2.33)
rs2619369	G/A	Intron 4	C	0.04650	0.04	0.02				0.0100	3.67 (1.33-10.14)	0.5427	1.24 (0.62-2.49)	0.3018	1.50 (0.70-3.24)	0.0879	2.04 (0.91-4.58)
rs1812923	C/A	Intron 4	C	0.57114	0.56	0.55	0.09606	0.52	0.57	0.4854	0.92 (0.74-1.15)	0.7176	0.97 (0.80-1.17)	0.5271	0.93 (0.75-1.16)	0.7958	0.98 (0.81-1.18)
IVS4+66A->G	G/A	Intron 4	C	0.99687	0.47	0.47				0.5906	1.08 (0.81-1.44)	0.6295	0.94 (0.74-1.20)	0.4813	0.91 (0.70-1.18)	0.4566	1.11 (0.85-1.45)
rs2301135	C/G	promotor	C	0.64281	0.47	0.48				0.9624	0.99 (0.75-1.32)	0.5557	0.93 (0.73-1.18)	0.2460	0.86 (0.66-1.11)	0.5978	1.07 (0.83-1.40)
rs2619364	A/G	5′- region	C	0.23243	0.70	0.73				0.0095	0.65(0.47-0.90)	0.5342	1.09(0.82-1.43)	0.3144	0.86(0.65-1.14)	0.5127	0.90(0.66-1.22)
rs2583987	C/A	5′- region	C	0.16770	0.29	0.27				0.0060	1.56 (1.13-2.14)	0.5977	0.93 (0.71-1.22)	0.2629	1.17 (0.89-1.54)	0.4234	1.13 (0.83-1.54)
rs2583988	C/T	5′- region	C	0.33576	0.70	0.73				0.0077	0.65(0.47-0.89)	0.3082	1.15(0.87-1.52)	0.3110	0.86(0.65-1.14)	0.7468	0.95(0.70-1.28)
Rep1 (rare alleles 261 bp / pooled)	rare	5′- region	C	0.31812	0.67	0.69				0.2613	1.19(0.87-1.62)	0.0265	0.74(0.57-0.96)	0.2143	0.83(0.63-1.10)	0.8102	0.96(0.73-1.27)
rs1372525	A/G	5′- region	C	0.91544	0.51	0.51				0.5885	0.92(0.69-1.23)	0.5365	1.08(0.84-1.38)	0.4587	1.10(0.84-1.43)	0.5424	0.92(0.70-1.19)
rs2737026	C/T	5′- region	D	0.12206	0.81	0.78	0.14492	0.81	0.78	**0.0007**	1.64 (1.22-2.19)	0.8612	1.02 (0.81-1.28)	**0.0017**	1.57 (1.18-2.08)	0.8695	1.02 (0.81-1.28)
rs2736994	C/T	5′- region	D	0.11149	0.82	0.79	0.16522	0.83	0.80	**0.0002**	1.75 (1.29-2.36)	0.8859	1.02 (0.81-1.28)	**0.0009**	1.63 (1.21-2.19)	0.8439	1.02 (0.82-1.28)

P-values in bold indicate significance (at the 5% significance level) after adjustment for multiple testing within a sample set. The adjusted significance threshold for sample set I is $p = 0.0022$, and for sample set II $p = 0.0078$.

SNPs and increased to 2.7% if two SNPs (one of each associated region) were combined.

Discussion

We have characterized the LD structure and association patterns with PD in a 178kb region on chromosome 4q21.3 containing the *SNCA* gene. We found that the *SNCA* gene and its flanking regions can be described by 2 large LD blocks of 50 (block B) and 93kb (block C). Although only about one third of the 44 common SNPs in the *SNCA* region genotyped by the current HapMap project [21] correspond to our SNP set, which makes a direct comparison difficult, the block structure indicated by the HapMap data is in line with our results. Several common SNPs within LD block B ranging from within intron 4 to about 10kb into the 3′-region of *SNCA* were strongly associated with PD in two independent data sets. It may be of interest that one tagging marker defined on the HapMap data by a standard procedure would have indicated the association signal, because of complete LD ($r^2 = 1$) with two of the associated loci. Although the LD blocks B and C are not completely independent, as indicated by the appearance of association signals in

Table 2. Haplotype Associaton Analysis on All Common Haplotypes (Frequency > 1%) and Evolutionary Related Haplotype Groups in LD Block B

Haplotype ID	Haplotype	Frequency in Cases	Frequency in Controls	Individual p-value	Odds ratio (95% CI)	Pooled haplotypes	Individual p-value	Odds ratio (95% CI)
Sample set I								
1	`TAGACAGCAT`	0.006	0.015	0.0760	0.37 (0.12 – 1.10)	1+2+3+4	**0.0004**	0.72 (0.60-0.86)
2	`CAGACAGCAT`	0.474	0.543	**0.0031**	0.75 (0.62 – 0.91)			
3	`CCGACAACAC`	0.025	0.029	0.5490	0.83 (0.46 – 1.50)			
4	`CAGACAACAC`	0.056	0.055	0.9317	1.01 (0.68 – 1.52)			
5	`TCAGTGACGC`	0.329	0.280	0.0245	1.25 (1.03 – 1.52)	5+6	**0.0002**	1.42 (1.18-1.71)
6	`CAGGTGATGC`	0.092	0.055	**0.0024**	1.72 (1.21 – 2.43)			
Sample set II								
1+2	`---ACA-CAT`	0.478	0.518	0.1437	0.84 (0.68-1.05)	1+2+3+4	**0.0047**	0.73 (0.58-0.91)
3+4	`---ACA-CAC`	0.060	0.097	**0.0120**	0.57 (0.37-0.88)			
5	`---GTG-CGC`	0.343	0.307	0.1700	1.17 (0.93-1.48)	5+6	**0.0084**	1.35 (1.08-1.68)
6	`---GTG-TGC`	0.106	0.070	**0.0230**	1.58 (1.06-2.34)			

P-values in bold are significant after Bonferroni correction.

both blocks, the accumulation of significant markers in the 3′-block suggests that the major causal variant in our population is located in the 3′-end rather than in the 5′-end of *SNCA*. Further evidence for the functional significance of the 3′-region comes from cross-species conservation patterns. A high number of strongly conserved elements are distributed across block B (see Fig). A detailed analysis of α-synuclein human and mouse genomic sequences showed a high conservation of protein-coding regions, the 3′-untranslated regions (up to 80.4% identical), and short regions (<60bp) within intron 4 with an average cross-species identity of at least 75%.[22] In contrast, the 5′-untranslated regions and 5′-upstream regions except for the promoter are not highly conserved between human and mouse.[22] These observations suggest biologically relevant regulatory functions within the LD block B.

The strongest associations with PD were found for common SNPs and haplotype groups, which were pooled according to their evolutionary relations. Because of the high allele frequencies of the significantly associated markers, the causal variant can be hypothesized to be old. The ancestral haplotype, on which the causal mutation appeared, existed long enough to diversify and produce related haplotypes observed in these haplotype groups. A balancing mechanism of selection could account for the stable existence of the two haplotype groups. By comparison, we found that the ape haplotypes of block B are more similar to the risk-associated than to the protective haplotype group. The risk haplotypes therefore can be suspected to be ancestral to the protective variants.

Aggregation of α-synuclein is thought to be a major mechanism for the pathogenesis of PD.[4] It is conceivable that more than one locus within *SNCA* could influence aggregation behavior, and it has been pointed out that heterogeneity in the set of causal loci may further complicate association studies of complex traits.[14] A specific causal set could depend on the interacting genetic and physiological environment, which may be influenced by sex or age. Indeed, we found an additional association signal specific for female and young subjects. This signal was assigned to a fragmented, but intercorrelated, region. The signal extends from within intron 4 (block C) to the intergenic region between *SNCA* and the *MMRN* gene, which codes for multimerin-1, a protein involved in homeostasis. The allelic frequency spectrum of this group of markers with a minor allele frequency of about 20% is different from the marker group with minor allele frequencies of about 40% (main association region), indicating a different age for these two putative causal variants.

Previous association studies have focused on polymorphisms (sequence repeat Rep1 and several SNPs) of the promoter region of *SNCA* as possible risk factors for PD.[7] Gene expression differences in vitro caused by the repeat polymorphism (Rep1), as well as by SNPs in the promoter region, have been described.[13,23] However, reports of the significance of the promoter region for PD are still conflicting. Both positive[7,10,11,24,25] and negative association results[12,13,26,27] have been published. We also analyzed the three markers (rs2583988, Rep1, and rs1372525) in the 15.3kb promoter region, which Pals and colleagues[11] indicated previously as the most promising region for association, but we could not corroborate the direct significance of this region for the development of PD in our population. This agrees with the results of Holzmann and colleagues,[13] who also investigated a German sample. Notably, most reports of a positive promoter association in *SNCA* were based on multivariate or haplotype analyses and found equally strong association signals in intron 4 polymorphisms or haplotypes.[7,10] This overall pattern fits well with our results of an associated fragmented region from intron 4 (block C) to the 5′-region of block D in a subgroup of patients. The intermediate promoter marker Rep1 is correlated (r^2 = 0.52) with the block D markers in our population. The neighboring marker rs2583988, however, shows some correlation with our main associated SNPs in block B (r^2 = 0.51). The complex correlation structure at the 5′-end of block C and the combination of different association signals in haplotype analyses could explain the stronger haplotype association results compared with single-marker results in most studies. It may also be that the fine-scale correlation structure in this region differs among German, Caucasian, Chinese or Belgian subjects who Farrer and colleagues,[10] Tan and colleagues,[7] or Pals and colleagues,[11] respectively, investigated. We have determined that LD patterns across eight European population samples indicate variability of the boundary strength in blocks B and C.[28] In Italian populations, the intron 4 boundary of block B is shifted by about 10kb to the 5′-end of the *SNCA* gene, and the LD block C is subdivided into two parts, diminishing the correlation between the two associated marker groups of intron 4 (block C) and block D. Association studies in these Southern European populations thus may be well suited to allocate our fragmented association signal either to intron 4 or the 5′-region of *SNCA*. Another potential source for differences in results of association studies is a variable sample composition. We have shown here that differing age or sex structures between samples can be responsible for different outcomes.

The functional relation between the Parkinson phenotype and the detected associated regions of *SNCA* remain unclear. Potential mediatory factors might be enhancer or silencer elements, posttranscriptionally regulated RNA stability, or elements affecting alternative splicing. A splice variant (NACP112) lacking exon 5, which is located within the strongly associated region of LD block B, has been described.[29] Sequence variants in the surrounding of the splice sites could determine the relative expression of the splice variants.[30] A causal variant also could be located in the 3′-untranslated regions. Two SNPs in this region were previously found to alter the expression of the *SNCA* gene in rats.[31] Further experiments will help to elucidate any potential impact of the risk-associated haplotype on alternative splicing, aggregation behavior, and expression level of α-synuclein. Identifying the functional relevance of genetic risk factors can provide powerful tools for an early intervention into the pathogenesis of PD.

This study was supported by the BFAM project from the German Federal Ministry of Education and Research (031U112C, J.C.M.), the German National Genome Network (German Ministry for Education and Research, grant 01GS0116), the Udall Center of Excellence in Parkinson's disease (NIH grant NS40256), and the Competence Network Parkinson's disease (German Ministry for Education and Research, grant 01Gl0201).

We acknowledge the participation of all probands. We thank U. Walter for the supply of samples and J. Favor for reviewing the manuscript.

References

1. Kruger R, Kuhn W, Muller T, et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 1998;18:106–108.
2. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997;276:2045–2047.
3. Zarranz JJ, Alegre J, Gomez-Esteban JC, et al. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol 2004;55:164–173.
4. Lashuel HA, Petre BM, Wall J, et al. Alpha-synuclein, especially the Parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. J Mol Biol 2002;322: 1089–1102.
5. Spillantini MG, Schmidt ML, Lee VM, et al. Alpha-synuclein in Lewy bodies. Nature 1997;388:839–840.
6. Nagar S, Juyal RC, Chaudhary S, et al. Mutations in the alpha-synuclein gene in Parkinson's disease among Indians. Acta Neurol Scand 2001;103:120–124.
7. Tan EK, Chai A, Teo YY, et al. Alpha-synuclein haplotypes implicated in risk of Parkinson's disease. Neurology 2004;62: 128–131.
8. Singleton AB, Farrer M, Johnson J, et al. Alpha-synuclein locus triplication causes Parkinson's disease. Science 2003;302: 841–843.
9. Miller DW, Hague SM, Clarimon J, et al. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. Neurology 2004;62:1835–1838.
10. Farrer M, Maraganore DM, Lockhart P, et al. Alpha-synuclein gene haplotypes are associated with Parkinson's disease. Hum Mol Genet 2001;10:1847–1851.
11. Pals P, Lincoln S, Manning J, et al. Alpha-synuclein promotor confers susceptibility to Parkinson's disease. Ann Neurol 2004; 56:591–595.
12. Izumi Y, Morino H, Oda M, et al. Genetic studies in Parkinson's disease with an alpha-synuclein/NACP gene polymorphism in Japan. Neurosci Lett 2001;300:125–127.
13. Holzmann C, Kruger R, Saecker AM, et al. Polymorphisms of the alpha-synuclein promotor: expression analyses and association studies in Parkinson's disease. J Neural Transm 2003;110: 67–76.
14. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. Am J Hum Genet 2004;75: 353–362.
15. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992;55:181–184.
16. Illig T, Bongardt F, Schöpfer A, et al. The endotoxin receptor TLR 4 polymorphism is not associated with diabetes or components of the metabolic syndrome. Diabetes 2003;52:2861–2864.
17. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296: 2225–2229.

18. Zhang K, Jin L. HaploBlockFinder: haplotype block analyses. Bioinformatics 2003;19:1300–1301.
19. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–989.
20. Zaykin DV, Westfall PH, Young SS, et al. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79–91.
21. The international HapMap consortium. The international HapMap project. Nature 2003;426:789–796.
22. Touchman JW, Dehejia A, Chiba-Falek O, et al. Human and mouse alpha-synuclein genes: comparative genomic sequence analysis and identification of a novel gene regulatory element. Genome Res 2001;11:78–86.
23. Chiba-Falek O, Nussbaum R. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. Hum Mol Genet 2001;10:3101–3109.
24. Tan EK, Khajavi M, Thornby JI, et al. Variability and validity of polymorphism association studies in Parkinson's disease. Neurology 2000;55:533–538.
25. Kruger R, Vieira-Saecker AM, Kuhn W, et al. Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. Ann Neurol 1999; 45:611–617.
26. Parsian A, Racette B, Zhang ZH, et al. Mutation, sequence analysis, and association studies of α-synuclein in Parkinson's disease. Neurology 1998;51:1757–1759.
27. Khan N, Graham E, Dixon P, et al. Parkinson's disease is not associated with the combined alpha-synuclein/apolipoprotein E susceptibility genotype. Ann Neurol 2001;49:665–668.
28. Mueller JC, Lohmussaar E, Mägi R, et al. Linkage disequilibrium patterns and tagSNP transferability among European populations. Am J Hum Genet 2005;76:387–398.
29. Ueda K, Saitoh T, Mori H. Tissue-dependent alternative splicing of mRNA for NACP, the precursor of non-A beta component of Alzheimer's disease amyloid. Biochem Biophys Res Commun 1994;205:1366–1372.
30. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3:285–298.
31. Liang T, Spence J, Liu L, et al. Alpha-synuclein maps to a quantitative trait locus for alcohol preference and is differentially expressed in alcohol-preferring and -nonpreferring rats. Proc Natl Acad Sci U S A 2003;100:4690–4695.
32. Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. In: Proceedings of the Seventh Annual International Conference on Computational Molecular Biology (RECOMB 2003). 2003:277–286.