Genetic and epigenetic study of monozygotic twins affected by Parkinson's disease

Yi-Min Sun, MD [1*], Wanli Yang, MSc [2*], Ekaterina Rogaeva, PhD [3,4], Anthony E. Lang, MD [4-6], Jian Wang, MD [1#], Ming Zhang, PhD [2,7,8#]

Supplemental materials

Supplementary Table S1. Haplotype results for the three twin pairs

STR markers	Twin pair 1		Twin pair 2		Twin pair 3	
	II -2	II-3	II-5	II-6	II-1	II-2
D3S1358	15, 16	15, 16	16, 16	16, 16	15, 17	15, 17
TH01	7, 7	7, 7	6, 9	6, 9	7, 9	7, 9
D21S11	28, 29	28, 29	31, 32	31, 32	31, 31	31, 31
D18S51	14, 15	14, 15	13, 21	13, 21	12, 14	12, 14
Penta E	11, 21	11, 21	10, 12	10, 12	15, 18	15, 18
D5S818	10, 13	10, 13	7, 11	7, 11	10, 12	10, 12
D13S317	8, 11	8, 11	10, 11	10, 11	8, 10	8, 10
D7S820	11, 14	11, 14	9, 13	9, 13	10, 11	10, 11
D16S539	11, 12	11, 12	9, 13	9, 13	9, 9	9, 9
CSF1PO	13, 13	13, 13	12, 12	12, 12	11, 12	11, 12
Penta D	11, 11	11, 11	10, 11	10, 11	9, 9	9, 9
AMEL	X	X	X	X	X	X
vWA	14, 18	14, 18	14, 17	14, 17	16, 17	16, 17
D8S1179	14, 16	14, 16	12, 14	12, 14	13, 14	13, 14
TPOX	8, 9	8, 9	8, 11	8, 11	8, 8	8, 8
FGA	19, 22	19, 22	21, 23	21, 23	22, 25	22, 25
D19S433	13, 15	13, 15	12, 16	12, 16	12, 13	12, 13
D12S391	19, 19	19, 19	18, 18	18, 18	20, 21	20, 21
D6S1043	13, 14	13, 14	12, 14	12, 14	13, 18	13, 18
D2S1338	19, 23	19, 23	20, 23	20, 23	16, 19	16, 19

D1S1656	12, 15	12, 15	12, 17	12, 17	16, 16	16, 16

Supplementary Table S2. List of known PD risk genes.

	Gene name	Location	Inheritance	Confidence as actual PD gene[#]
1	*SNCA*	4q22.1	AD	Very high
2	*LRRK2*	12q12	AD	Very high
3	*VPS35*	16q11.2	AD	Very high
4	*EIF4G1*	3q27.1	AD	Low
5	*CHCHD2*	7p11.2	AD	Low
6	*UCHL1*	4p13	AD	Low
7	*HTRA2*	2p13.1	AD or risk factor	Low
8	*DNAJC13*	3q22.1	AD	Low
9	*LRP10*	14q11.2	AD	Low
10	*GIGYF2*	2q37.1	AD or risk factor	Low
11	*RIC3*	11p15.4	AD	Low
12	*POLG*	15q26.1	AD	High
13	*PSAP*	6p21.2	AD	Low
14	*PRKN*	6q26	AR	Very high
15	*PINK1*	1p36.12	AR	Very high
16	*DNAJC6*	1p31.3	AR	Low
17	*DJ-1*	1p36.23	AR	Very high
18	*ATP13A2*	1p36.13	AR	Very high
19	*FBXO7*	22q12.3	AR	Very high
20	*PLA2G6*	22q13.1	AR	Very high
21	*SYNJ1*	21q22.11	AR	High
22	*VPS13C*	15q22.2	AR	High
23	*RAB39B*	Xq28	X-linked	Low
24	*GBA*	1q22	Risk factor	Very high
25	*GCH1*	14q22.2	Risk factor	Low
26	*SMPD1*	11p15.4	Risk factor	Low
27	*TMEM230*	20p13-p12.3	Risk factor	Low
28	*NUS1*	6q22.1	Risk factor	Low
29	*ARSA*	22q13.33	Risk factor	Low
30	*ATP10B*	5q34	Risk factor	Low

AD: Autosomal dominant; AR: Autosomal recessive; #: According to the rating method in review (Blauwendraat C. et al. 2020), which included the evidence of multiple independent families reported, functional evidence and negative reports published.

Supplementary Table S3. List of primers used to confirm the *PRKN* variants.

Primer name	Sequence (5’ to 3’)
PRKN -F	GAGAAGGCTTCGGGACCC
PRKN -R	ATTTAACCCAGGAGAGCCGC

Supplementary Table S4. Repeat length of 28 pathogenic STRs in three identical twin pairs.

CHROM	POS	REF	Gene	Repeat unit	Repeats risk cutoff	TP in Family A carrying the repeats > risk cutoff	TP in Family B carrying the repeats > risk cutoff	TP in Family C carrying the repeats > risk cutoff
chr1	149390802	A	*NOTCH2NL*	GGC	61	No	No	No
chr2	191745598	C	*GLS*	GCA	680	No	No	No
chr3	63898360	G	*ATXN7*	GCA	37	No	No	No
chr3	128891419	A	*CNBP*	CAGG	50	No	No	No
chr4	3076603	C	*HTT*	CAG	36	No	No	No
chr4	39350044	A	*RFC1*	AARRG	400	No	No	No
chr4	41747989	A	*PHOX2B*	GCN	25	No	No	No
chr5	146258290	A	*PPP2R2B*	GCT	43	No	No	No
chr6	16327864	G	*ATXN1*	TGC	39	No	No	No
chr6	170870994	G	*TBP*	GCA	43	No	No	No
chr9	27573526	C	*C9ORF72*	GGCCCC	30	No	No	No
chr9	71652202	A	*FXN*	GAA	66	No	No	No
chr12	7045879	C	*ATN1*	CAG	48	No	No	No
chr12	50898784	T	*DIP2B*	GGC	75	No	No	No
chr12	112036753	G	*ATXN2*	GCT	32	No	No	No
chr13	70713515	A	*ATXN8OS*	CTG	74	No	No	No
chr14	23790681	G	*PABPN1*	GCG	12	No	No	No
chr14	92537353	C	*ATXN3*	GCT	55	No	No	No
chr16	87637893	C	*JPH3*	CTG	41	No	No	No
chr18	53253386	G	*TCF4*	CAG	50	No	No	No
chr19	13318672	C	*CACNA1A*	CTG	20	No	No	No
chr19	46273462	C	*DMPK*	CAG	50	No	No	No
chr20	2633379	G	*NOP56*	GGCCTG	650	No	No	No
chr21	45196324	G	*CSTB*	CGCGGGGCGGGG	30	No	No	No
chr22	46191234	G	*ATXN10*	ATTCT	280	No	No	No
chrX	66765158	T	*AR*	GCA	38	No	No	No
chrX	146993568	G	*FMR1*	CGG	200	No	No	No
chrX	147582151	A	*AFF2*	GCC	200	No	No	No

Supplementary Table S5. SNVs of 113 Parkinson's disease related genes in three MZ twin pairs. The variants were filtered based on MAF<0.05 and CADD>10.

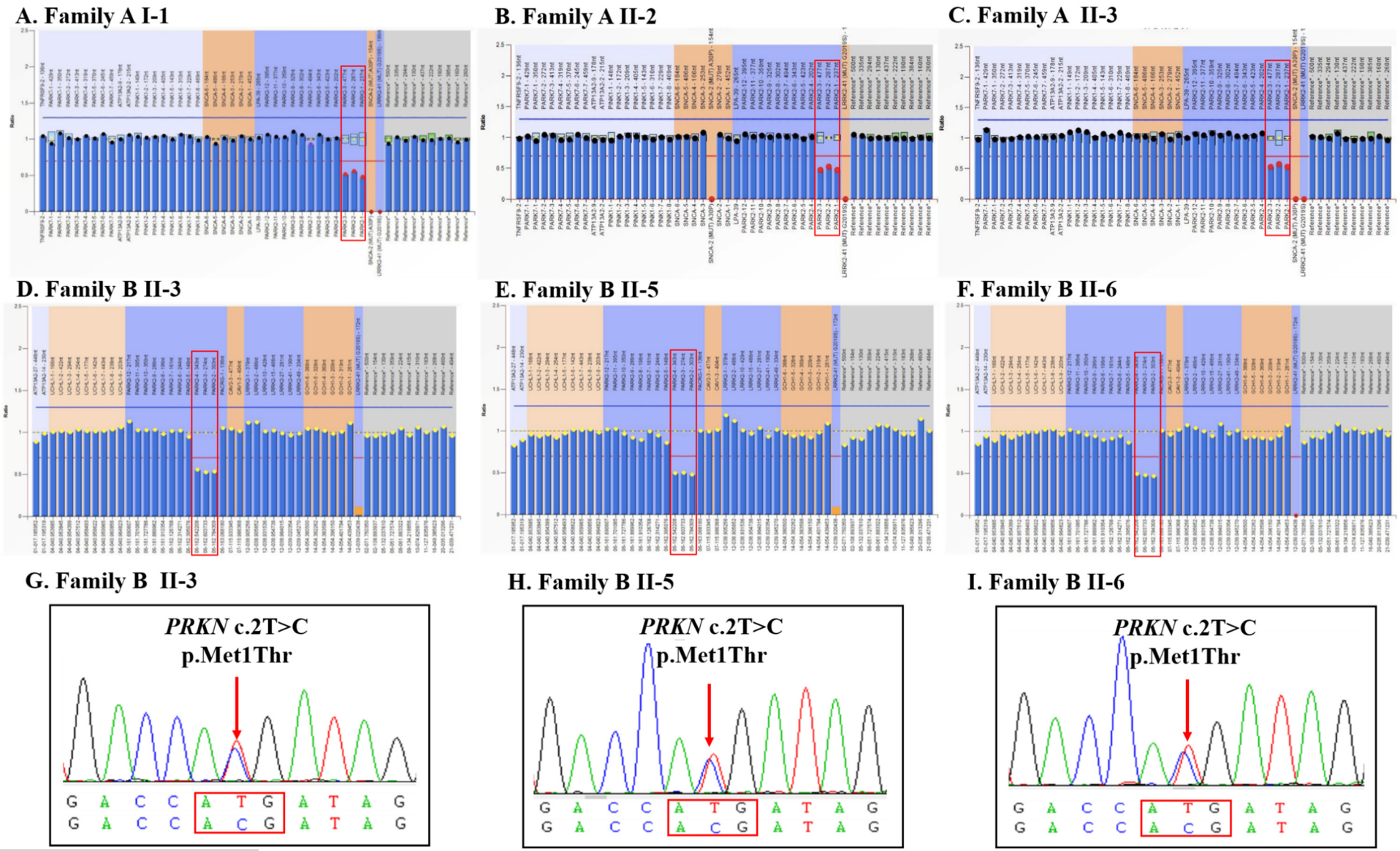
Gene	Genomic Position (hg19)	Nucleotide change	AA change	MAF		Functional prediction				Family A		Family B		Family C	
				gnomAD_EAS	ChinaMAP	CADD	SIFT	PolyPhen2	MutationTaster	II-2	II-3	II-5	II-6	II-1	II-2
ITPKB	chr1:226924241	c.919G>T	p.Ala307Ser	0.01299	0.014592	14.18	D	B	N	0/0	0/0	0/1	0/1	0/0	0/0
MED12L	chr3:151105707	c.5093G>A	p.Arg1698Gln	0.01454	0.0160559	28.9	D	D	D	0/1	0/1	0/0	0/0	0/0	0/0
GAK	chr4:861215	c.2401A>G	p.Lys801Glu	0.001852	0.000802796	10.17	D	B	N	0/0	0/0	0/0	0/0	0/1	0/1
BST1	chr4:15709192	c.374G>A	p.Arg125His	0.03433	0.0316396	23.2	T	D	P	0/1	0/1	0/0	0/0	0/0	0/0
PRKN	chr6:163148699	c.2T>C	p.Met1Thr	0.0001420	0.000330657	27.1	D	D	D	0/0	0/0	0/1	0/1	0/0	0/0
CTSB	chr8:11702653	c.1001A>C	p.Gln334Pro	0.03268	0.0387231	23.1	D	B	D	0/1	0/1	0/0	0/0	0/0	0/0
SEC23IP	chr10:121685689	c.2263G>A	p.Val755Met	0.02130	0.0259256	16.3	T	B	D	0/0	0/0	0/0	0/0	0/1	0/1
SMPD1	chr11:6413290	c.995C>G	p.Pro332Arg	0.005263	0.00481677	26	D	P	D	0/0	0/0	0/0	0/0	0/1	0/1
CHD9	chr16:53242622	c.1453-772G>T	-	NA	4.72233e-05	NA	NA	NA	NA	0/0	0/0	0/0	0/0	0/1	0/1
CHD9	chr16:53337907	c.5989G>A	p.Ala1997Thr	0.003171	0.00401398	24.4	D	B	D	0/0	0/0	0/0	0/0	0/1	0/1
DNAH17	chr17:76567034	c.914C>T	p.Thr305Met	0.003137	0.0020306	19.13	D	P	N	0/0	0/0	0/1	0/1	0/0	0/0
DDRGK1	chr20:3180666	c.491G>T	p.Arg164Leu	0.003064	0.00288062	27.8	D	D	D	0/0	0/0	0/0	0/0	0/1	0/1

AA:amino acids; MAF: minor allele frequency; Mutation assessment by SIFT and PolyPhen2: D: damaging; T:tolerated; P:possibly damaging; B:benign; Mutation assessment by MutationTaster: D: Disease causing; N and P: Polymorphism; NA: not available; 0/1: means the genotype of the variant is heterozygous.

Supplementary Table S6. Mutect2 analysis suggested rare somatic SNVs in family A and B. The variants were filtered based on MAF<0.001, AD >10 and CADD>10.

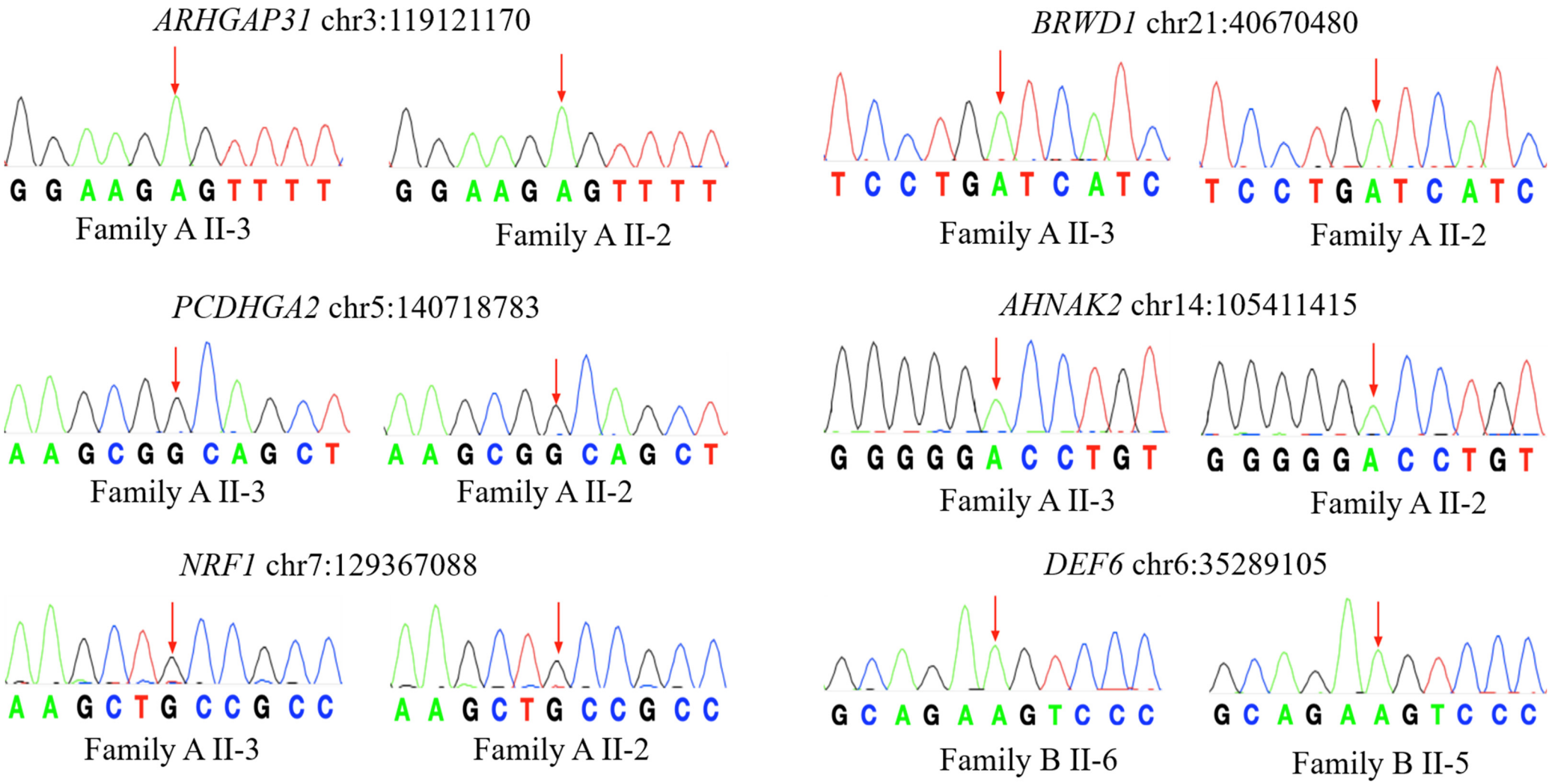
Gene	Genomic Position (hg19)	Nucleotide change	AA change	Genotype	MAF		Functional prediction				Subject ID
					gnomAD_EAS	ChinaMAP	CADD	SIFT	PolyPhen2	MutationTaster	
ARHGAP31	chr3:119121170	c.A1571G	p.E524G	Het	NA	NA	24.4	D	B	D	FamilyA II-3
PCDHGA2	chr5:140718783	c.G245C	p.G82A	Het	NA	NA	24.5	D	D	D	FamilyA II-3
NRF1	chr7:129367088	c.G1231C	p.A411P	Het	NA	NA	28.8	D	D	D	FamilyA II-3
AHNAK2	chr14:105411415	c.A10373G	p.D3458G	Het	NA	NA	21.1	D	P	N	FamilyA II-3
BRWD1	chr21:40670480	c.A227G	p.D76G	Het	NA	NA	23.3	D	D	D	FamilyA II-3
DEF6	chr6:35289105	c.A1814G	p.K605R	Het	NA	NA	18.3	T	B	N	FamilyB II-5

MAF: minor allele frequency; AD: allele depth; AA: amino acids; Het: heterozygous; Mutation assessment by SIFT and PolyPhen2: D: damaging; T: tolerated; P: possibly damaging; B: benign; Mutation assessment by MutationTaster: D: Disease causing; N: Polymorphism;

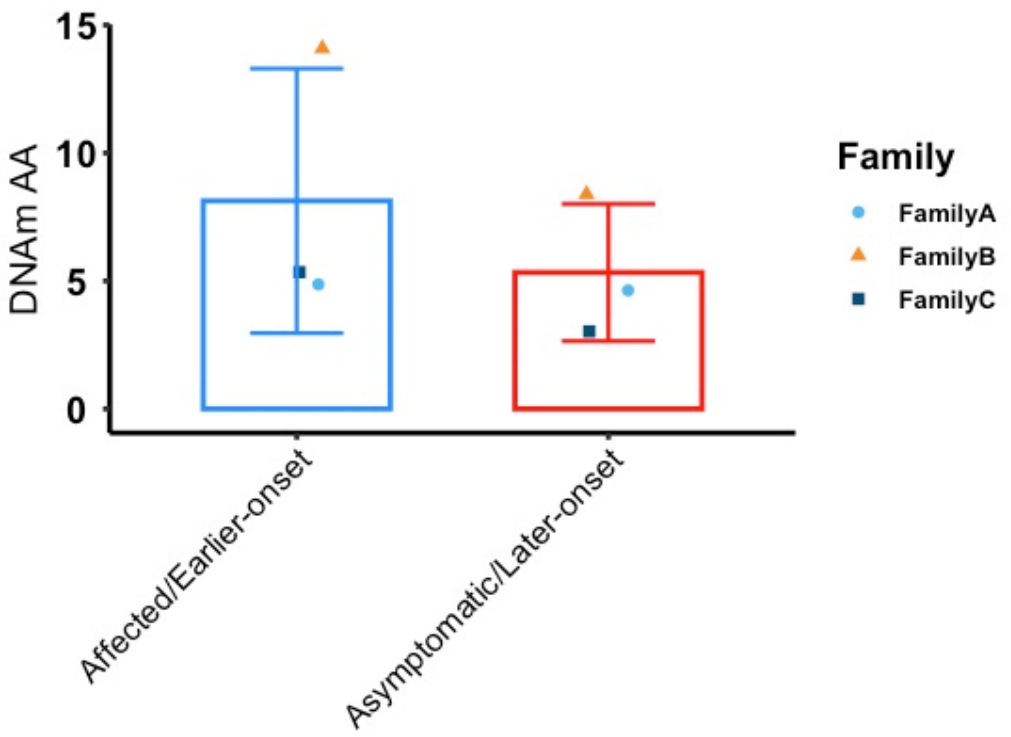


Supplementary Figure S1. Genetic findings for the MZ twins and their family members. A-C. Heterozygous *PRKN* exon 1-3 deletions in case I-1, II-2 and II-3 from family A; D-F. multiplex ligation-dependent probe amplification results of the heterozygous *PRKN* exon 2-4 deletion in case II-3, II-5 and II-6 from family B. The red

boxes indicate that the normalization ratios for the deletion were below the threshold of 0.65, which represent a heterozygous deletion. G-I. The Sanger sequencing results of *PRKN* p.Met1Thr in II-3, II-5, II-6 from family B.



Supplementary Figure S2. Sanger sequencing did not validate somatic mutations in family A and B.



Supplementary Figure S3. Scatter plot of the DNAm age acceleration. The affected/earlier onset twins had a trend of higher DNAm AA when compared to the asymptomatic/later onset twins (mean of DNAm AA: 8.13 vs 5.3 year; SD of DNAm AA: 5.1 vs 2.6 year).