Supplemental Material for the article:

Excess of singleton loss-of-function variants in Parkinson's Disease contributes to genetic risk

Dheeraj R. Bobbili, Peter Banda, Rejko Krüger, Patrick May

Content **Page**

Number of cases	340
Number of controls	146
Number of variants	459391
Number of exonic/splicing variants	218987
Ti/Tv ratio of exonic/splicing variants	3,07

Supplementary Table 1: Numbers for the PPMI dataset after QC.

SNP	Candidate gene	A1	A2	P-value	OR

1:155135036	GBA	G	A	2.59e-35	0.58
3:52816840	ALAS1,TLR9,DNAH1,BAP1,PHF7,NISCH,STAB1,ITIH3,ITIH4	G	A	2.25e-7	0.68
17:43994648	ARHGAP27,CRHR1,SPPL2C,MAPT,STH,KANSL1	T	C	1.26e-68	0.78
2:169110394	STK39	C	T	5.68e-26	0.83
3:182762437	MCCC1	A	G	2.11e-30	0.85
6:32666660	HLA-DRB6,HLA-DQA1	T	C	1.26e-13	0.85
1:205723572	NUCKS1,SLC41A1	C	T	1.12e-2	0.89
2:135539967	TMEM163,CCNT2	T	C	8.24e-24	0.89
4:15737101	FAM200B,CD38	C	A	1.22e-19	0.90
12:123303586	OGFOD2	G	A	2.05e-20	0.90
7:23293746	KLHL7,NUPL2,GPNMB	G	A	3.51e-18	0.91
8:16697091	MICU3	A	G	2.38e-11	0.91
9:17579690	SH3GL2	T	G	1.99e-12	0.91
14:55348869	GCH1	T	C	4.30e-16	0.91
15:61994134	VPS13C	G	A	3.94e-14	0.91
1:226916078	ITPKB	C	T	2.40e-10	0.92
4:77198986	FAM47E	T	C	1.43e-14	0.92
3:48748989	NCKIPSD,CDC71	G	T	6.80e-8	0.93
10:15569598	FAM171A1	C	A	2.37e-8	0.93
11:83544472	DLG2	A	G	3.72e-9	0.93
2:166133632	SCN3A	T	C	9.73e-7	0.94
8:22525980	SORBS3,PDLIM2,C8orf58,BIN3	T	C	9.06e-7	1.06
2:102413116	IL1R2	C	T	3.83e-8	1.07
16:19279464	COQ7	T	G	1.46e-9	1.07
20:3168166	DDRGK1	A	G	1.99e-6	1.07
14:88472612	GALC	T	C	1.20e-9	1.08
16:31121793	ZNF646,KAT8	A	G	5.44e-12	1.08
16:52599188	TOX3	T	C	8.29e-8	1.08
19:2363319	LSM7	T	C	6.64e-7	1.08
11:133765367	MIR4697	T	C	1.11e-13	1.09
14:67984370	TMEM229B	T	A	9.61e-11	1.09
8:11707174	CTSB	A	G	9.54e-11	1.10
18:40673380	SYT4	G	A	5.56e-16	1.10
3:18277488	SATB1	G	T	3.02e-9	1.11
1:232664611	SIPA1L2	T	C	8.41e-13	1.12
6:27681215	ZNF184	A	G	3.44e-13	1.12
4:114360372	ANK2,CAMK2D	C	T	2.11e-9	1.14
5:60273923	ELOVL7	C	A	1.69e-11	1.15
12:40614434	LRRK2	T	C	1.21e-19	1.15

3:87520857	CHMP2B	C	G	1.22e-4	1.21
4:951947	TMEM175,DGKQ	C	T	1.47e-50	1.23
4:90626111	SNCA	G	A	5.21e-123	1.33
10:121536327	BAG3	A	G	2.23e-19	1.65

Supplementary Table 2: Summary statistics of SNVs used to generate PRS. The variants are represented according to the GRCh37 human reference genome. SNP = chromosome and position of SNP on the genome, Candidate gene=Nearest gene/locus, A1 = reference allele, A2 = alternate allele, OR = odds ratio and P-value.

Features	Controls (340)	Cases (146)	F-statistics	ANOVA

				Chi-sq P-value
Clinical features				
UPDRS Total Score	4.372 (4.141)	32.317 (12.762)	666.62	6.361 E-94
UPSIT Raw Score	34.417 (4.442)	22.247 (8.255)	280.91	2.421 E-50
Symbol Digit Modalities Total Correct	47.424 (10.94)	41.3(9.73)	38.24	1.302 E-9
S Anxiety	27.452 (7.582)	32.920 (10.086)	36.45	3.069 E-9
REM Sleep Behavior Score	2.856 (2.240)	4.238 (2.691)	32.04	2.551 E-8
SCOPA-AUT Total Autonomic	7.910 (6.897)	12.117 (8.682)	30.27	6.014 E-8
MoCA Total Score	28.246 (1.123)	27.152 (2.286)	29.52	8.673 E-8
T Anxiety	28.541 (6.822)	32.230 (9.368)	19.02	1.571 E-5
Total Semantic Fluency Score	51.883 (11.056)	48.7 (11.535)	7.88	5.182 E-3
Benton Summary Score	13.184 (1.922)	12.858 (2.119)	2.38	1.234 E-1
ESS Score	5.655 (3.472)	5.897 (3.522)	1.15	2.837 E-1
QUIP Score	0.267 (0.717)	0.280 (0.625)	0.13	7.157 E-1
Non-clinical features				
Gender	96 (68.57)	229 (67.352)	0.13	7.193 E-1
Age	61.184 (10.294)	62.032 (9.533)	0.89	3.472 E-1
PRS	0.0911 (0.0066)	0.092 (0.0078)	6.41	1.168 E-2
Singleton Count	8.315 (3.041)	9.561 (3.544)	10.62	1.199 E-3
PRS_LRRK2	-0.0146 (0.00686)	-0.0118 (0.0066)	18.30	2.268 E-5
PD Family History	7 (4.794)	86 (25.294)	27.52	1.553 E-7

Supplementary Table 3: Summary statistics and predictive ability of various clinical and non-clinical scores and features available from the PPMI consortium and the genetic features generated in this study. For independence/significance testing we applied ANOVA for continuous data and Chi-square for binary data. The values in brackets indicate standard deviation values unless stated otherwise. A description of features can be found in Supplemental Table 4.

Montreal Cognitive Assessment (MoCA)	MoCA is a measure to detect mild cognitive impairment (MCI). It has a duration of 10-minutes and is graded based on 30-points. MoCA comprises of various examination categories such as short-term memory recall (5 points), visuospatial ability (4 points), executive function (4 points), attention and working memory (6 points), language (5 points), and orientation to time and place (6 points). The general cut-off applied for detecting MCI is 26 out of 30 (1).
Benton judgement of line orientation	Benton score includes a 30-item task in order to assess the ability to discriminate the direction in which the lines are presented (2). The choice of response comprised of a series of 11 lines that are each separated by an angle of 18 degrees (3). Every stimulus comprises of two lines which represent either the proximal, middle, or distal half of a response-choice line. Based on the number of correct responses the performance is scored.
Semantic fluency	Semantic fluency score is used to judge semantic memory. First, an individual is instructed to name as many items as possible from a given category in the fixed time (one minute per category). The score is generated based on the number of names recalled by the subject (4). The circumlocutions and repetitions are excluded and the scores could range from 0 to 20.
Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire (SCOPA-AUT)	To assess the burden and frequency of autonomic dysfunction in PD, the score generated based on SCOPA-AUT can be employed (5). It is an easily self-administered questionnaire containing 25 questions and it generally lasts about 10 minutes. It comprises of questions from various domains including gastrointestinal, urinary, cardiovascular, thermoregulatory, pupillomotor, skin, respiratory, and sexual. All the domains correlate with the severity of the disease except sexual dysfunction.
Rapid eye movement (REM) sleep behavior score	The REM sleep behavior score is used to diagnose REM sleep behavior disorder (RBD). It is generated using a questionnaire which comprises of a validated 10-items and the patient has to self-rate the questionnaire which covers the clinical features of RBD (6). The score has a maximum total score of 13 points and a cut-off of 5 is generally considered to be suggestive of RBD (7).
Anxiety scores	In order to measure the anxiety, a combination of two subscales is employed for both adults and children. The two sub-scales are S-anxiety and T-anxiety (8). Together, both the scales have 40 items in total of which 20 items are allocated to each of the subscales and they are described below.

	S-Anxiety: State Anxiety Scale (S-Anxiety) assesses the present state of anxiety. It is generated by enquiring the subject about how he/she is feeling at the time of enquiry. To measure this score various features are used that represent the subjective feelings of apprehension, tension, nervousness, worry, and activation/arousal of the autonomic nervous system. **T-Anxiety:** The Trait Anxiety Scale (T-Anxiety), on the other hand checks the conditions that relatively stable such as general states of calmness, security, and confidence.
Symbol Digit Modalities Test (SDMT)	The SDMT is employed as a measure to detect neurological dysfunction. For this test, subjects were required to use a coded key to match nine abstract symbols paired with numerical digits (9). The total duration of the test is 90s and the final score is the based on the correct number of substitutions in 90s. The scores could range from 0 and 110.
Movement disorder society-Unified Parkinson's disease rating scale (MDS-UPDRS)	The MDS-UPDRS is a modified version of the UPDRS (10). It is measured based on the assessment of 50 questions with regard to both the motor and non-motor symptoms associated with PD. MDS-UPDRS is measured in four parts: Part I (non-motor experiences of daily living), Part II (motor experiences of daily living), Part III (motor examination) and Part IV (motor complications).
University of Pennsylvania Smell Identification Test (UPSIT) score	The University of Pennsylvania Smell Identification Test (UPSIT) score measures the olfactory functioning of the subject and has become a 'gold standard' in this setting. The UPSIT score is measured based on the response from forty different smells that is released by scratching a panel of microencapsulated odorants using a pencil lead (11). The responses are scored and an aggregated score is generated at the end of the test.

Supplementary Table 4: Clinical and non-clinical scores available for the PPMI dataset.

Variant	Allele	OR	lower	upper	#	#	P_{glm}	adj	P_{emp}	adj

class	frequency		CI	CI	cases	controls		P_{glm}		P_{emp}
LoF	**rare**	1.003	0.989	1.017	88.826	87.308	0.758	0.908	0.757	0.796
CADD20	**rare**	0.999	0.99	1.01	140.435	140.103	0.631	0.908	0.627	0.796
NONSYN	**rare**	0.999	0.994	1.004	303.506	303.226	0.801	0.908	0.796	0.796
SYN	**rare**	0.999	0.992	1.006	199.526	199.288	0.964	0.161	0.967	0.186
LoF	**singleton**	1.058	1.013	1.106	18.688	17.164	0.013	**0.037**	0.014	**0.041**
CADD20	**singleton**	1.015	0.998	1.032	75.282	71.993	0.154	0.086	0.151	0.15
NONSYN	**singleton**	1.01	1.001	1.02	135.897	128.774	0.083	0.063	0.084	0.126
SYN	**singleton**	1.009	0.997	1.022	78.588	74.342	0.185	0.742	0.186	0.967

Supplementary Table 5: Overview about burden analysis for rare and singleton variants for four different variant classes (Loss-of-function: LoF; missense variants: NONSYN, exonic variants with CADD score ≥ 20, synonymous variants: SYN). Significant P-values are highlighted in red, significant P-values after adjustment in bold red. OR=Odds Ration, CI=confidence interval, P_{glm}=P-value linear model, P_{emp}=empirical P-value, adj= adjusted. # is the mean number of qualifying variants in either cases or controls.

LoF type	OR	lower CI	upper CI	# case	# control	P_{glm}	adj P_{glm}	P_{emp}	adj P_{emp}
frameshift deletion	1.191	1.06	1.344	3.15	2.637	0.002	**0.02**	0.003	**0.014**
frameshift insertion	0.974	0.83	1.146	1.344	1.39	0.63	0.81	0.631	0.789
splicing	1.027	0.972	1.086	11.097	10.5	0.239	0.592	0.239	0.399
stopgain	1.141	1.023	1.279	3.459	2.973	0.013	0.05	0.014	**0.035**
stoploss	0.924	0.497	1.826	0.097	0.089	0.825	0.81	0.819	0.819

Supplementary Table 6: Overview about burden analysis for different singleton Loss-of-function (LoF) variant types. Significant P-values are highlighted in red, significant P-values after adjustment in bold red. OR=Odds Ration, CI=confidence interval, P_{glm}=P-value linear model, P_{emp}=empirical P-value, adj= adjusted. # is the mean number of qualifying variants in either cases or controls.

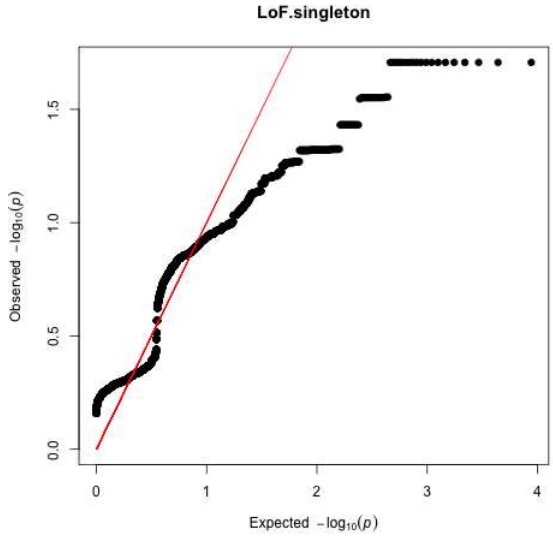
Supplementary Table 7: Number of qualifying singleton variants per gene for the different variant classes (Lof, missense, CADD20, synonymous) in cases vs controls.

File: SupplementalTable7.Variants_per_gene_case_control_counts.xlsx

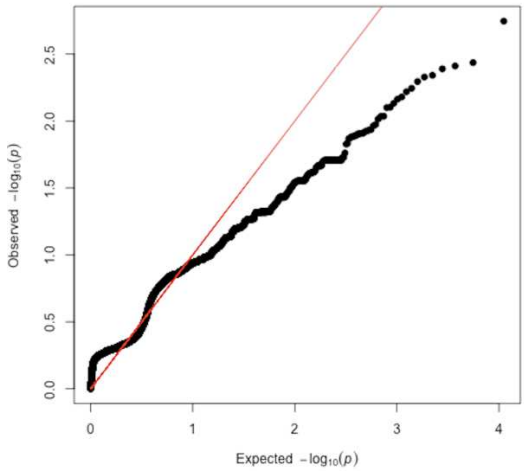
Supplementary Table 8: Number of qualifying singleton variants per individual per variant class (Lof, missense, CADD20, synonymous).

File: SupplementalTable8.Variant_counts_per_individual.xlsx

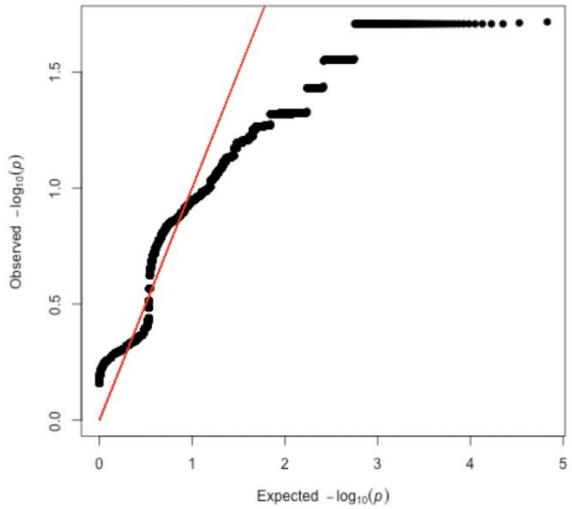
Supplementary Figure 1: Population stratification. Top: Ethnicity of samples in the PPMI study. The PPMI samples were represented along with samples from the 1000g study. Each color represents different ethnicities and each shape represents the 1000g superpopulation to which the samples belong to. The abbreviations of the legend are given below. ASW: Americans of African Ancestry in SW USA, CEU, CHB: Han Chinese in Beijing, China, CHS: Southern Han Chinese, FIN: Finnish in Finland, GBR: British in England and Scotland, JPT: Japanese in Tokyo, Japan, LWK: Luhya in Webuye, Kenya, MXL: Mexican Ancestry from Los Angeles, PUR: Puerto Ricans from Puerto Rico, TSI: Toscani in Italia, YRI: Yoruba in Ibadan, Nigeria. AFR: African, AMR: Ad Mixed American, EAS: East Asian, EUR: European. Bottom: PPMI samples included in the analyses after final QC.



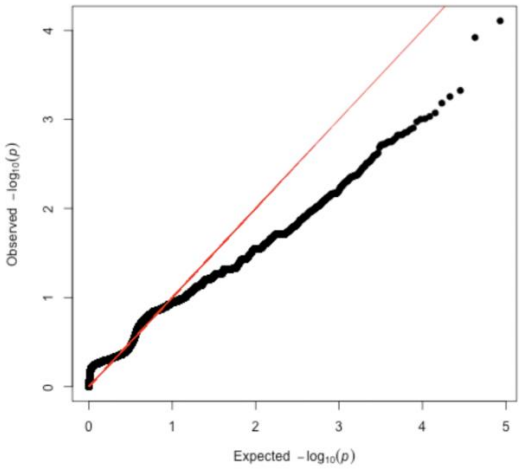
Supplementary Figure 2: QQ-plot of LoF singleton variants. The P-values are generated by using the score method of rvtests package.



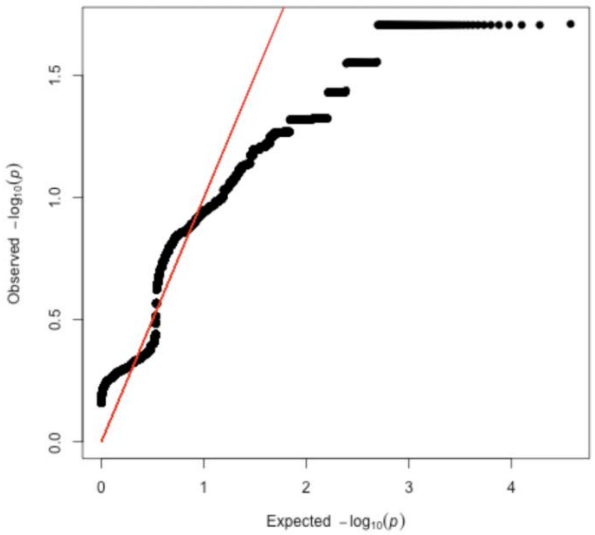
Supplementary Figure 3: QQ-plot of rare LoF variants. The P-values are generated by using the score method of rvtests package.



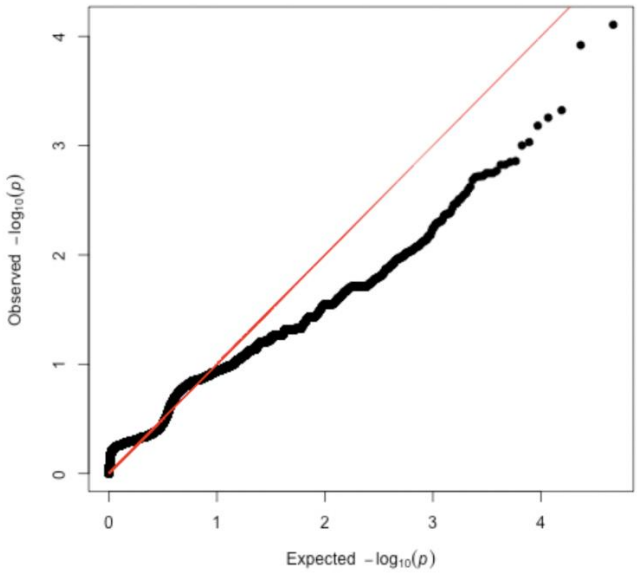
Supplementary Figure 4: QQ-plot of NONSYN singleton variants. The P-values are generated by using the score method of rvtests package.



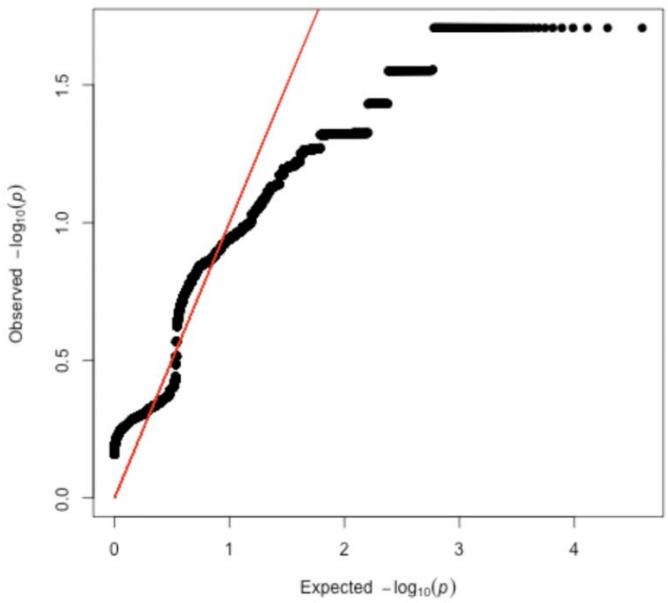
Supplementary Figure 5: QQ-plot of rare NOSYN variants. The P-values are generated by using the score method of rvtests package.



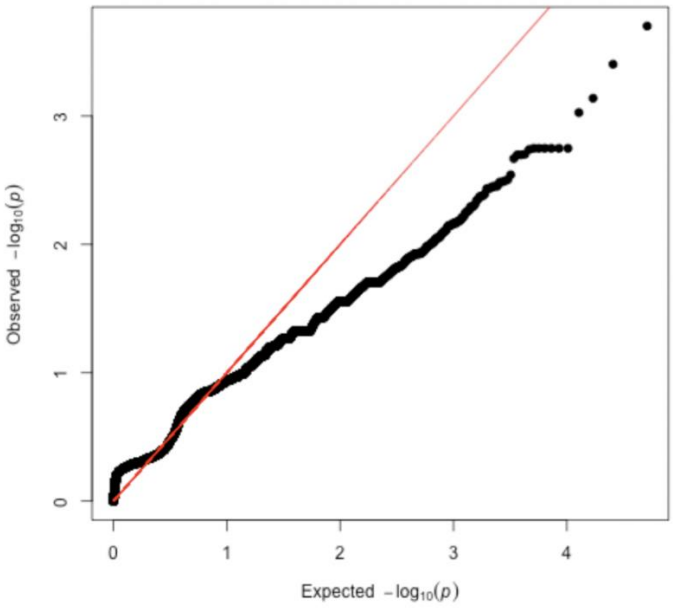
Supplementary Figure 6: QQ-plot of CADD20 singleton variants. The P-values are generated by using the score method of rvtests package.



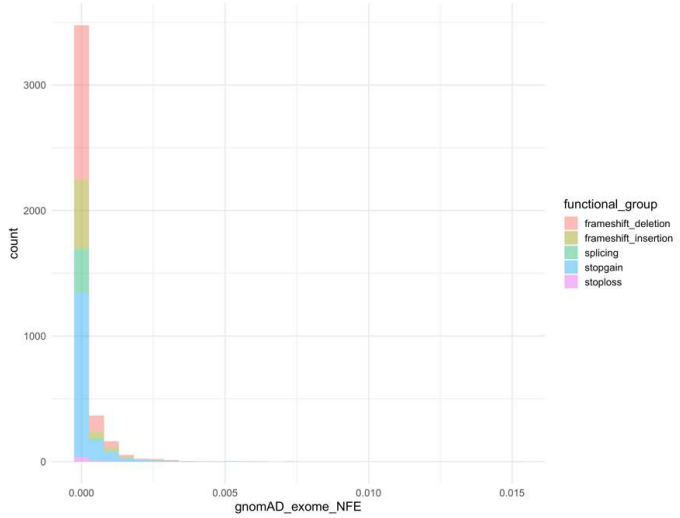
Supplementary Figure 7: QQ-plot of rare CADD20 variants. The P-values are generated by using the score method of rvtests package.



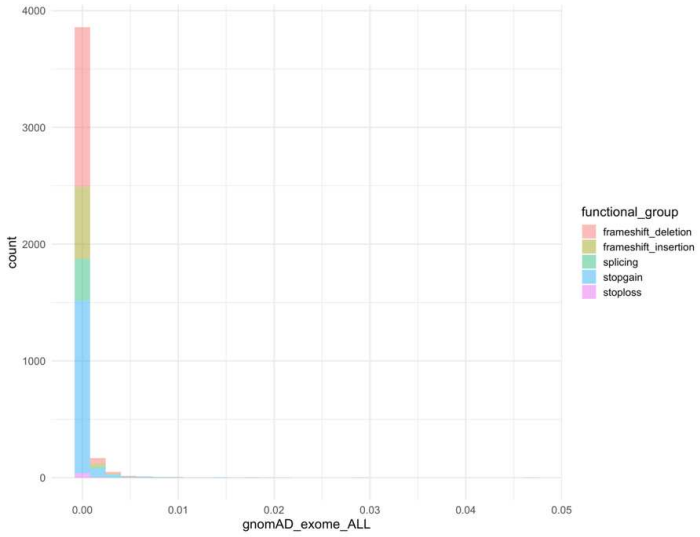
Supplementary Figure 8: QQ-plot of SYN singleton variants. The P-values are generated by using the score method of rvtests package.



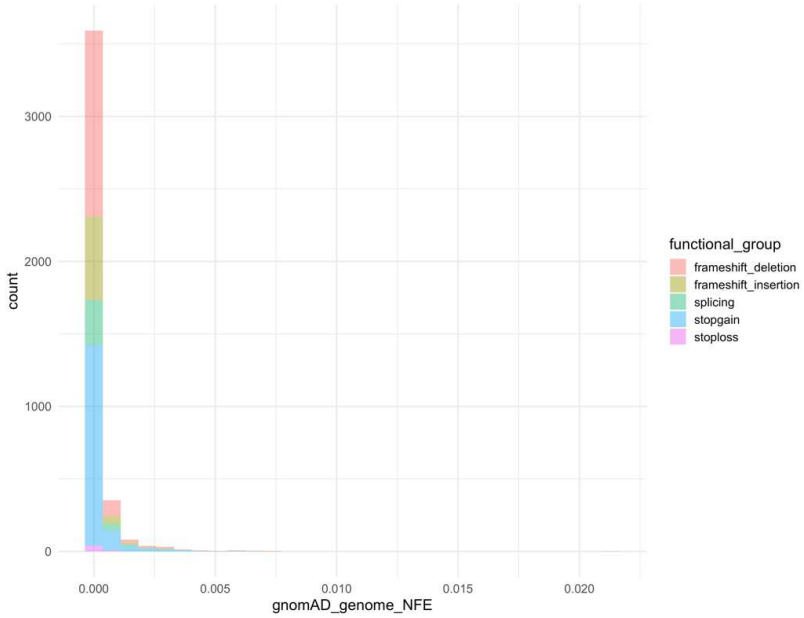
Supplementary Figure 9: QQ-plot of rare SYN variants. The P-values are generated by using the score method of rvtests package.



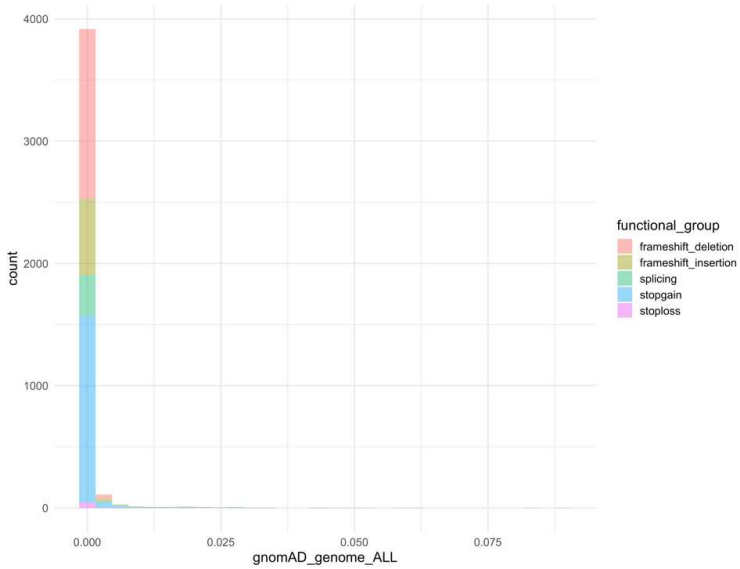
Supplementary Figure 10: Plot showing the observed allele frequencies of singleton LoF variants in the exome data of Non-Finnish European (NFE) population in the gnomAD database. They are separated per sub-functional group.



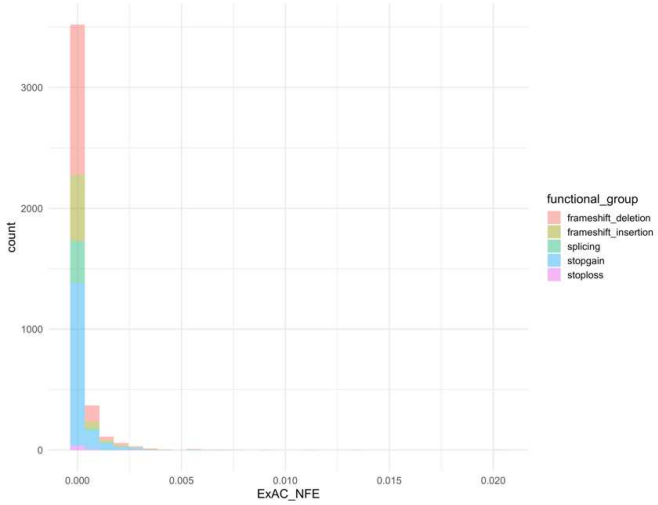
Supplementary Figure 11: Plot showing the observed allele frequencies of singleton LoF variants in exome data of all populations (ALL) in the gnomAD database. They are separated per sub-functional group.



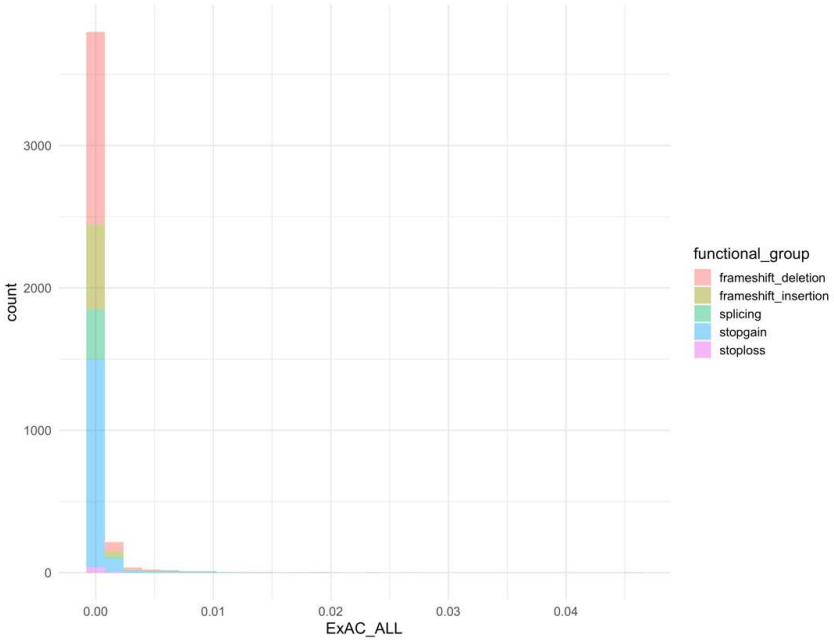
Supplementary Figure 12: Plot showing the observed allele frequencies of singleton LoF variants in whole genome data of the Non-Finnish European (NFE) population of the gnomAD database. They are separated per sub-functional group.



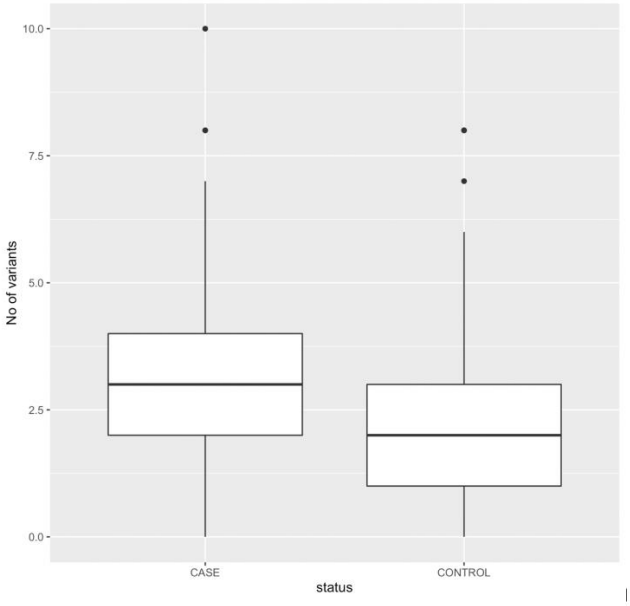
Supplementary Figure 13: Plot showing the observed allele frequencies of singleton LoF variants in whole genome data of all populations (ALL) of the gnomAD database. They are separated per sub-functional group.



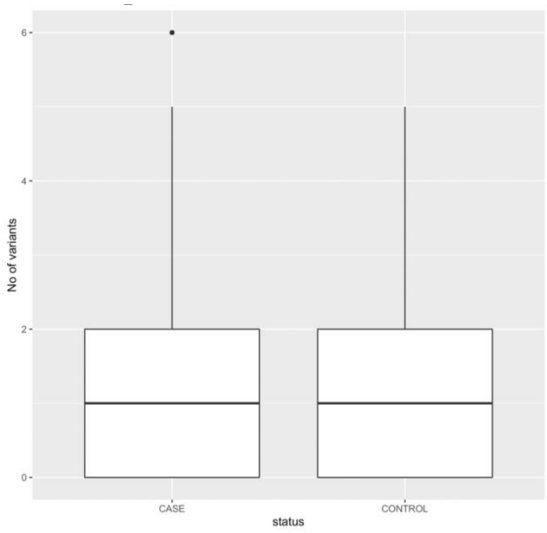
Supplementary Figure 14: Plot showing the observed allele frequencies of singleton LoF variants in exome data of the Non-Finnish European (NFE) population in the ExAC database. They are separated per sub-functional group.



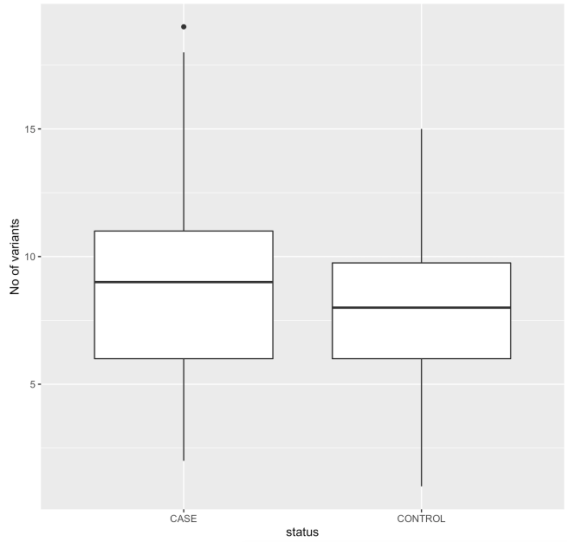
Supplementary Figure 15: Plot showing the observed allele frequencies of singleton LoF variants in exome data of all populations (ALL) in the ExAC database. They are separated per sub-functional group.



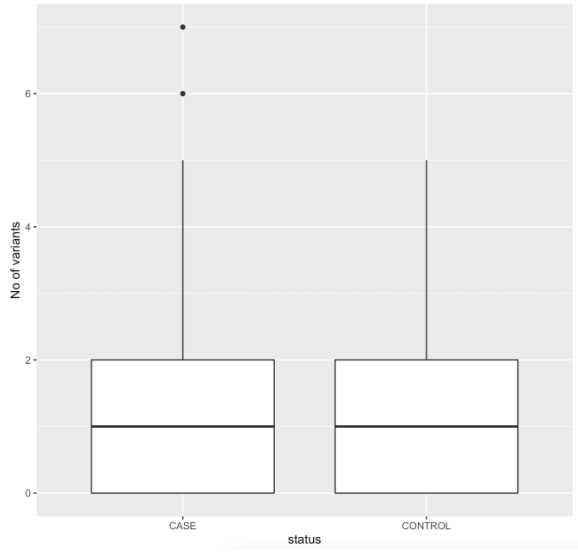
Supplementary Figure 16: Box plot showing the number of LoF frameshift deletion singleton variants in cases versus controls.



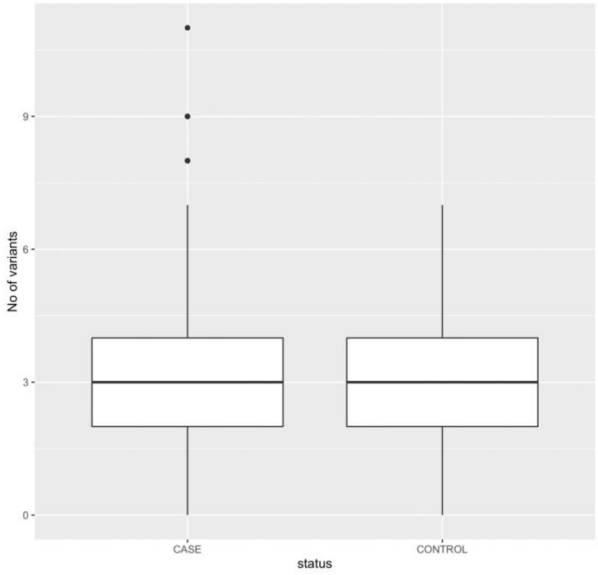
Supplementary Figure 17: Box plot showing the number of LoF frameshift insertion singleton variants in cases versus controls.



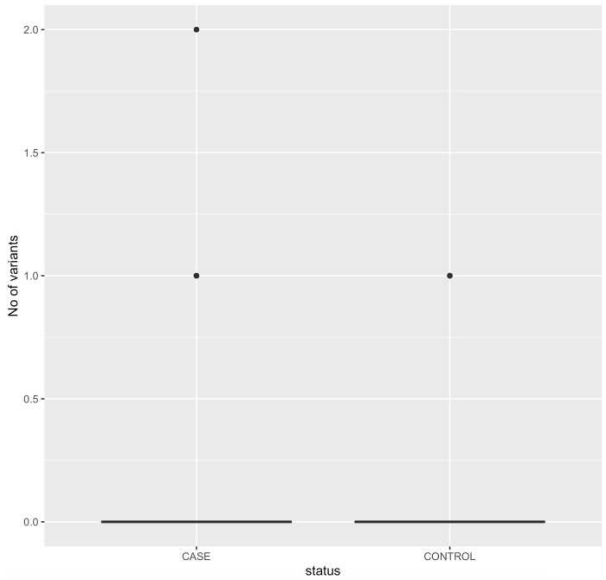
Supplementary Figure 18: Box plot showing the number of LoF singleton variants in cases versus controls.



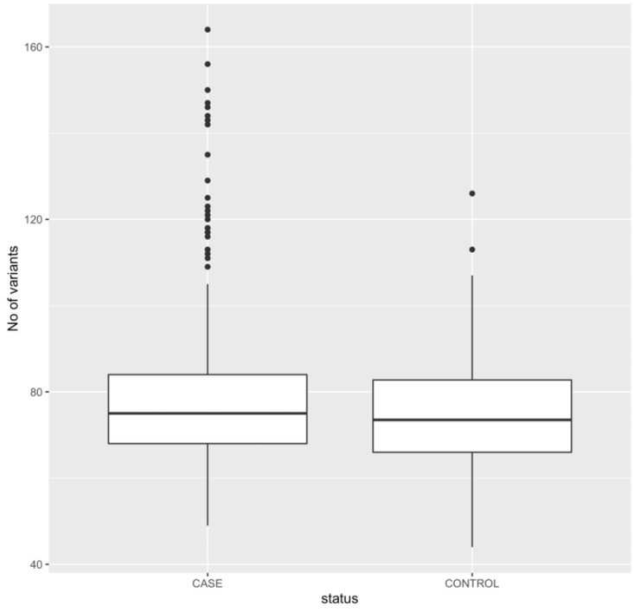
Supplementary Figure 19: Box plot showing the number of LoF splicing singleton variants in cases versus controls.



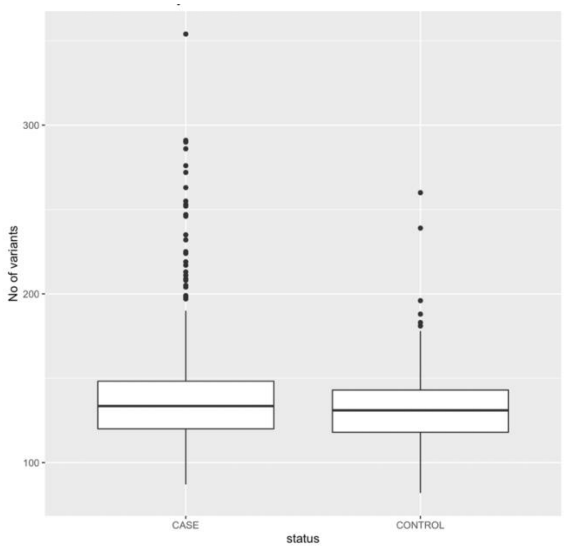
Supplementary Figure 20: Box plot showing the number of LoF stopgain singleton variants in cases versus controls.



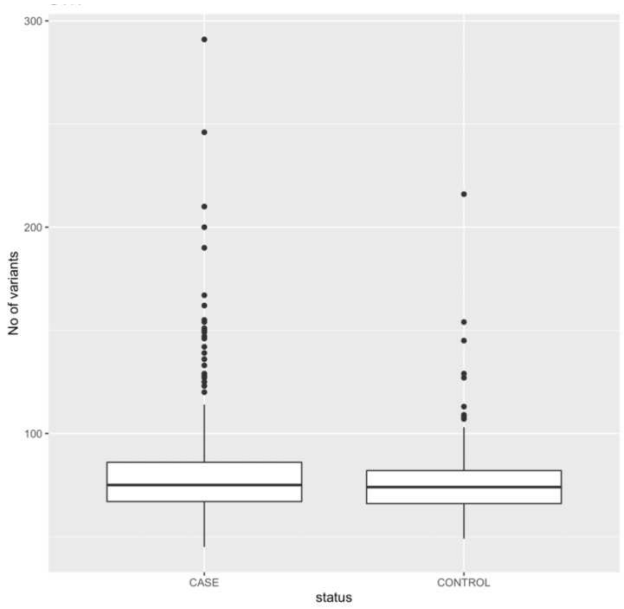
Supplementary Figure 21: Box plot showing the number of LoF stoploss singleton variants in cases versus controls.



Supplementary Figure 22: Box plot showing the number of CADD20 singleton variants in cases versus controls.



Supplementary Figure 23: Box plot showing the number of NONSYN singleton variants in cases versus controls.



Supplementary Figure 24: Box plot showing the number of SYN singleton variants in cases versus controls.

References

1. Borland E, Nägga K, Nilsson PM, Minthon L, Nilsson ED, Palmqvist S. The Montreal Cognitive Assessment: Normative Data from a Large Swedish Population-Based Cohort. J Alzheimers Dis. 59(3):893–901.
2. Benton AL, Varney NR, Hamsher KD. Visuospatial judgment. A clinical test. Arch Neurol. 1978 Jun;35(6):364–7.
3. Calamia M, Markon K, Denburg NL, Tranel D. Developing a Short Form of Benton's Judgment of Line Orientation Test: An Item Response Theory Approach. Clin Neuropsychol. 2011 May;25(4):670–84.
4. Araujo NB de, Barca ML, Engedal K, Coutinho ESF, Deslandes AC, Laks J. Verbal fluency in Alzheimer's disease, Parkinson's disease, and major depression. Clinics. 2011;66(4):623–7.
5. Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. Mov Disord. 2004;19(11):1306–1312.
6. Högl B, Stefani A. REM sleep behavior disorder (RBD). Somnologie. 2017;21(Suppl 1):1–8.
7. Rolinski M, Szewczyk-Krolikowski K, Tomlinson PR, Nithi K, Talbot K, Ben-

Shlomo Y, et al. REM sleep behaviour disorder is associated with worse quality of life and other non-motor features in early Parkinson's disease. J Neurol Neurosurg Psychiatry. 2014;jnnp–2013.
8. JULIAN LJ. Measures of Anxiety. Arthritis Care Res [Internet]. 2011 Nov [cited 2018 Apr 10];63(0 11). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3879951/
9. Kiely KM, Butterworth P, Watson N, Wooden M. The Symbol Digit Modalities Test: Normative Data from a Large Nationally Representative Sample of Australians. Arch Clin Neuropsychol. 2014 Dec 1;29(8):767–75.
10. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord Off J Mov Disord Soc. 2008 Nov 15;23(15):2129–70.
11. Doty RL, Shaman P, Kimmelman CP, Dann MS. University of Pennsylvania Smell Identification Test: a rapid quantitative olfactory function test for the clinic. The Laryngoscope. 1984 Feb;94(2 Pt 1):176–8.