PARKINSON'S PROGRESSION MARKERS INITIATIVE

Play a Part in Parkinson's Research

Allelic status of selected Parkinson disease-associated variants for PPMI subjects with available whole-genome sequencing data

William Matloff, BE and Arthur Toga, PhD

Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA 90032, USA

Summary

The alleles of selected Parkinson disease-associated variants for PPMI subjects with available whole-genome sequencing data were compiled into a csv file to enable the quick examination of these variants in the data set. Genotype information for these selected variants was extracted from the hg38 aligned cohort VCF files from the whole-genome sequencing data (Project 118 Cohort Variants hg38 aligned ChrNN.vcf). The first column of this file contains the subject ID. The additional columns show the number of copies of the minor allele of each variant for each subject. The selected variants included in the csv file are shown in **Table 1**.

Method

The program BCFtools[1] was used to extract the variants specified in **Table 1** and compile them together into a single VCF file. Multi-allelic sites were split into two also using this software. PLINK v1.90b5[2] was then used to convert this VCF file into the binary PLINK format, enforce a minimum genotyping quality (GQ) score threshold of 20, and recode the data into a text file containing the number of copies of the minor allele of each variant for each subject. For subjects whose genotype for a variant did not meet this GQ threshold, NA was recorded for their value. Variants not present in the cohort VCF files were assumed to be homozygous for the reference allele. Four such variants were rs104893875, rs104893878, rs35801418, and rs121434567. The final csv file was compiled together using R version 3.4.1[3]. The format of the column headers is CHROM:POS:REF:ALT_MA_GENE(S)_RSID, where

CHROM:	Chromosome number
POS:	Base pair position in hg38 coordinates
REF:	Reference allele for the variant
ALT:	Alternate allele for variant
MA:	Minor allele of variant (based on the PPMI sample). Values in the table represent the number of copies a subject has of this minor allele
GENE(S):	Associated gene(s) and descriptor if available
RSID:	rsid for the variant

As an example: chr1:155235252:A:G_G_GBA_L444P_rs421016. For this variant with an rsid of rs421016, the column values represent the number of copies of the G allele at chr1:155235252.

Table 1: Parkinson disease-associated variants selected for inclusion in the table

	rsid	chr	bp_hg38	Variant Name/Implicated Gene(s)
1	rs114138760	chr1	154925709	PMVK
2	rs421016	chr1	155235252	GBA_L444P
3	rs76763715	chr1	155235843	GBA_N370S
4	rs75548401	chr1	155236246	GBA_T408M
5	rs2230288	chr1	155236376	GBA_E365K
6	rs104886460	chr1	155240629	GBA_IVS2+1
7	rs387906315	chr1	155240660	GBA_84GG
8	rs823118	chr1	205754444	NUCKS1
9	rs4653767	chr1	226728377	ITPKB
10	rs10797576	chr1	232528865	SIPA1L2
11	rs34043159	chr2	101796654	IL1R2/MAP4K4
12	rs6430538	chr2	134782397	ACMSD/TMEM163
13	rs353116	chr2	165277122	SCN3A/SCN2A
14	rs1955337	chr2	168272635	STK39
15	rs4073221	chr3	18235996	SATB1
16	rs12497850	chr3	48711556	NCKIPSD/CDC71/IP6K2
17	rs143918452	chr3	52782824	ITIH1
18	rs1803274	chr3	165773492	BuChE
19	rs12637471	chr3	183044649	MCCC1
20	rs34884217	chr4	950422	TMEM175
21	rs34311866	chr4	958159	TMEM175
22	rs11724635	chr4	15735478	BST1
23	rs6812193	chr4	76277833	FAM47E/STBD1
24	rs356181	chr4	89704988	SNCA
25	rs3910105	chr4	89761420	SNCA
26	rs104893877	chr4	89828149	SNCA_A53T
27	rs104893875	chr4	89828170	SNCA_E46K
28	rs104893878	chr4	89835580	SNCA_A30P
29	rs4444903	chr4	109912954	EGF
30	rs121434567	chr4	110004540	EGF
31	rs78738012	chr4	113439216	ANK2/CAMK2D

32	rs2694528	chr5	60978096	ELOVL7/NDUFAF2
33	rs9468199	chr6	27713436	ZNF184
34	rs8192591	chr6	32218019	NOTCH4_G1739S
35	rs115462410	chr6	32698883	HLA_DBQ1
36	rs199347	chr7	23254127	GPNMP
37	rs1293298	chr8	11854934	CTSB
38	rs591323	chr8	16839582	MICU3/FGF20
39	rs2280104	chr8	22668467	BIN3
40	rs13294100	chr9	17579692	SH3GL2
41	rs10906923	chr10	15527599	FAM171A1/ITGA8
42	rs118117788	chr10	119950976	MIR4682
43	rs329648	chr11	133895472	MIR4697
44	rs76904798	chr12	40220632	LRRK2
45	rs33939927	chr12	40310434	LRRK2_R1441G
46	rs33939927	chr12	40310434	LRRK2_R1441C
47	rs33949390	chr12	40320043	LRRK2_R1628P/H
48	rs35801418	chr12	40321114	LRRK2_Y1699C
49	rs34637584	chr12	40340400	LRRK2_G2019S
50	rs34778348	chr12	40363526	LRRK2_G2385R
51	rs11060180	chr12	122819039	OGFOD2/CCDC62
52	rs11158026	chr14	54882151	GCH1
53	rs8005172	chr14	88006268	GALC/GPR65
54	rs2414739	chr15	61701935	VPS13C
55	rs11343	chr16	19268142	COQ7/SYT17
56	rs14235	chr16	31110472	ZNF646/KAT8/BCKDK
57	rs4784227	chr16	52565276	TOX3/CASC16
58	rs11868035	chr17	17811787	SREBF1
59	rs17649553	chr17	45917282	MAPT
60	rs12456492	chr18	43093415	SYT4/RIT2
61	rs55785911	chr20	3172857	DDRGK1
62	rs737866	chr22	19942586	COMT
63	rs174674	chr22	19946502	COMT
64	rs5993883	chr22	19950115	COMT
65	rs740603	chr22	19957654	COMT
66	rs165656	chr22	19961340	COMT

67	rs6269	chr22	19962429	COMT
68	rs4633	chr22	19962712	COMT
69	rs2239393	chr22	19962905	COMT
70	rs4818	chr22	19963684	COMT
71	rs4680	chr22	19963748	COMT
72	rs165599	chr22	19969258	COMT

References

1. https://samtools.github.io/bcftools/
2. www.cog-genomics.org/plink/1.9/
3. https://www.r-project.org/

About the Authors

This document was prepared by William Matloff, Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA. For more information please contact William Matloff by email at William.Matloff@loni.usc.edu.

Notice: This document is presented by the author(s) as a service to PPMI data users. However, users should be aware that no formal review process has vetted this document and that PPMI cannot guarantee the accuracy or utility of this document.