Published in final edited form as:
Am J Med Genet B Neuropsychiatr Genet. 2010 June 5; 0(4): 973–979. doi:10.1002/ajmg.b.31061.

SNPs in CAST are associated with Parkinson disease: A confirmation study

Andrew S. Allen, PhD and
The Department of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University, Durham

Glen A. Satten, PhD
the Centers for Disease Control and Prevention, Atlanta

Abstract

Using data from the National Institutes of Neurological disease and Stroke's (NINDS) study of Parkinson disease (PD), we recently reported that single nucleotide polymorphisms (SNPs) in a region containing the Calpastatin (CAST) gene were associated with PD. Here we follow up this finding with an analysis of the Center for Inherited Disease Research's (CIDR) genome-wide association study in familial PD. After adjusting for population stratification and multiple testing, we find a significant association (p=0.0167) between PD and SNP rs1559085 in CAST. These findings confirm CAST/PD associations in a second, independent, dataset and suggest that CAST be prioritized for further investigation.

Although there are now several single-gene forms of Parkinson disease (PD) known (Klein and Schlossmacher, 2006), the genetic factors that predispose to idiopathic Parkinson disease are largely unknown (de Lau and Breteler, 2006). Two genome-wide association studies (GWASs) have been recently undertaken to address this question. The National Institute of Neurological Disease and Stroke (NINDS) PD study (Fung et al., 2006) uses cases and controls from the NINDS repository, while the Center for Inherited Disease Research (CIDR) genome-wide association study in familial PD study combines data on cases from the PROGENI and GenePD studies with controls (different from those used in NINDS PD) from the NINDS repository (Pankratz et al. 2009). Data from both studies is available through the database of genotypes and phenotypes (dbGap; http://view.ncbi.nlm.nih.gov/dbgap).

Previously, we performed a genome-wide analysis of the NINDS study data (dbGAP accession number phs000089) using three statistical tests; the single-SNP Mantel-Haenszel (MH) test as well as two novel tests based on haplotype sharing, in which the patterns of genetic similarity surrounding each SNP genotyped is compared between case and control participants (Allen and Satten, 2009). We reported a genome-wide significant association between single nucleotide polymorphisms (SNPs) in the calpastatin (CAST [MIM *114090]) gene and PD for one of the haplotype sharing tests. Here we report on the CAST/PD association we find in the CIDR familial PD study.

The CIDR familial PD data, publically released through dbGAP (accession number phs000126.v1.p1), is comprised of genotypes measured on 1943 individuals, of which 1048 are patients with PD and 895 are neurologically normal controls. The Illumina, haplotype

Correspondence to: Dr Andrew Allen, Department of Biostatistics and Bioinformatics, DUMC 2721, 2424 Erwin Road, Suite 1102, Durham, NC 27710; Tel 919-668-8043; Fax 919-688-5888; andrew.s.allen@duke.edu.

tagging human370CNV version 1C array and BeadStudio version 3.1.14 calling algorithm were used for genotyping; 344,301 SNP genotypes were publically released. In comparison, the NINDS PD study augmented the haplotype tagging HumanHap300 array with the gene-centric Illumina Infinium I, resulting in 408,803 unique SNP genotypes.

Following the recommendation of the CIDR PD study investigators, we excluded 176 samples from our analyses due to LRRK2 mutation status, population substructure, and cryptic relatedness. The study investigators also recommended that the data of 38 cases whose DNA was extracted by whole genome amplification be carefully scrutinized. We confirmed that these cases have much lower call rates, lower genotype quality scores, and lower average heterozygosity values than the rest of the data (see supplemental figures 1-4). Thus we elected to exclude these samples from subsequent analyses.

We further subjected the CIDR data to the same quality control and data cleaning measures implemented in our analysis of the NINDS PD study. In particular, we excluded data from SNPs that had extensive missingness (missingness>10%), deviations from Hardy-Weinberg equilibrium (p-value<0.001 in controls), and low minor allele frequency (<0.2%). After this quality control (QC) filtering, data on 334,225 SNPs remained. Using the software package PLINK (Purcell et al., 2007) we searched for but did not find individuals who share more than 12.5% of their SNPs identical by descent, i.e. more than would be reasonably expected in a study of unrelated persons. We also looked for persons whose genotypes suggested that their ancestry was markedly different from the study population using the spectral graph theory approach of Lee et al. (Lee et al., 2009). Following Lee et al. (2009), we filtered the cleaned SNP data to arrive at an approximately independent set of markers distributed throughout the genome. To do so, SNPs were eliminated from regions of known high linkage disequilibrium (LD) and LD was further reduced by pruning SNPs using the '--independent-pairwise' command in PLINK (Purcell et al., 2007). Using this subset of 68,288 approximately independent SNPs, we applied the eigen-analysis methods detailed in Lee et al. (2009) and found that the first 6 eigenvectors were sufficient to describe the genetic variation found in the sample. No further outliers were identified.

We adjusted for possible population stratification as in our NINDS PD analysis by using the stratification score (Epstein et al., 2007), which we calculated as the probability of being a case given only the genomic information in the 6 eigenvectors described above. Using the values of the stratification score calculated for each individual, we then divided the sample into 5 equal strata based on quintiles of the stratification score. All subsequent statistical tests are based on these five strata. This approach has been shown to successfully correct for population stratification while maintaining good power (Epstein et al., 2007; Sarasua et al., 2009).

After quality control, data cleaning and adjustment for possible confounding due to population stratification, we examined SNPs within a region containing the CAST gene where we previously found association in the NINDS PD study (see Figure 1—the region, which spans 95967.1 kB to 96187.7 kB on chromosome 5, is indicated by solid red vertical lines). We identified 48 SNPs that passed QC and were within 50kB of CAST. Twenty of these SNPs were found within CAST (both rs2785, the SNP found most associated in NINDS PD, and rs155908, the SNP found most associated here, are among these 20 SNPs). We added two additional SNPs found at either end of the CAST±50kB boundary to arrive at 50 SNPs approximately centered at the peak of the signal found in the NINDS PD analysis. We note that all 50 of these SNPs were also included in the NINDS PD analysis while 23 additional SNPs in this region were included in the NINDS PD analysis but were not genotyped in the CIDR PD study.

We tested association at each of the 50 SNPs using the same three stratified tests we used to analyze the NINDS PD data: the standard single-SNP MH test, and the *p* and *cross* tests of Allen and Satten (2009) that compare patterns of haplotype similarity between cases and controls. As in our previous analysis of the NINDS data, haplotype similarity tests were calculated using 15-SNP windows centered at each of the 50 SNPs we studied (data from SNPs outside of our 50 SNP region was used to calculate haplotype similarity for windows that were within 7 SNPs of the boundaries of our region). We explicitly adjusted for the fact that we are conducting 150 tests (3 different tests centered at each of 50 SNPs) using the permutation-based step-down procedure of Westfall and Young (1993).

The results of our analysis can be found in table I. After adjusting for multiple comparisons, we found the single SNP MH test to be significant (rs1559085—p=0.0167; unadjusted p=0.0002). In our previous evaluation of the NINDS PD data, we found the *cross* test was genome-wide significant. In the CIDR PD data, neither haplotype sharing test was significant after adjustment for multiple testing: *p* test (rs1559085—p=0.1028; unadjusted p=0.0019); *cross* test (rs152280—p=0.3089; unadjusted p=0.0076). However, the most associated SNP in CIDR PD study (rs1559085) is only 32.7kB away from the SNP (rs27852) that was the center of the association peak we found in the NINDS PD data with the *cross* test. The two SNPs also show no evidence of recombination between them as the D' between the SNPs is at its maximal value of 1. The effect size conveyed by rs1559085 is also similar in the two studies with the CIDR PD study showing an allelic odds ratio of 1.43 [95% unadjusted CI=(1.18, 1.74)] while the NINDS PD study shows an allelic odds ratio of 1.56 [95% unadjusted CI=(1.09, 2.21)]. We further note that, although the MH test failed to reach genome-wide significance in the NINDS PD data, the MH test does exhibit some signal in the NINDS PD data. The minimum MH unadjusted *p* value over the 73 NINDS PD SNPs found in the region from 95967.1 kB to 96187.7 kB on chromosome 5 (indicated by solid red lines in figure 1) was 7.9×10^{-5} (at rs10053056, which is 35.3kB away from rs1559085 and only 2.6 kB from rs27852). Additionally, 37% of these SNPs (27 out of 73) have unadjusted *p*-values below 0.05.

Our analysis of the CIDR familial PD data confirms the association between Parkinson's disease and the genomic region we previously identified in the NINDS PD data. Further, the effect sizes and *p*-values seen in the two studies are quite similar. We are unsure why the strongest haplotype sharing signal in the NINDS PD data (the *cross* statistic) produced the weakest signal in the CIDR PD data. One reason may be that the NINDS PD study had a higher density of SNPs, especially in gene-centric regions, making it easier to detect sharing patterns. Also, the NINDS PD study sampled isolated PD cases without extensive PD family history (individuals with 3 or more relatives with Parkinsonism were excluded), while the CIDR PD study sampled only probands with close PD relatives. Thus a possible explanation for the difference in the strength of the sharing signal may be due to different types of Parkinson's risk variants being captured in the two study designs. Individuals with close PD relatives (as in CIDR PD) are more likely to have their genetic PD risk explained by highly penetrant but rare variants than are those without close PD relatives (as in NINDS PD). These rare variants may only segregate within a single family in the study population, which would dilute the haplotype sharing signal about such loci (recall that the relatives with PD were not included in the study).

Our analysis of the NINDS and CIDR PD GWASs find PD association in a genomic region containing the calpastatin (CAST) gene. Calpastatin inhibits calpain, a calcium-dependent protease participating in a variety of physiologic processes (Goll et al., 2003). Calpains have been implicated in neurodegenerative disorders such as PD, Alzheimer's and multiple sclerosis (Saito et al., 1993; Mouatt-Prigent et al., 1996; Tsuji et al., 1998; Shields et al., 1999; Adamec et al., 2002; Raynaud and Marcilhac, 2006). The mid-brains of PD patients

have been found to have higher levels of calpain (Mouatt-Prigent et al., 1996) and calpain overexpression may play a role in neuronal death (Shukla et al., 2006; Camins et al., 2006). Inhibiting calpain in animal models of PD prevents neuronal and behavioral deficits (Crocker et al., 2003). Thus, variation in calpastatin expression that affects calpain inhibition may explain how CAST genotypes affect PD risk. Our confirmation of the CAST locus/PD association in a second data set suggests that CAST involvement in PD etiology should be further investigated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the NINDS and the CIDR whole genome association study in familial Parkinson's Disease study investigators for providing the CIDR PD data through dbGap. Funding support for the CIDR PD study and the genotyping of samples was provided by the NINDS (Foroud/Myers, PI). The dataset used in analyses described in this manuscript was obtained from the NINDS Database found at http://view.ncbi.nlm.nih.gov/dbgap through dbGaP accession number phs000126.v1.p1. A.S.A. acknowledges support from the NIH through NHLBI grant K25 HL077663 and NIMH grant R01 MH084680.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

Adamec E, Mohan P, Vonsattel JP, Nixon RA. Calpain activation in neurodegenerative diseases: confocal immunofluorescence study with antibodies specifically recognizing the active form of calpain 2. Acta Neuropathol. 2002; 104:92–104. [PubMed: 12070670]

Allen AS, Satten GA. A novel haplotype-sharing approach for genome-wide case-control association studies implicates the calpastatin gene in Parkinson's disease. Genet Epidemiol. 2009 Published online: Apr 13.

Camins A, Verdaguer E, Folch J, Pallàs M. Involvement of Calpain Activation in Neurodegenerative Processes. CNS Drug Reviews. 2006; 12:135–148. [PubMed: 16958987]

Crocker SJ, Smith PD, Jackson-Lewis V, et al. Inhibition of calpains prevents neuronal and behavioral deficits in an MPTP mouse model of Parkinson's disease. The Journal of Neuroscience. 2003; 23:4081–4091. [PubMed: 12764095]

de Lau LML, Breteler MMB. Epidemiology of Parkinson's Disease. Lancet Neurology. 2006; 5:525–535. [PubMed: 16713924]

Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case-control studies. Am J Hum Genet. 2007; 80:921–30. [PubMed: 17436246]

Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurology. 2006; 5:911–916. [PubMed: 17052657]

Goll DE, Thompson VF, Li H, et al. The calpain system. Physiol Rev. 2003; 83:731–801. [PubMed: 12843408]

Klein C, Schlossmacher MG. Parkinson disease, 10 years after its genetic revolution: Multiple clues to a complex disorder. Neurology. 2007; 69:2093–2104. [PubMed: 17761553]

Lee AB, Luca D, Klei L, et al. Discovering genetic ancestry using spectral graph theory. Genetic Epidemiology. 2009 Published Online: May 19.

Mouatt-Prigent A, Karlsson J, Agid Y, Hirsch E. Increase m-calpain expression in the mesencephalon of patients with Parkinson's disease but not in neurodegenerative disorders involving the mesencephalon: a role in nerve cell death? Neuroscience. 1996; 73:979–987. [PubMed: 8809817]

Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, Doheny KF, Gusella JF, Nichols WC, Foroud T, Myers RH. the PSG-PROGENI and GenePD Investigators, Coordinators

and Molecular Genetic Laboratories. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet. 2009; 124:593–605. [PubMed: 18985386]

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

Raynaud F, Marcilhac A. Implication of calpain in neuronal apoptosis:A possible regulation of Alzheimer's disease. FEBS Journal. 2006; 273:3437–3443. [PubMed: 16884489]

Saito K, Elce J, Hamos J, Nixon R. Widespread activation of calcium-activated neutral proteinase (calpain) in the brain in Alzheimer disease: A potential molecular basis for neuronal degeneration. Proc Nat Acad Sci USA. 1993; 90:2628–2632. [PubMed: 8464868]

Sarasua SM, Collins JS, Williamson DM, et al. Effect of population stratification on the identification of significant SNPs in genome wide association studies. To appear in BMC Proceedings. 2009

Shields DC, Schaecher KE, Saido TC, Banik NL. A putative mechanism of demyelination in multiple sclerosis by a proteolytic enzyme, calpain. Proc Natl Acad Sci USA. 1999; 96:11486–11491. [PubMed: 10500203]

Shukla M, Rajgopal Y, Babu PP. Activation of calpains, calpastatin and spectrin cleavage in the brain during the pathology of fatal murine cerebral malaria. Neurochemistry International. 2006; 48:108–113. [PubMed: 16236382]

Tsuji T, Shimohama S, Kimura J, Shimizu K. m-Calpain (calcium-activated neutral proteinase) in Alzheimer's disease brains. Neurosci Lett. 1998; 248:109–112. [PubMed: 9654354]

Westfall, PH.; Young, SS. Resampling-Based Multiple Testing. New York: J Wiley & Sons; 1993. p. 340

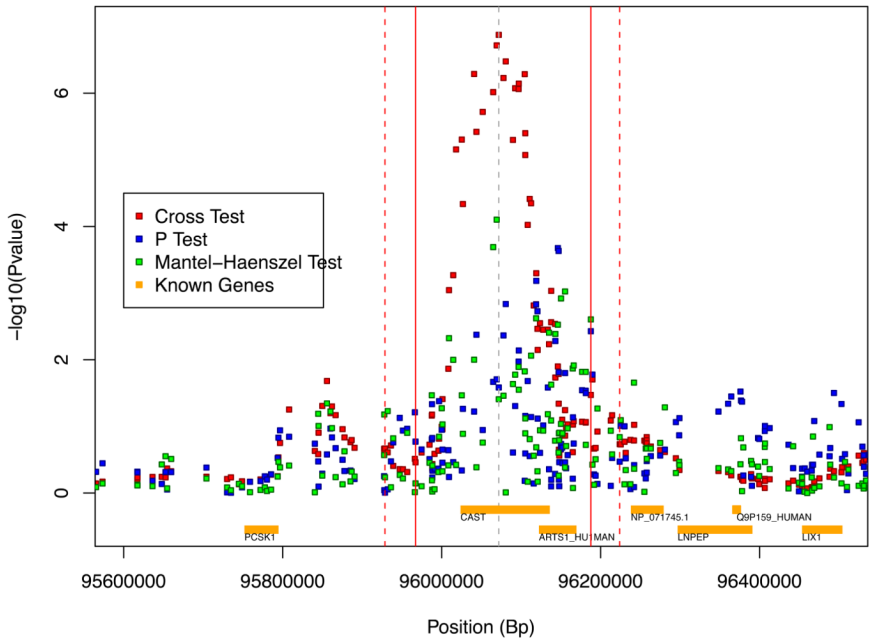

Figure 1.

Analysis results from NINDS data reported in Allen and Satten (2009) highlighting follow-up region. Solid red vertical lines denote region comprised of 50 SNPs in the CIDR PD data. Dotted red vertical lines denote extent of SNPs used in haplotype sharing tests.

Table I

Results of CAST region follow-up study in CIDR PD data. Fifty loci were tested by 3 different tests (150 tests total). Results sorted by raw p-values. Bold entry denotes significant locus after adjusting for multiple comparisons. Tests: mh—Mantel Haenszel test; p—haplotype sharing *p* test; c—haplotype sharing *cross* test.

SNP	Physical position (Bp)	Test	Raw P-value	Adjusted P-value
rs1559085	**96104458**	**mh**	**0.0002**	**0.0167**
rs1559085	96104458	p	0.0019	0.1022
rs152280	96187698	c	0.0076	0.3054
rs27434	96155268	c	0.01	0.3713
rs10515248	96137761	mh	0.0101	0.3761
rs13362120	96096950	p	0.012	0.4246
rs10515248	96137761	p	0.0176	0.5373
rs10053056	96069176	mh	0.018	0.5442
rs28129	96181274	c	0.0224	0.6054
rs34753	96176667	c	0.0225	0.6074
rs155056	96000101	mh	0.0247	0.6411
rs26618	96156592	c	0.025	0.6411
rs261212	95967121	c	0.0257	0.6495
rs42398	96146211	c	0.0257	0.6495
rs30187	96150086	c	0.0268	0.6624
rs469783	96147280	c	0.0268	0.6626
rs18036	96181364	c	0.0279	0.6712
rs10050860	96147966	c	0.0307	0.7024
rs27037	96120450	mh	0.0357	0.7471
rs155040	95995910	p	0.0446	0.8173
rs27042	96142696	c	0.046	0.8238
rs6885297	95974758	c	0.0565	0.8713
rs17086432	95996726	mh	0.0717	0.9214
rs261234	95988209	p	0.073	0.9217
rs27042	96142696	mh	0.082	0.9446
rs3853203	95987808	p	0.0834	0.9447
rs2611713	95984706	c	0.0861	0.9482
rs4869304	96009225	mh	0.0951	0.9618
rs10515248	96137761	c	0.0991	0.9654
rs10515240	95987585	p	0.1011	0.9684
rs261235	95988173	p	0.1066	0.9742
rs4869305	96018123	mh	0.1155	0.9801
rs469783	96147280	mh	0.1208	0.984
rs10515240	95987585	c	0.1219	0.9841
rs2611713	95984706	p	0.1227	0.9846
rs28081	96118746	c	0.1296	0.9873
rs17086408	95990991	p	0.1331	0.9888
rs27037	96120450	c	0.1344	0.9893

SNP	Physical position (Bp)	Test	Raw P-value	Adjusted P-value
rs1559085	96104458	c	0.1366	0.9903
rs152005	96051603	mh	0.1368	0.9904
rs27037	96120450	p	0.1388	0.9904
rs18036	96181364	p	0.1411	0.9915
rs261212	95967121	mh	0.1589	0.9955
rs10053056	96069176	c	0.159	0.9955
rs152005	96051603	c	0.1666	0.9955
rs27852	96071795	mh	0.2042	0.9992
rs26618	96156592	p	0.2053	0.9993
rs27429	96120729	c	0.2141	0.9993
rs27524	96127700	c	0.2145	0.9993
rs469532	96077746	c	0.2188	0.9993
rs27851	96108157	c	0.2253	0.9993
rs27524	96127700	mh	0.2255	0.9993
rs28096	96135000	c	0.2414	0.9997
rs151835	96096738	c	0.2417	0.9997
rs18036	96181364	mh	0.2421	0.9997
rs4254932	96000943	mh	0.2425	0.9997
rs34753	96176667	p	0.2454	0.9997
rs28129	96181274	p	0.253	0.9998
rs469783	96147280	p	0.2564	0.9999
rs10050860	96147966	mh	0.2578	0.9999
rs17086432	95996726	p	0.261	0.9999
rs3756623	96110704	mh	0.2622	0.9999
rs469532	96077746	mh	0.2645	0.9999
rs27852	96071795	c	0.2698	0.9999
rs261234	95988209	c	0.2711	1
rs27429	96120729	p	0.2734	1
rs28081	96118746	p	0.28	1
rs2611713	95984706	mh	0.2848	1
rs27772	96115732	mh	0.2863	1
rs27851	96108157	p	0.2912	1
rs13362120	96096950	c	0.2921	1
rs17086512	96040813	mh	0.2924	1
rs26482	96043712	c	0.2975	1
rs27434	96155268	p	0.3043	1
rs10515241	95988347	p	0.3059	1
rs27772	96115732	c	0.3157	1
rs27434	96155268	mh	0.3212	1
rs17086512	96040813	c	0.3221	1
rs261235	95988173	c	0.3266	1
rs4869304	96009225	c	0.3305	1

SNP	Physical position (Bp)	Test	Raw P-value	Adjusted P-value
rs10515240	95987585	mh	0.3309	1
rs3853203	95987808	c	0.3383	1
rs151835	96096738	mh	0.3395	1
rs10515241	95988347	mh	0.3432	1
rs155054	96001058	mh	0.3442	1
rs3853203	95987808	mh	0.3451	1
rs28096	96135000	p	0.3471	1
rs4869305	96018123	p	0.3498	1
rs3756623	96110704	c	0.3651	1
rs26618	96156592	mh	0.3688	1
rs6885297	95974758	p	0.3747	1
rs1421911	96026703	c	0.3784	1
rs27042	96142696	p	0.3866	1
rs261235	95988173	mh	0.389	1
rs155056	96000101	c	0.3976	1
rs17086408	95990991	c	0.4017	1
rs13362120	96096950	mh	0.405	1
rs30187	96150086	mh	0.4112	1
rs155054	96001058	p	0.4116	1
rs17086408	95990991	mh	0.4189	1
rs27524	96127700	p	0.4315	1
rs26482	96043712	mh	0.4319	1
rs28129	96181274	mh	0.4348	1
rs4254932	96000943	c	0.4368	1
rs34753	96176667	mh	0.4405	1
rs4254932	96000943	p	0.442	1
rs469532	96077746	p	0.4447	1
rs1862182	96025282	c	0.4459	1
rs1421911	96026703	mh	0.4528	1
rs152280	96187698	mh	0.4644	1
rs261212	95967121	p	0.4652	1
rs1862609	96112446	c	0.4658	1
rs10515242	96008431	p	0.4659	1
rs10515241	95988347	c	0.4756	1
rs30187	96150086	p	0.4791	1
rs10515242	96008431	mh	0.4925	1
rs26482	96043712	p	0.4989	1
rs4869305	96018123	c	0.4991	1
rs10515242	96008431	c	0.5142	1
rs152005	96051603	p	0.5199	1
rs155054	96001058	c	0.5322	1
rs42398	96146211	mh	0.539	1

SNP	Physical position (Bp)	Test	Raw P-value	Adjusted P-value
rs4400148	96014543	c	0.5499	1
rs1862182	96025282	mh	0.565	1
rs1421911	96026703	p	0.5676	1
rs4869304	96009225	p	0.6087	1
rs17086432	95996726	c	0.6249	1
rs42398	96146211	p	0.6261	1
rs27852	96071795	p	0.6867	1
rs6885297	95974758	mh	0.6879	1
rs3756623	96110704	p	0.6963	1
rs155040	95995910	c	0.6983	1
rs261234	95988209	mh	0.7214	1
rs1862609	96112446	mh	0.731	1
rs152280	96187698	p	0.7346	1
rs1862609	96112446	p	0.755	1
rs10050860	96147966	p	0.7579	1
rs1862182	96025282	p	0.7805	1
rs10053056	96069176	p	0.7833	1
rs27772	96115732	p	0.7925	1
rs17086512	96040813	p	0.7937	1
rs28096	96135000	mh	0.8419	1
rs4400148	96014543	p	0.8439	1
rs155040	95995910	mh	0.8586	1
rs28081	96118746	mh	0.8802	1
rs4400148	96014543	mh	0.8874	1
rs151835	96096738	p	0.9023	1
rs155056	96000101	p	0.9488	1
rs27429	96120729	mh	0.9603	1
rs27851	96108157	mh	0.9974	1