Rare coding SNP in DZIP1 gene associated with late-onset sporadic Parkinson’s disease

André X. C. N. Valente1,2,3, Joo H. Shin4, Abhijit Sarkar5 & Yuan Gao4

1Systems Biology Group, Biocant - Biotechnology Innovation Center, Cantanhede, Portugal, 2CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal, 3Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia, USA, 4Lieber Institute for Brain Development, Johns Hopkins Medical Campus, 855 N. Wolfe Street, Suite 300, Baltimore, Maryland 21205, 5Department of Physics and Vitreous State Laboratory, Catholic University of America, Washington, DC, USA.

An association between a rare, coding, non-synonymous SNP variant in the gene DZIP1 and Parkinson’s disease was found, based on an analysis of the existing NGRC genome-wide association study dataset. The statistical analysis utilized the hypothesis-rich, targeted search unbiased assessment approach, rather than the hypothesis-free, genome-wide agnostic search paradigm. The association of DZIP1 with Parkinson’s disease is discussed in the context of a Parkinson’s disease stem-cell ageing theory.

Family genetic linkage studies have associated six genes with Mendelian inheritable forms of Parkinson’s disease (PD)1. However, these monogenic forms account for fewer than 10% of PD cases. Further, they lead mostly to juvenile or early onset forms of PD (before age 50). Given that no decisive environmental causative factors have been found either, the etiology of late-onset PD (comprising over 90% of all PD cases) remains essentially underdetermined. A range of hypotheses are being explored2–6. We have proposed the theory that i) sporadic PD is best defined as a characteristic deviation from normality in the expression program of a cell (the PD-state) and ii) this PD-state can originate as a case of hematopoietic stem-cell program defect7.

At present, considerable efforts are focused on finding differential genetic susceptibility to late-onset PD via the genome-wide association study (GWAS)8. In a GWAS, a set of patients and controls is genotyped at known SNP sites in the human genome. Mathematically, this assigns individuals to locations within a high-dimensional SNP space (Figure 1). Genetic susceptibilities are inferred from statistically significant differences in the placement of patients and controls in this SNP space. Large enough differential disease risks constitute practical predictive genetic markers. So far though, susceptibilities found have been typically weak (some 85% of trait associated SNPs reported have an odds ratio in the 0.5–2 range)9. Nonetheless, such findings can still be invaluable as indicators of the involvement of particular genes or biological processes in the disease mechanics. As of today, GWASs have reported about a dozen, modest effect (odds ratio in the 0.5–2 range), susceptibility loci for PD10–17.

The hypothesis-free paradigm currently dominates GWAS statistical data analysis8. It has been previously described why this is a poor choice18–21. Biological knowledge and insightful hypotheses are as crucial in the analysis of a GWAS as they are in the analysis of any classical biological experiment22–25. The alternative hypothesis-rich mathematical theory recognizes this fact and allows biological thought to maximize statistical power21,26. Key in the approach is the concept of Rational Class (RC), a set of candidate laws (markers in the GWAS context) that share an underlying common rationale.

In this article, we analyze the late-onset sporadic PD GWAS NGRC dataset of Hamza et al.10, under the hypothesis-rich framework (the late-onset, sporadic qualifier will be henceforth subsumed)26,27. In the Methods section, the focus is on describing the RCs constructed specifically for this PD GWAS analysis. Findings are summarized in the Results section. Finally, in the Discussion section, we review relevant biological information to contextualize our findings.

Results

The significant findings from the hypothesis-rich analysis of the Hamza et al. dataset are presented in Table 1. For these SNPs, the null hypothesis was that the two regions defined by the split mode (see Figure 3: 1-dimensional split modes) present no differential susceptibility to PD. Now, pure chance in the finite sampling of individuals
from the population could create the false impression of differential susceptibility. Assuming the null hypothesis, the reference probability indicates the ease of such stochasticity producing an unwarranted call (as per the hypothesis-rich framework) of differential susceptibility\textsuperscript{27}. We emphasize that the quoted reference probabilities already take into consideration the presence of multiple-hypotheses testing. For easiness of comparison, the arbitrariness in defining odds ratio (given the validity of the inverse of any choice) was settled by making every odds ratio larger than unity. The minor allele effect entry then indicates which region carries the greater risk of PD.

The reported SNPs in the SNCA region and in the HLA-DRA region had all been noted as significant in previous GWAS\textsuperscript{10,28,29}. The SNPs reported in the chromosome 17 q21.31 region (usually categorized as the MAPT region) validate the previous GWAS based association of this region with PD (most of these MAPT region SNPs have been specifically previously reported, though we could not confirm all)\textsuperscript{10–17}. The novel finding is the increased susceptibility to PD conferred by a rare, coding, non-synonymous SNP variant in the DZIP1 gene (Figure 4).

### Discussion

The PD working theory we put forward in previous work\textsuperscript{7} provides a possible context for the connection of DZIP1 with PD found in this analysis. Therefore, we start by reviewing it. Firstly, PD would be defined in terms of the PD-state, a characteristic deviation from the normal expression program of a healthy cell. Singular cellular manifestations of PD would therefore be de-emphasized in favor of this systems-level definition. Crucially, the PD-state would be a generic cell state, not restricted to neurons. Secondly, the PD-state would be a generic cell state, not restricted to neurons. Secondly, the PD-state would be a generic cell state, not restricted to neurons.

---

**Table 1** Summary of the findings from hypothesis-rich analysis of the Hamza et al. GWAS PD dataset. See the Results main text section for meaning of the entries.

| Rational Class | Gene | SNP | Location | Minor allele frequency | Minor allele effect | Split mode | Odds ratio | Reference probability |
|----------------|-----------------|------|----------|-----------------------|---------------------|------------|------------|----------------------|
| RC #2 coding region minor allele freq: 10%–30% | IMP5, MAPT, CRHR1, KIAA1267, C17orf69, NSF | rs12373123, rs12185235, 17651549, rs16940665, 12185268, kgp6408681, rs1052551, rs16940674, rs36076725, rs17652121, kgp3974170, rs17574604, kgp3365508, rs10445337, rs1881193, rs3583914, rs1052553, kgp4886152, rs199533 | chr. 17, q21.31 in coding regions synonymous & non-synonymous substitutions | 19% thru 21% | protective | minor allele dominant | 1.2 thru 1.3 | 0.06 thru 0.09 |
| RC #7 hematopoietic coding region minor allele freq. < 10% | DZIP1 | kgp1112497 | chr. 13, q32.1 coding region non-synonymous substitution | 0.7% | harmful | minor allele dominant | 4.4 | 0.03 |
| RC #15 generic (non-coding/non-UTR) | HLA-DRA | rs3129822 | chr. 6, p21.3 intronic | 44% | harmful | extreme | 1.5 | 0.04 |
| 30% < minor allele freq. | SNCA-GPRIN3, SNCA | rs356220, rs2736990, rs356168 | chr. 4, q22.1 intergenic (SNCA-GPRIN3) intronic (SNCA) | 40% thru 49% | harmful | minor allele dominant | 1.3 thru 1.4 | 0.04 thru 0.07 |
We now describe what is known at present about the biological role of DZIP1. The gene DZIP1 encodes a C2H2-type zinc finger protein. Its acronym stands for DAZ-interacting protein 1, as DZIP1 was originally identified in a screen for protein interaction partners of the DAZ (deleted in azoospermia) protein 30. Its expression in human embryonic, stem, fetal and adult germ cells was thus well noted. Zebrafish mutants in *iguana* (the DZIP1 ortholog in Zebrafish) have been invaluable in characterizing the gene. A *iguana* mutant (*jo10a*) displayed ultrastructural defects in perivascular mural cell recruitment and subsequent hemorrhage, thus linking vascular stability and DZIP1. Work with Zebrafish *iguana* mutants also revealed DZIP1 to be a component of the Hedgehog (Hh) signaling pathway. Within the Hh pathway, DZIP1 acts downstream of Smoothened, modulating the activity of the Gli family of transcription factors. DZIP1 has further been implicated in primary ciliogenesis and its role in Hh signaling may occur in this context. Hh plays a vital part in directing embryonic pattern formation. However, it continues regulating adult stem cells beyond embryogenesis. Studies have specifically implicated Hh in the adult maintenance of hematopoietic stem cells, epithelial stem cells in the gastrointestinal tract, neuronal stem cells in the subventricular zone (SVZ) and the hippocampal dentate gyrus, hair follicle stem cells, mammary stem cells and mesenchymal stem cells. Besides its role in neurogenesis, Hh has also shown neurotrophic properties, in particular regarding dopaminergic neuron survival. Under a cancer stem-cell hypothesis interpretation, this is consistent with the role of Hh in adult stem cell homeostasis. The aberrant Hh signaling would contribute the conversion of adult stem cells (or perhaps their early progeny) into cancer stem cells, cells endowed with stem-cell properties and trapped in a pathological state of constant renewal. Now, under our PD hypothesis, PD also originates in a stem-cell program defect. However, while in the cancer stem cell hypothesis the pathology progresses via physical replication of the cancer stem cells themselves, in PD we are proposing propagation solely of the PD phenotype as having a causative effect on PD susceptibility, which clearly would not be the case here. The issue may also arise merely by study recruitment centers in areas with distinct subpopulations not enrolling identical ratios of PD to control subjects. Note

**Methods**

We analyzed the NGRC GWAS dataset of Hamza et al. consisting of 1986 control subjects and 2000 sporadic late-onset PD patients. All individuals were Americans of self-reported European ancestry. As in any GWAS, a concern is the presence of population structure in the cohort data. Likely the European population, due to historical and geographical factors, does not constitute, mating-wise, a single uniformly mixed population. Now, suppose the existing subpopulations have distinct susceptibilities to PD. This could be due to differences in genetic background, culture (e.g., diet), or physical environment. Regardless, a genetic marker of a subpopulation (e.g., a SNP variant typical of a subpopulation) would then effectively mark a distinct susceptibility to PD. This poses a problem, in that we would like to interpret markers as having a causative effect on PD susceptibility, which clearly would not be the case here. The issue may also arise merely by study recruitment centers in areas with distinct subpopulations not enrolling identical ratios of PD to control subjects. Note
that although for simplicity we allude above to discrete subpopulations, generically the mixing makeup will have a continuous character.

To analyze the dataset of Hamza et al., we used the SNP space coordinate system shown in Figure 1. The relative overall location of individuals in SNP space (Euclidean distance wise) reflects the cohort population structure. Namely, relative locations are consistent with parental country of origin for those subjects that reported such information and for whom both parents had a common origin. This is visually clear upon a change of coordinates to principal component coordinates (Figure 2).

A variety of methods exist for mitigating the population structure problem\(^7\). We chose to homogenize the population regarding the PD to control subject ratio, via individual weight knock-down. Briefly, this involves reducing the statistical weight of selected individuals to locally level the ratio of PD to control subjects throughout SNP space. A separate article describes in detail both the method and its application to this particular dataset\(^7\). The homogenization procedure reduced the dataset to a net weight of 1904 PD patients and 1802 controls (a 7% size reduction). This will be the reference dataset henceforth.

We utilized the hypothesis-rich framework to investigate the dataset\(^26,27\). The hypothesis-rich framework provides a targeted search, unbiased approach to the analysis of GWAS data. The targeted search assertion follows from biological considerations guiding the statistical search for genetic susceptibility factors. Specifically, biological information enters the mathematical analysis via the concept of Rational Class (RC), a set of candidate genetic markers that share a common rationale. Yet, in spite of the biased search, an unbiased assessment is obtained from a proper mathematical treatment of multiple hypotheses testing\(^26,27\).

We now describe the RCs constructed for the PD GWAS problem. Throughout, recall that separating markers into distinct RCs can be statistically advantageous if the resulting RCs have different True Quality Distributions and Correlation Structures (short-hand, TQDs)\(^26\). This can be the case whenever a biological rationale underlies the marker separation. On the other hand, RCs must be rank ordered and statistical resolution decreases with increasing rank, thus overly liberal RC creation is pointless\(^26-27\).

A total of 23 RCs were constructed (Table 2). The first 15 RCs, containing individual SNPs, were based on the following factors:

**Genomic region.** We grouped SNPs by whether they fell in a coding region, in the UTR or in the remainder of the genome. Confirming the distinct biological roles of these regions, past GWASs show the incidence of trait associated SNPs in them is not uniform\(^2\).

**SNP allele frequencies.** These frequencies are affected by the degree of selective pressure on the associated haplotypes. Thus, on average, the character of SNPs with different allele frequency ratios may be distinct. We divided SNPs into three broad groups, based on their minor variant frequency: <10%, 10–30% and >30%. Also, note that we are comparatively more interested in larger odds ratio markers. Given two SNP markers showing the same statistical significance (ordinarily, same p-value), the one with the lower minor variant frequency necessarily shows a larger differential trait susceptibility (larger odds ratio). Thus, as an additional benefit, the above frequency breakdown effectively protects the search for rare variant, high odds ratio markers.

**Hematopoietic fingerprints.** Given our PD working theory, the set of SNPs occurring in genes with a function in the hematopoietic system are of particular relevance. We recorded 2253 SNPs spread across 662 so called hematopoietic fingerprint genes\(^2\). The genes were identified by Chambers et al. via global gene expression profiling of murine hematopoietic stem cells and their major differentiated lineages (NK-cells, T-cells, B-cells, monocytes, neutrophils and nucleated erythrocytes\(^3\)).

Combination of the above factors yielded RCs 1 thru 15 (Table 2). In these RCs, SNPs were tested for association with differential PD risk three separate times, each time based on a different mode of splitting the SNP space (Figure 3, 1-dimensional split modes). The dominant and recessive modes were motivated by their well known biological counterparts. However, a situation where phenotype is significantly more assured only under homogeneous alleles is also biologically plausible. The extreme mode accommodates these cases by excluding individuals with heterogeneous alleles from the statistical comparison. In every case, the null hypothesis was that the two regions defined by the split mode present no different susceptibility to PD. Statistical comparison between the two chosen regions was done via the Fisher exact test.

RCs 16 thru 19 were based on SNP pairs. Given there are on the order of 10^5 SNPs, potential SNP pairs are on the order of 10^10. A RC containing such a large number of entries is unlikely to have a favorable TQD\(^2\). It is therefore fundamental to prioritize SNP pairs. We generated one list of SNP pairs based on protein-protein physical interactions. For every two interacting proteins on different chromosomes, all SNP pairs with one SNP in each of the interacting proteins respective coding gene region were added to the list. The exclusion of protein pairs on the same chromosome excludes pairs of SNPs potentially in linkage disequilibrium. Protein-protein interactions were obtained from HPRD (~39000 interactions)\(^4\). The SNP pairs were tested for association with differential PD risk five times, each time based on a different mode of splitting the SNP space (Figure 3, 2-dimensional split modes). In every case, the null hypothesis was that the two regions defined by the split mode present no different susceptibility to PD. Statistical comparison between the two chosen regions was done via the Fisher exact test. The results of the tests were assigned to RC 16 or to RC 17 depending on whether the associated odds ratio was smaller or larger than 3. Once more, this has the benefit of safeguarding the search for high odds ratio markers.

A second list of SNP pairs was constructed based on the hematopoietic fingerprint genes. Based on the expression profiling, Chambers et al. had further divided the hematopoietic fingerprint genes into the following subclasses: hematopoietic stem cells, B-cells, naive T-cells, NK-cells, monocytes, granulocytes, nucleated erythrocytes, differentiated shared fingerprint, lymphoid shared fingerprint and myeloid shared fingerprint 57. We generated hematopoietic gene pairs by considering every possible pairing of genes within the same hematopoietic subclass, exclusive of gene pairs in the same chromosome. The procedure described above for protein pairs was then applied to the hematopoietic gene pairs, thus generating RCs 18 and 19.

RCs 20 thru 22 contained SNP tuplets generated from protein complexes. Human protein complexes were obtained from the CORUM database (~1300 complexes)\(^6\). Consider first RC 20, containing 2-tuplets generated from complexes of up to 4 proteins. The 2-tuplets for RC 20 were generated as follows:

| Table 2 | The Rational Classes (RCs) constructed to analyze the PD GWAS data |
|---|---|
| **Rational Class Rank** | **Rational Class Description** |
| 1 | coding | allele freq.: < 10% | dominant, recessive and extreme |
| 2 | coding | allele freq.: 10% to 20% | dom., rec. extr. |
| 3 | coding | allele freq.: > 30% | dom., rec. extr. |
| 4 | UTR | allele freq.: < 10% | dom., rec. extr. |
| 5 | UTR | allele freq.: 10% to 20% | dom., rec. extr. |
| 6 | UTR | allele freq.: > 30% | dom., rec. extr. |
| 7 | hematopoietic | coding | allele freq.: < 10% | dom., rec. extr. |
| 8 | hematopoietic | coding | allele freq.: 10% to 20% | dom., rec. extr. |
| 9 | hematopoietic | coding | allele freq.: > 30% | dom., rec. extr. |
| 10 | hematopoietic | UTR | allele freq.: < 10% | dom., rec. extr. |
| 11 | hematopoietic | UTR | allele freq.: 10% to 20% | dom., rec. extr. |
| 12 | hematopoietic | UTR | allele freq.: > 30% | dom., rec. extr. |
| 13 | non-coding/non-UTR | allele freq.: < 10% | dom., rec. extr. |
| 14 | non-coding/non-UTR | allele freq.: 10% to 20% | dom., rec. extr. |
| 15 | protein pairwise interactions | coding | OR > 3 | 5 split modes |
| 16 | protein pairwise interactions | coding | OR < 3 | 5 split modes |
| 17 | hematopoietic pairwise interactions | coding | OR > 3 | 5 split modes |
| 18 | hematopoietic pairwise interactions | coding | OR < 3 | 5 split modes |
| 19 | protein complexes | sizes 2 thru 4 | top 2 proteins |
| 20 | protein complexes | sizes 3 thru 9 | top 3 proteins |
| 21 | protein complexes | sizes 4 thru 16 | top 4 proteins |
| 22 | gene expression sig. | coding and UTR | all freqs. | dom., rec. extr. |
1. Given a complex, consider the SNPs that fall in the coding region of the protein members of the complex. Denote them as CSNPs. Add every possible (CSNP A, CSNP B) 2-tuplet to RC 20, provided CSNP A and CSNP B are located in different chromosomes.

2. Repeat for every complex of up to 4 proteins.

Each 2-tuplet was tested for association with differential PD risk 3 separate times, as follows:

1. Under the dominant 1-dimensional split mode, assign a Fisher exact test based p-value to each CSNP in the tuplet in the standard fashion (i.e., considering the CSNP as an individual SNP, as in the RCs 1 thru 15). We formalize it by writing p-value = p(CSNP; dominant mode).

2. The p-value associated with the 2-tuplet is (max(p(CSNP A; dominant mode), p(CSNP B; dominant mode)))1/2 (i.e., squared).

3. Assign two more p-values to the tuplet, as above, but now utilizing the recessive and extreme 1-dimensional split modes.

RC 21 was similar to RC 20, except that:

1. It was based on complexes of sizes 3 thru 9.

2. It contained 3-tuplets (CSNP A, CSNP B, CSNP C).

3. The p-value associated with a 3-tuplet is (max(p(CSNP A; split mode), p(SNP B; split mode), p(SNP C; split mode))) (i.e., cubed).

4. RC 22 was similar to RC 20, except that:

   1. It was based on complexes of sizes 4 thru 16.

   2. It contained 4-tuplets (CSNP A, CSNP B, CSNP C, CSNP D).

5. The p-value associated with a 4-tuplet is (max(p(CSNP A; split mode), p(SNP B; split mode), p(SNP C; split mode), p(SNP D; split mode))) (i.e., to the fourth power).

   In these complex based RCs, in every case the null hypothesis is that none of the SNPs in the tuplet shows differential susceptibility to PD between the two regions defined by the split mode. The anticipation is that a complex mechanistically involved in PD may produce a tuplet (or tuplets) of particular low p-value under the above tuplet p-value definition. RCs 20, 21 and 22 are kept separate to preserve potentially distinct TQDs.

   Finally, RC 23 was based on genes in the blood gene expression signature for PD (involving 18 genes) we developed in earlier work 26. RC 23 contained:

   1. All SNPs in the coding or UTR regions of the genes present in the expression signature.

   2. All SNP pairs, exclusive of pairs in the same chromosome, with one SNP in the coding or UTR regions of one expression signature gene and the other SNP in the coding or UTR region of a second expression signature gene.

   The individual SNPs were tested for association with differential risk of PD under the three 1-dimensional split modes via the Fisher exact test, as in RCs 1 thru 15. The SNP pairs were tested for association with differential risk of PD under the five 2-dimensional split modes via the Fisher exact test, as in RCs 16 thru 19. All tests were placed in a single RC, given their low number.

**Quality control.** At the outset, a quality control procedure was applied to the Hamza et al. dataset that excluded the following SNPs from the entire analysis:

- SNPs with a p-value less than 10–4 under the Hardy-Weinberg test.

- SNPs with less than a 99.9% call rate.

The quality control was implemented using the program Plink®. A total of 748807 SNPs passed the quality control.

1. Bekris, L. M., Mata, I. F. & Zabetian, C. P. The genetics of Parkinson’s disease. *Journal of Geriatric Psychiatry and Neurology* 23 (4), 228–242 (2010).

2. Hawkes, C. H., Tredici, K. D. & Braak, H. Parkinson’s disease: a dual-hit paradigm. *Neurotoxicology* 106 (23), 9362–9367 (2009).

3. Lerner, A. & Bagic, A. Olfactory pathogenesis of idiopathic Parkinson disease. *The New England Journal of Medicine* 356 (1), 127–138 (2007).

4. Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *The New England Journal of Medicine* 363 (2), 166–176 (2010).

5. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the USA* 106 (23), 9362–9367 (2009).

6. Halliwell, B. & Gutteridge, J. M. C. (Oxford University Press, New York, 1999)

7. Beal, M. F. & Zabetian, C. P. The genetics of Parkinson’s disease. *Neurotoxicology* 106 (23), 9362–9367 (2009).

8. Moore, F. L., Jaruzelska, J., Dorfman, D. M. & Reijo-Pera, R. A. Identification of a novel gene, DZIP (DAZ-interacting protein) that encodes a protein that interacts with DAZ (deleted in azoospermia) and is expressed in embryonic stem cells and germ cells. *Genomics* 83, 834–843 (2004).

9. Edwards, T. L. et al. Genomewide association study confirms SNPs in SNCA and suggests a locus on 12q24 as the risk loci for Parkinson’s disease in the European population. *PLoS Genetics* 6 (5), e1002142 (2011).
51. Tsuboi, K. & Shults, C. W. Intrastriatal injection of Sonic Hedgehog reduces behavioral impairment in a rat model of Parkinson’s disease. *Experimental Neurology* **173**, 95–104 (2002).

52. Bak, M. et al. Mutation analysis of the Sonic hedgehog promoter and putative enhancer elements in Parkinson’s disease patients. *Molecular Brain Research* **126**, 207–211 (2004).

53. Merchant, A. A. & Matsui, W. Targeting hedgehog-a cancer stem cell pathway. *Clinical Cancer Research* **16**(12), 3130–3140 (2010).

54. Subramaniam, D., Ramalingam, S., Houchen, C. W. & Anant, S. Cancer stem cells: a novel paradigm for cancer prevention and treatment. *Mini-Reviews in Medicinal Chemistry* **10**, 359–371 (2010).

55. Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nature Genetics* **36**, 512–517 (2004).

56. Valente, A. X. C. N., Zischkau, J., Gao, Y. & Sarkar, A. GWAS heterogeneous population normalization via subject weight knock-down. *In preparation* (2011).

57. Chambers, S. M. et al. Hematopoietic Fingerprints: An Expression Database of Stem Cells and Their Progeny. *Cell Stem Cell* **1**, 578–591 (2007).

58. Keshava Prasad, T. S. et al. Human Protein Reference Database - 2009 Update. *Nucleic Acids Research* **37**(database issue), D767–72 (2009).

59. Ruemp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Research* **38**(database issue) D497–501 (2010).

60. Purcell, S. et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**(3), 559–575 (2007).

**Author contributions**

A.X.C.N.V. and Y.G. conceived the study; A.X.C.N.V. wrote the manuscript; J.H.S. and A.S. gave technical support and conceptual advice.

**Additional information**

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

How to cite this article: Valente, A.X.C.N., Shin, J.H., Sarkar, A. & Gao, Y. Rare coding SNP in DZIP1 gene associated with late-onset sporadic Parkinson’s disease. *Sci. Rep.* **2**, 256; DOI:10.1038/srep00256 (2012).