Human Genetic Variation and Parkinson’s Disease

Sun Ju Chung
Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Parkinson’s disease (PD) is a chronic neurodegenerative disorder with multifactorial etiology. In the past decade, the genetic causes of monogenic forms of familial PD have been defined. However, the etiology and pathogenesis of the majority of sporadic PD cases that occur in out-bred populations have yet to be clarified. The recent development of resources such as the International HapMap Project and technological advances in high-throughput genotyping have provided new basis for genetic association studies of common complex diseases, including PD. A new generation of genome-wide association studies will soon offer a potentially powerful approach for mapping causal genes and will likely change treatment and alter our perception of the genetic determinants of PD. However, the execution and analysis of such studies will require great care.

Key Words: Parkinson’s disease, Genome, Genetic variation, Genome-wide association study.

In 2001, two reference versions of the human genome were published. One human genome sequence was reported by the Human Genome Sequencing Consortium and reflected the assembly of sequences derived from numerous donors, whereas the other genome sequence, released by Celera Genomics, was a consensus sequence derived from five individuals. However, both versions of the genome sequence represented the human genome as a haploid sequence, and generic variation was not annotated. Therefore, many researchers have studied how genetic variants contribute to phenotype diversity and have conducted large-scale studies to identify and catalogue nucleotides that differ among individuals. Initial studies focused largely on understanding the range of patterns and frequencies of single nucleotide polymorphisms (SNPs). As their prevalence and contribution to human traits and biology were realized, several consortia were formed, and systemic studies were performed to improve our understanding of diverse human genomic variants.

The first complete human genome sequence of a single individual, Levy et al. was published in 2007. Shortly thereafter, the second complete genome sequence of an individual, Watson, determined with next-generation sequencing technology, was published. Subsequently, three additional genomes from anonymous individuals were sequenced: one Han Chinese (Asian), one Nigerian (African), and one Korean (Asian). Although these data have rapidly increased our knowledge of the various forms of human genetic variation, our understanding of the location and frequencies of structural variants across the genome is still limited. However, an enormous amount of effort is being expended to identify the common genetic variations that contribute to the development of common complex diseases.

This review is a general overview of human genetic variation and its contribution to Parkinson’s disease (PD).

Classes of Human Genetic Variation

Common vs. rare variants

Human genetic variants are typically referred to as either common or rare to denote the frequency of the minor allele in the human population. Common variants are synonymous with polymorphisms, defined as genetic variants with a minor allele frequency of at least 1% in the
population, whereas rare variants have a minor allele frequency of less than 1%.

**Single nucleotide polymorphisms**

A SNP is a single base change in the DNA sequence at a particular point compared with the "common" or "wild-type" sequence. SNPs are the most prevalent class of genetic variation among individuals. It has been estimated that the human genome contains at least 11 million SNPs, with about 7 million of these occurring with minor allele frequencies exceeding 5% and the remaining having minor allele frequencies between 1 and 5%.

**Structural variants**

Structural variants are defined as all base pairs that differ between individuals and that are not single nucleotide variants. These include insertion-deletion variants (indels), block substitutions, inversions of DNA sequences, and copy number differences. The technical ability to detect structural variants in the human genome has only recently emerged.

---

**Genetic Association Studies in Parkinson’s Disease**

Investigators conducting genetic association studies may target genes for investigation according to the known or postulated biology and previous results, an approach known as candidate gene association. As a large-scale candidate gene association study, Chung et al. investigated the association of common variants in *PARK* loci and related genes with PD susceptibility and age at onset in an outbred population (unpublished data: correspondence to Dr. Maraganore at NorthShore University Hospital System, Chicago, USA). They matched 1,103 PD cases from the upper Midwest, USA, individually with unaffected siblings (*n* = 654) or unrelated controls (*n* = 449) from the same region. Using a sequencing approach in 25 cases and 25 controls, SNPs in species-conserved regions of *PARK* loci and related genes were detected. Additional tag SNPs were selected from the HapMap. A total of 235 SNPs and two variable-number tandem repeats (VNTRs) in the *ATP13A2, DJ1, LRRK1, LRRK2, MAPT, Omi/HtrA2, PARK2, PINK1, SNCA, SNCB, SNCG, SPR*, and *UCHL1* genes were genotyped in all 2,206 subjects. Case-control analyses were performed to study the association with PD susceptibility, whereas case-only analyses were used to study the association with age at onset. Only *MAPT* SNP rs2435200 was associated with PD susceptibility after correcting for multiple testing [odds ratio (OR) = 0.74, 95% confidence interval (CI) = 0.64-0.86, uncorrected *p* < 0.0001, log additive model]; however, 16 additional *MAPT* variants, seven *SNCA* variants, and one *LRRK2, PARK2*, and *UCHL1* variant each had significant uncorrected *p*-values (Table 1). No significant associations were found for age at onset after correcting for multiple testing. These results confirmed the association of the *MAPT* and *SNCA* genes with PD susceptibility, but showed limited association of other *PARK* loci and related genes with PD.

Alternatively, we may screen the entire genome for association, an approach that has recently transformed the field of genetic association studies. Such a "genome-wide association study (GWAS)" is hypothesis-free, as there is no bias or presumptive list of candidate genes that are being tested. GWAS has greatly accelerated the pace of discovery of genetic association.

As testing so many potential genes simultaneously carries the risk of finding many spurious associations, genetic variants that seem to have strong or suggestive statistical signals in an initial GWAS need to be tested for replication in other large data sets or studies.

The boundaries between candidate gene studies and GWAS can become blurred, and the two types of study are not mutually exclusive.

---

**Genome-Wide Association Study in Parkinson’s Disease**

Six GWAS of PD have been published (Table 2). The study by Maraganore et al. included 775 PD cases and 775 matched controls. This study genotyped 198,345 informative genomic SNPs, and found that a SNP within the *semaforin 5A gene (SEMA5A)* had the lowest combined *p*-value (*p* = 7.62 × 10⁻⁸). The authors also reported some suggestive findings for *MAPT* and *SNCA*, as well as other *PARK* loci and related genes. However, none of the findings was significant after correcting for multiple testing. The study by Fung et al. examined more SNP markers (408,000 SNPs), but also failed to observe an association of any genetic variation with PD susceptibility after correcting for multiple testing; however, that study included only 276 PD cases and 276 unmatched controls. The study by Pankratz et al. enrolled 857 familial PD cases and 867 controls, and observed suggestive associations for the *GAK/DGKQ* region on chromosome 4 (additive model: OR = 1.69; *p* = 3.4 × 10⁻⁵), *MAPT* SNPs (recessive model: OR = 0.56; *p* = 2.0 × 10⁻³), and the *SNCA* SNPs (additive model: OR = 1.35; *p* = 5.5 × 10⁻⁵). Despite enriching their sample for genetic load (familial PD cases), none of the SNPs was significant after correcting for multiple testing.

Recently, three GWAS confirmed that common variants in *SNCA* and *MAPT* genes increase PD susceptibility. The study by Satake et al. (2,011 cases and 18,381 controls) reported strong associations at *SNCA* on 4q22 (rs11931074, *OR* = 1.37, *p* = 7.35 × 10⁻¹⁵), *PARK16* on 1q32 (OR = 1.52 × 10⁻⁵), *BST1* on 4q11 (OR = 3.94 × 10⁻⁵), and *LRRK2* on 12q12 (*p* = 2.72 × 10⁻⁹). The study by Simón-Sánchez et al. (5,074 cases and 8,551 controls) observed two strong association signals in the *SNCA* gene (rs2736990, OR = 1.23, *p* = 2.24 × 10⁻⁶) and
| Chromosome | SNP      | Position | Gene   | Type of variant | Allele | Minor allele frequencies<sup>a</sup> | Trend mode OR<sup>d</sup> | Trend test p value<sup>e</sup> |
|------------|----------|----------|--------|-----------------|--------|--------------------------------------|--------------------------|-----------------------------|
| 17         | rs2435200| 41,427,688| MAPT   | Intronic SNP    | A/G    | 0.372/0.422                          | 0.74 (0.64–0.86)         | <0.0001                     |
| 4          | rs2736990| 90,897,564| SNCA   | Intronic SNP    | C/T    | 0.490/0.470                          | 1.27 (1.09–1.47)         | 0.0017                      |
| 17         | rs1765212| 41,429,810| MAPT   | Synonymous      | C/T    | 0.164/0.196                          | 0.76 (0.63–0.91)         | 0.0035                      |
| 17         | rs4792891| 41,329,294| MAPT   | 5' UTR SNP      | G/T    | 0.284/0.320                          | 0.79 (0.68–0.93)         | 0.0036                      |
| 17         | rs1769160| 41,326,456| MAPT   | Intronic SNP    | G/T    | 0.164/0.196                          | 0.76 (0.64–0.92)         | 0.004                       |
| 17         | rs1052587| 41,458,449| MAPT   | 3' UTR SNP      | C/T    | 0.165/0.196                          | 0.77 (0.64–0.92)         | 0.0041                      |
| 17         | rs1757436| 41,464,049| MAPT   | Conserved       | A/G    | 0.164/0.197                          | 0.77 (0.64–0.92)         | 0.0041                      |
| 17         | rs1765154| 41,417,115| MAPT   | Conserved       | C/T    | 0.163/0.194                          | 0.76 (0.63–0.92)         | 0.0041                      |
| 17         | H1/H2    | –        | MAPT   | Intragenic VNTR | –      |                                    |                          | 0.0042                      |
| 17         | rs1777034| 41,329,480| MAPT   | Intronic SNP    | C/T    | 0.164/0.196                          | 0.77 (0.64–0.92)         | 0.0046                      |
| 17         | rs1052551| 41,424,761| MAPT   | Synonymous      | A/G    | 0.165/0.196                          | 0.77 (0.64–0.92)         | 0.0047                      |
| 17         | rs1215024| 41,467,460| MAPT   | Conserved       | A/G    | 0.164/0.196                          | 0.77 (0.64–0.92)         | 0.0048                      |
| 17         | rs1757428| 41,460,355| MAPT   | 3' UTR SNP      | C/T    | 0.165/0.196                          | 0.77 (0.64–0.92)         | 0.0049                      |
| 17         | rs9468   | 41,457,408| MAPT   | 3' UTR SNP      | C/T    | 0.164/0.196                          | 0.77 (0.64–0.92)         | 0.005                       |
| 17         | rs1765091| 41,395,527| MAPT   | 5' UTR SNP      | A/G    | 0.164/0.196                          | 0.77 (0.64–0.92)         | 0.0053                      |
| 4          | rs1372520| 90,976,528| SNCA   | Intronic SNP    | C/T    | 0.171/0.198                          | 0.77 (0.64–0.93)         | 0.0056                      |
| 17         | rs1694086| 41,459,672| MAPT   | 3' UTR SNP      | A/G    | 0.165/0.196                          | 0.77 (0.64–0.93)         | 0.0059                      |
| 17         | rs1052553| 41,429,726| MAPT   | Synonymous      | A/G    | 0.164/0.195                          | 0.78 (0.65–0.93)         | 0.0072                      |
| 4          | rs2572324| 90,897,821| SNCA   | Intronic SNP    | C/T    | 0.338/0.307                          | 1.24 (1.05–1.45)         | 0.009                       |
| 4          | rs3775423| 90,876,514| SNCA   | Intronic SNP    | C/T    | 0.399/0.381                          | 1.41 (1.09–1.82)         | 0.009                       |
| 4          | REP1     | 91,124,217| SNCA   | 5' UTR VNTR     | –      |                                    |                          | 0.0118                      |
| 4          | rs356186 | 90,924,387| SNCA   | Intronic SNP    | A/G    | 0.158/0.180                          | 0.78 (0.64–0.95)         | 0.0119                      |
| 12         | rs1748426| 38,984,953| LRRK2  | Intronic SNP    | A/G    | 0.083/0.102                          | 0.73 (0.57–0.93)         | 0.0128                      |
| 4          | rs1051700| 40,959,306| UCHL1  | Intronic SNP    | A/C    | 0.406/0.381                          | 1.19 (1.02–1.39)         | 0.0228                      |
| 4          | rs356218 | 90,856,033| SNCA   | Conserved       | A/G    | 0.367/0.342                          | 1.17 (1.01–1.37)         | 0.0419                      |
| 6          | rs1217440| 16,225,918| PARKIN | Conserved       | C/T    | 0.052/0.040                          | 1.43 (1.01–2.04)         | 0.0435                      |

<sup>a</sup> NCBI build 36 of the human genome, <sup>b</sup> the REP1 variant is a variable-number tandem repeat; common allele lengths are 259, 261, and 263 bp, <sup>c</sup> note that REP1 has three common alleles: the frequency of the 259-, 261-, and 263-bp allele was 0.24, 0.68, and 0.08, respectively. The frequency of the MAPT H1 and H2 haplotypes was 0.82 and 0.18, respectively, <sup>d</sup> the OR for REP1 was coded using the score test method, <sup>e</sup> only the MAPT gene variant rs2435200 remained significant after Bonferroni or permutation correction for multiple comparisons. SNP: single nucleotide polymorphism, UTR: untranslated region, VNTR: variable-number tandem repeat, OR: odds ratio, CI: confidence interval.
Table 2. Genome-wide association studies in Parkinson’s disease

| Authors                  | Journal (year)       | Ethnicty      | Chromosome | Gene       | Platform (SNP) | Replication sample | Replication sample | Exploratory sample | Exploratory sample | p-value | OR     | SEMA5A 5p15.2 | 1.7 | 7.62 × 10^{-10} | Patrojen (Illumina 98,345) | 2.5 | 2 × 10^{-10} | 332/332 North American White cases/controls | 1001.21 | 332/332 North American White cases/controls | 5.0 | 7 × 10^{-10} | 56/56 Asian cases/controls |
|-------------------------|----------------------|--------------|-------------|------------|----------------|--------------------|--------------------|--------------------|--------------------|---------|-------|----------------|-----|-----------------|--------------------------|------|--------------|-----------------------------|---------|---------------|-----------------------------|
| Maraganore et al.       | Am J Hum Genet (2005) | North American White                             | 10q11.2     | SNCA       | Illumina       | 857/867 Asian cases/controls | 3.36/14.573         | 172/172 European White cases/controls | 1.73/3.978          | 9802.521 | 9802.521 | 1.37 | 7 × 10^{-10} | 1.37 | 7 × 10^{-10} | 132/132 European White cases/controls | 12q12 | 132/132 European White cases/controls | 1.52 | 7 × 10^{-8} | 1.52 | 7 × 10^{-8} | 132/132 European White cases/controls |
| Fung et al.             | Ann Neurol (2006)     | North American White                             | 17q21.31    | MAPT       | Illumina       | 1,713/3,978 European White cases/controls | 1.75/1.745           | 1.75/1.745 European White cases/controls | 1.75/1.745          | 1,713/3,978 | 1,713/3,978 | 1.23 | 2 × 10^{-10} | 1.23 | 2 × 10^{-10} | 12q12 | 12q12 European White cases/controls | 1.52 | 7 × 10^{-8} | 1.52 | 7 × 10^{-8} | 12q12 European White cases/controls |
| Pankratz et al.         | Nat Genet (2009)      | Asian                                                 | 17q21.31    | PARK16     | Illumina       | 3,36/14.573         | 1.73/3.978          | 9802.521 | 9802.521 | 1.37 | 7 × 10^{-10} | 1.37 | 7 × 10^{-10} | 132/132 European White cases/controls | 12q12 | 132/132 European White cases/controls | 1.52 | 7 × 10^{-8} | 1.52 | 7 × 10^{-8} | 132/132 European White cases/controls |
| Simon-Sanchez et al.    | Nat Genet (2009)      | None                                                   | 17q21.31    | PARK16     | Illumina       | 1,713/3,978 European White cases/controls | 1.75/1.745           | 1.75/1.745 European White cases/controls | 1.75/1.745          | 1,713/3,978 | 1,713/3,978 | 1.23 | 2 × 10^{-10} | 1.23 | 2 × 10^{-10} | 12q12 | 12q12 European White cases/controls | 1.52 | 7 × 10^{-8} | 1.52 | 7 × 10^{-8} | 12q12 European White cases/controls |
| Edwards et al.          | Ann Hum Genet (2010)  | European White                                        | 17q21.31    | LRRK2      | Illumina       | 1,713/3,978 European White cases/controls | 1.75/1.745           | 1.75/1.745 European White cases/controls | 1.75/1.745          | 1,713/3,978 | 1,713/3,978 | 1.23 | 2 × 10^{-10} | 1.23 | 2 × 10^{-10} | 12q12 | 12q12 European White cases/controls | 1.52 | 7 × 10^{-8} | 1.52 | 7 × 10^{-8} | 12q12 European White cases/controls |

The GWAS approach still has substantial limitations. Enormous gaps remain in the ability to provide a biological explanation for why a genomic interval tracks with a complex trait. Although a tag SNP for a linkage disequilibrium (LD) bin is statistically associated with a trait, we have no idea of the precise variants in the bin that have a causal role in contributing to variation in the trait. Moreover, tag SNPs are in LD not only with other SNPs, but also with common structural variants, the majority of which have not yet been identified. The causative variants underlying GWAS test associations are likely to be regulatory rather than coding. Therefore, experiments should be conducted that simultaneously assay global gene expression and genome-wide variation in a large number of individuals to map genetic factors underlying differences in expression levels. These datasets may be valuable tools for identifying the causative variants and biological bases for many loci associated with a complex trait through GWAS.

The definite evaluation of the functions of these genetic variations awaits further investigation.

Limitation of Genome-Wide Association Study in Identifying Causative Variants

The GWAS approach still has substantial limitations. Enormous gaps remain in the ability to provide a biological explanation for why a genomic interval tracks with a complex trait. Although a tag SNP for a linkage disequilibrium (LD) bin is statistically associated with a trait, we have no idea of the precise variants in the bin that have a causal role in contributing to variation in the trait. Moreover, tag SNPs are in LD not only with other SNPs, but also with common structural variants, the majority of which have not yet been identified. The causative variants underlying GWAS test associations are likely to be regulatory rather than coding. Therefore, experiments should be conducted that simultaneously assay global gene expression and genome-wide variation in a large number of individuals to map genetic factors underlying differences in expression levels. These datasets may be valuable tools for identifying the causative variants and biological bases for many loci associated with a complex trait through GWAS.
genetic variant may differ from population to population.

**Issues in Genome-Wide Association Study**

We need to consider several issues to conduct GWAS properly. Genotyping error, genotype proportions (Hardy-Weinberg equilibrium), multiple comparisons, replication, population stratification, genetic risk prediction, and the manipulation and interpretation of information should be addressed adequately. Publication bias (negative results tend to be not published) is another big problem.

**Future Directions**

Although the discovery of GWAS signals is exciting, the amount of work required to achieve and confirm causal variants should not be underestimated. However, we predict that GWAS will identify common generic risk variants for PD and other common complex diseases. Future genomic technologies, including whole genome sequencing and genome-wide measures of epigenetic variability and somatic variation, are likely to change the treatment strategy of PD and alter our perception of the genetic determination of the disease. Therefore, clinicians will need to have solid knowledge of genetic principles and of the interpretation of complex genetic information.

**REFERENCES**

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291:1304-1351.
3. International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437:1299-1320.
4. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-861.
5. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, et al. Whole-genome patterns of common DNA variation in three human populations. Science 2005;307:1072-1079.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006;444:444-454.
7. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. Nature 2008;453:56-64.
8. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol 2007;5: e254.
9. Wheeler DA, Srinivasas M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872-876.
10. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, et al. The diploid genome sequence of an Asian individual. Nature 2008;456:60-65.
11. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-59.
12. Kim JJ, Ju YS, Park H, Kim S, Lee S, Yi JH, et al. A highly annotated whole-genome sequence of a Korean individual. Nature 2009;460:1011-1015.
13. Human Genome Structural Variation Working Group, Eichler EE, Nickerson DA, Altshuler D, Bowcock AM, Brooks LD, et al. Completing the map of human genetic variation. Nature 2007;447:161-165.
14. Maragatore DM, de Andrade M, Lesnick TG, Strain KJ, Farrar MJ, Rocca WA, et al. High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 2005;77:685-693.
15. Fung HC, Schoiz S, Matarin M, Simón-Sánchez J, Hernandez D, Britton A, et al. Genome-wide genotyping in Parkinson’s disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 2006;5:911-916.
16. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet 2009;124:593-605.
17. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson’s disease. Nat Genet 2009;41:1303-1307.
18. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbss JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. Nat Genet 2009;41:1308-1312.
19. Edwards TL, Scott WK, Almonte C, Burt A, Powell HE, Beecham GW, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann Hum Genet 2010;74:97-109.