Common Neurodegenerative Diseases: Dissection by Genome-wide Association

Vincenzo Bonifati, MD, PhD

Corresponding author
Vincenzo Bonifati, MD, PhD
Department of Clinical Genetics, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail: v.bonifati@erasmusmc.nl

Current Neurology and Neuroscience Reports 2007, **7:**425–427
Current Medicine Group LLC ISSN 1528-4042
Copyright © 2007 by Current Medicine Group LLC

Genome-wide association studies have already yielded great results. The success of this method in several common diseases leaves little doubt that it will aid in deciphering the genetic bases of the most common and devastating neurodegenerative disorders.

Introduction

In the past two decades, spectacular progress has been made in the identification of mutations underlying Mendelian (monogenic) neurodegenerative diseases. Unbiased, genome-wide linkage analysis and positional cloning in large pedigrees have been keys to success. In some instances, the identification of genes causing rare Mendelian forms provided important insights for understanding the pathogenesis of a disease in general.

However, monogenic models appear inadequate to explain most cases of common neurodegenerative disorders, such as Alzheimer's diseases (AD), Parkinson's disease (PD), or amyotrophic lateral sclerosis (ALS). A more plausible hypothesis is that such "complex" diseases are determined by several genetic factors (each contributing a part of the total genetic risk), as well as by nongenetic factors and their interplay. The number of genetic factors involved and their individual effect size (the genetic architecture), as well as their interaction, are likely to be specific not only for a certain disease but also for a certain population. In this context, another leading hypothesis is that the susceptibility to common diseases is determined by gene variants that are also common in the population.

Linkage analysis approaches are of limited validity for the dissection of complex diseases. A more promising approach is to directly compare large series of cases and unrelated controls using genetic markers (case-control study) [1,2]. If a certain genetic variant (allele A) is significantly more frequent (or more rare) among cases than among controls, the A variant (or another variant located nearby on the chromosome), said to be "in linkage disequilibrium" with A might be a risk (or a protective) factor for the disease. In the past, this approach has been extensively applied to candidate genes, but only rarely have the results been replicated.

Genome-wide Association Studies

Some fundamental scientific and technological developments of the past few years now allow the extension of the allelic association design at the level of the whole genome (genome-wide association [GWA]) [1,2]. The study design remains simple: compare the cases and controls for a very high number of DNA variants (in this case single nucleotide polymorphisms [SNPs]), aiming at best possible coverage of the genome. In the ideal situation, all the several million known SNPs would be interrogated. Although this is still unfeasible, recent progress allows some good compromises.

First, the completion of the Human Genome Project led to the identification and annotation of the required large number of SNP variants. Second, the International Hap Map project has provided information about the relatedness between each SNP and the surrounding variants so that one might focus on fewer SNPs, which are enough to capture most of the common variability in each of the discrete segments composing our genome (linkage disequilibrium blocks). These "intelligent" SNPs are also called "haplotype-tagging." The rarer SNPs (minor allele frequency < 1%–5%) are not relevant under the hypothesis of "common variants explaining common diseases" and can be ignored for the purposes of this approach, at least in a first phase.

Third, development in high-throughput genotyping platforms allows us to test the required number of SNPs in large numbers of cases and controls at affordable costs. The technology of DNA arrays currently allows hundred of thousands SNPs to be interrogated in a single experiment, with arrays for 1 million SNPs expected soon. Some arrays focus on SNPs located within genes (gene-centric) and some focus on SNPs coding for protein variants. In this way, one might concentrate on variants with a high likelihood of biologic effects. Other arrays capitalize on

haplotype-tagging SNPs, whereas others interrogate randomly spaced SNPs in the genome. An array of 500,000 SNPs might yield an average inter-marker distance of 5 kb, allowing the large "noncoding genome" to also be systematically investigated.

Most GWA studies will require thousands of cases and controls; therefore, the collaborative efforts of multiple research centers are crucial to success. GWA studies also require relevant data storage and computational power, and a growing list of dedicated statistical tools is being developed. The enormous number of tests performed raises unprecedented statistical challenges for the biologist. The threshold required to declare significant association at genome-wide level depends on the number of SNPs being tested, but it easily reaches orders of $P = 10^{-7}$ or 10^{-8}. There is debate on the best statistical treatment for GWA data, and the Bonferroni correction might be too stringent. Others prefer permutation or Bayesian statistics. In any case, due to a high chance of false-positive findings, it is of crucial importance that results are confirmed in independent, large replication studies [1,2].

Will the GWA study approach be successful to dissect complex diseases? The results of pioneer studies generate optimism. In different complex diseases, including bipolar disorder, type 1 and type 2 diabetes, coronary artery disease, rheumatoid arthritis, Crohn's disease, breast cancer, and others, large-scale GWA studies have reported the successful confirmation of many previously known disease-risk genes and the identification (including replication) of several novel ones [3–6].

However, easy generalizations might be dangerous. In each case, the success of the GWA approach will depend on the genetic architecture of the disease under study (which is largely unknown), and particularly how many genes are involved, how large the effect-size is of each gene, and how homogeneous is the disease etiology in the target population. The easiest scenario will be that of diseases with a small number of genes involved, each of moderate to large effect size (odds ratio > 2) and distributed homogeneously in the patients from the population. However, other diseases could be determined by many more genes, some of which may have very small effects, and the etiologic structure might be heterogeneous within the population. In the last scenario, even a very large GWA study might be underpowered. It is also important to remember that current array technology for the GWA study approach is largely based on frequent SNPs, according to the "common disease – common variants" hypothesis. However, if many rare risk alleles are involved (frequency of < 5% in the population), these will likely remain undetected. Lastly, it is unlikely that a SNP yielding an association signal is the biologically relevant variant; more likely, it indicates another variant located closely within the linkage disequilibrium block. The real size effects (odd ratios) might therefore be diluted when looking at the initial SNPs signals. In other words, the power of the study also depends on how efficiently a certain "detecting" SNP tags the biologically relevant variant.

GWA Studies for Neurodegenerative Diseases

Within this framework, the first GWA studies in common neurodegenerative diseases have been recently completed. Two studies examining PD [7,8] and one examining ALS [9] each involved less than 300 case-control pairs from the North American population and tested up to 400,000 SNPs. None of these studies identified association at the level required for genome-wide significance after Bonferroni correction. The SNPs giving the strongest association signals did not overlap in the two PD studies [7,8], and the most strongly associated SNPs identified in the first PD study [7] were not replicated in independent case-control samples [10,11]. It is important to consider that these GWA studies were underpowered to detect common variants of small effect sizes (odds ratios of < 1.5). However, one could be tempted to conclude that there is not a single common gene for PD or ALS with a moderate to large effect size. However, caution is warranted because in these studies [7–9] the genomic coverage cannot be considered complete even for common variants, the etiologic heterogeneity in the studied population was likely high, and rare variants would have remained undetected.

The great potential and also the limitations of current GWA studies are well illustrated by the first reported study of this kind in AD [12]. Five-hundred thousand SNPs were tested on 664 pathologically verified AD cases and 422 controls collected at multiple centers, mostly in the United States. The apolipoprotein E (ApoE) locus emerged with spectacular statistical significance (uncorrected P value = 1.06×10^{-39}, Bonferroni corrected P value for 500,000 tests = 5.30×10^{-34}). The GWA approach is therefore able to detect genetic determinants of complex, late-onset neurodegenerative disease, if these determinants exist, and if enough SNPs are studied in large, well-characterized samples. No other genome-wide significant signals (stronger that 10^{-8}) were detected, but this, once again, is not definitive evidence for the absence of other genes for AD. Remarkably, the ApoE locus was detected by a single SNP located 14 kb away from the ApoE biologically relevant variants. Thus, ApoE would also have been missed if that single SNP was not included in the array. This tells us that even 500,000 SNP arrays might not be enough for adequate genome coverage, especially in general (outbred) populations. Other genes for AD might exist, perhaps even of the same size effect of ApoE or greater.

In an extension of the same GWA study [13], the subset of cases and controls carrying the ApoE-ε4 risk allele was mined for additional genetic determinants. Consistent, genome-wide evidence for association was found at several SNPs, nominating the GRB-associated binding protein 2 (*GAB2*) gene as a novel risk gene for

late-onset AD, in combination with the ApoE-ε4 allele. Functional evidence supports a role for the Gab2 protein in a pathway that controls the phosphorylation of the protein tau, a well-known player in AD pathogenesis [13]. The association of the *GAB2* variants with ApoE-ε4–positive AD needs further confirmation, but a novel chapter in the genetics of AD might have been opened, also yielding novel therapeutic targets.

Conclusions

We are only at the beginning of the GWA era, and this approach is already yielding results. Of note, more and more GWA teams are making raw genotype data publicly available to the research community for further mining and meta-analysis. It is likely that many more GWA studies will be performed in the near future in large samples from well-characterized populations. The success obtained in several common diseases leaves little doubt that GWA studies will also contribute significantly in deciphering the genetic bases of the most common and devastating neurodegenerative disorders.

Acknowledgments

This work was supported by a grant from the "Internationaal Parkinson Fonds" (The Netherlands).

References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. Palmer LJ, Cardon LR: **Shaking the tree: mapping complex disease genes with linkage disequilibrium.** *Lancet* 2005, **366:**1223–1234.
2. Balding DJ: **A tutorial on statistical methods for population association studies.** *Nat Rev Genet* 2006, **7:**781–791.
3. Sladek R, Rocheleau G, Rung J, et al.: **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature* 2007, **445:**881–885.
4. Rioux JD, Xavier RJ, Taylor KD, et al.: **Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.** *Nat Genet* 2007, **39:**596–604.
5. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447:**661–678.
6. Easton DF, Pooley KA, Dunning AM, et al.: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, 447:1087–1093.
7. Maraganore DM, de Andrade M, Lesnick TG, et al.: **High-resolution whole-genome association study of Parkinson disease.** *Am J Hum Genet* 2005, **77:**685–693.
8. 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 Neurol* 2006, **5:**911–916.
9. Schymick JC, Scholz SW, Fung HC, et al.: **Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data.** *Lancet Neurol* 2007, **6:**322–328.
10. Elbaz A, Nelson LM, Payami H, et al.: **Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study.** *Lancet Neurol* 2006, **5:**917–923.
11. Myers RH: **Considerations for genomewide association studies in Parkinson disease.** *Am J Hum Genet* 2006, **78:**1081–1082.
12. Coon KD, Myers AJ, Craig DW, et al.: **A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease.** *J Clin Psychiatry* 2007, **68:**613–618.
13. Reiman EM, Webster JA, Myers AJ, et al.: **GAB2 alleles modify alzheimer's risk in APOE-e4 carriers.** *Neuron* 2007, **54:**713–720.