In the last 10 years, rapid advances in high-throughput genotyping technologies as well as an increase in computational power have enabled the large-scale assessment of genetic risk factors for complex diseases, by genotyping hundreds of thousands to millions of genetic variants—usually single nucleotide polymorphisms (SNPs)—in large cohorts. These studies have become widely known as genome-wide association studies (GWAS) and allow for an unbiased interrogation of the entire human genome regarding any conceivable trait with a heritable component.

This approach has been particularly productive for Parkinson disease (PD). Numerous GWAS, including 3 large meta-analyses, have identified up to 28 independent loci that modify disease risk and pointed to large meta-analyses, have identified up to 28 independent loci that modify disease risk and pointed to genome-wide association studies (GWAS) and allow for an unbiased interrogation of the entire human genome regarding any conceivable trait with a heritable component.  

This approach has been particularly productive for Parkinson disease (PD). Numerous GWAS, including 3 large meta-analyses, have identified up to 28 independent loci that modify disease risk and pointed to genome-wide association studies (GWAS) and allow for an unbiased interrogation of the entire human genome regarding any conceivable trait with a heritable component.  

In this issue of Neurology®, Beecham et al. argue that one of the reasons for this “missing heritability” may be the uncontrolled degree of heterogeneity of PD in studies relying on clinical diagnosis alone. To reduce heterogeneity, they performed a GWAS in which only PD cases with autopsy-confirmed Lewy body (LB) pathology, and controls with exclusion of PD neuropathology, were included. Despite the relatively small number of samples used, they found evidence suggesting that common variation in a small region of chromosome 1p32 is associated with LB PD with a p value just below genome-wide significance level $(6.18 \times 10^{-8})$ and an odds ratio of 0.64 for the protective allele, which is comparable with that found for SNCA and MAPT in other studies. This association peak lies within PARK10, which was originally identified in a set of Icelandic families, and subsequently associated with age at onset of PD. Of note, association at this locus has not been detected in any of the larger PD GWAS.  

Although spurious association driven by undetected population stratification remains a possibility until this finding has been replicated in other studies, another possible explanation for this discrepancy is that the PARK10 locus is only associated with a specific subgroup of PD and that its effect size is strong enough to yield statistical association when a selection of LB PD cases is made, but not when larger series of clinical PD cases are studied. This would also explain why it was initially identified in the Icelandic population, which is known to be largely homogeneous and thus possibly enriched for this particular subtype.  

One of the limitations of the study design is obviously the small sample size in comparison to other PD GWAS. The study failed to detect genome-wide association at any of the 28 loci associated with PD in the latest meta-analysis. Only SNCA, MAPT, and GAK—among the strongest associated loci in other GWAS—are nominally associated in this dataset, with p values of $5.51 \times 10^{-4}$, $7.76 \times 10^{-4}$, and $1.71 \times 10^{-4}$, respectively. These p values are within the expected range given the sample size of fewer than 500 cases, indicating that the reduction of heterogeneity did not affect these loci. The most likely explanation for this phenomenon is that SNCA, MAPT, and GAK are associated with a more common, or with multiple, variants of PD, while other loci, in this case PARK10, are specific for a subset of cases.  

So it is “lumping” or “splitting” again. Is one strategy better than the other? In fact, both GWAS strategies may be useful and complementary. While the approach presented by Beecham et al. allows the identification of risk loci for a specific group of PD cases, large clinical series overpower heterogeneity with larger numbers of samples and detect variants of a more general effect, with the downside of diluting out those variants associated with subgroups only. Performing GWAS in different ethnic populations may be another “splitting approach” that may prove useful in the future.
Another limitation of this study (and GWAS in general) is the relatively small effect size associated with the identified loci. This precludes useful individual disease prediction, and is a problem for risk classification and personalized medicine. Small effect sizes may simply be attributable to imperfect tagging of the causal variant by the SNPs on the array, because in most cases, it is not the actual risk variant itself that is detected, but rather a nearby SNP that is located close enough on the chromosome to be inherited together with the causal variant over many generations (a phenomenon called “linkage disequilibrium”). In addition, current arrays are not very effective in detecting structural variation (duplications, deletions, or inversions), which frequently occur in the human genome and have been associated with several conditions, such as schizophrenia or autism.\textsuperscript{9,10} To identify the actual risk variant(s), GWAS should be complemented with next-generation sequencing efforts capable of detecting rare variants of larger effect size and, depending on the read length and the algorithm used, structural variants.

Future applications of the data derived from this and other GWAS include the possibility to build genetic risk profiles for a disease of interest. These profiles have the potential to identify at-risk individuals and apply different therapeutic strategies depending on the specific genetic underpinnings of the disease in a given individual. As we learn from the work by Beecham et al., certain risks may only be associated with specific subtypes of PD. Thus, proper understanding of the complex genetic architecture of PD is needed. Clearly, the days in which we thought of PD as a single clinicopathologic entity are over.

**STUDY FUNDING**
No targeted funding reported.

**DISCLOSURE**
The authors report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

**REFERENCES**