HHS Public Access

For complex disease genetics, collaboration drives progress

Andrew B. Singleton and **Bryan J. Traynor**

Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA

A principal tool in the development of etiology-based therapies is the identification of the genetic determinants of disease. The logic of this approach rests on the expectation that knowledge of the genetic variants underlying disease will enhance our understanding of the molecular pathogenesis of disease and reveal viable points for therapeutic intervention. Because of the general adoption of this paradigm, genetics has been a dominant force in disease investigation over the past 25 years. Success in this area has come in waves, each driven by new methods and technology. Linkage (identifying segments of the genome that are associated with given traits), positional sequencing (sequencing specific candidate genes based on their location), genome-wide association (examining genetic variants in a genome-wide manner across individuals for association with a trait), and family-based next-generation sequencing have each produced startling new phases of genetic discovery. The study by Cirulli et al. (1), reported on page 1436 of this issue, marks an early success in another phase of gene discovery: the application of exome sequencing in large case-control cohorts to identify genetic factors involved in complex disease. This is one of the first successes in this area and provides insights into amyotrophic lateral sclerosis (ALS), as well as broader lessons for disease gene discovery.

The design of Cirulli et al.'s effort to identify new genes that associate with ALS was essentially that of a two-phase exome-wide association study. Several approaches were used in the initial stage, each centering on the identification of rare variants within the protein-coding regions of the genome (exomes) and a burden analysis aimed at prioritizing genes with an excess of variants in cases over controls. The authors focused on three different models for selecting rare variants from the discovery set of 2843 ALS cases and 4310 controls: investigating all nonsynonymous coding variants (when a nucleotide is substituted, thereby producing a different amino acid) and canonical splice variants; looking at nonsynonymous variants that are predicted to be damaging variants; and assessing loss-of-function variants. On the basis of these analyses, 51 genes were taken forward to a replication effort in an additional set of 1318 ALS cases and 2371 controls.

Although an attempt was made to exclude ALS cases with known mutations, it is notable that the top hit in the discovery effort was SOD1, a gene encoding superoxide dismutase 1, which is known to contain mutations that cause ALS. Other genes previously associated with ALS were also associated with disease in the analysis, including TAR DNA BINDING PROTEIN (TARDBP), OPTINEURIN (OPTN), VALOSIN CONTAINING PROTEIN (VCP), and SPASTIC PARAPLEGIA (SPG11).

Cirulli et al. performed a large number of analyses and identified a number of interesting genes potentially involved in risk for ALS. However, the most immediately important and

compelling finding is the identification of a new association between ALS and variants in TBK1, which encodes a noncanonical IκB kinase family member, TANK-binding kinase 1. The authors report an overall excess of rare TBK1 variants in cases under the "dominant not benign" model (dominant inherited variants that are predicted to be damaging to protein function), with a 0.19% allele frequency in controls versus 1.1% in cases (combined discovery and replication P value = $3.63 \times 10–11$). Although these data intuitively suggest that TBK1 variants play a role in ~0.9% of the ALS cases examined in the study, it is important to recognize that this does not suggest that TBK1 mutation is sufficient to cause disease, nor does it explain the cause of 0.9% of ALS cases. A great deal of additional work is required to establish whether disease-linked variants in this gene are risk variants, or causal mutations, or both (2).

From a functional perspective, the linkage of TBK1 to ALS is interesting, as it highlights the importance of autophagy and degradation of ubiquitinated proteins in motor neuron degeneration. TBK1 phosphorylates the proteins encoded by genes previously linked to ALS, OPTN and SEQUESTOSOME 1 (SQSTM1). Phosphorylation by Tbk1 enhances the ability of these proteins to shepherd ubiquitinated proteins to autophagosomes for destruction. Vcp and ubiquilin 2 (Ubqln2), also encoded by ALS-linked genes, are involved in later stages of the same cellular pathway, reinforcing the long-held view that ubiquitin-proteasome and autophagy pathways are central to ALS pathogenesis.

The genetics field has changed perhaps more than any other scientific discipline over the past two decades. This transformation has been focused not only on methodology, but also on a more fundamental shift in the way gene hunting is executed. Traditionally a highly competitive field, gene hunting has evolved into a collaborative endeavor in which consortia and open data sharing are central tenets. The type of collaboration typified by Cirulli et al. has become a requirement for success. Other examples include large collaborative efforts in Parkinson's disease and Alzheimer's disease (3, 4). Another key to the success of Cirulli et al. was their access to publicly available data. Reference data from the Exome Aggregation Consortium (5) and the 1000 Genomes Project (6) were key steps in their filtering approach. The availability of such control and population sequence data is an essential resource for progress in disease research, and one that promotes efficiency and collaboration.

As Cirulli et al. rightly point out, much remains to be done to more fully understand the genetics of ALS (see the figure). their work opens at least one new avenue into the investigation of molecular pathogenesis of disease, linking the protein product of TBK1 with known ALS proteins. It is likely that their genetic data will also serve as the foundation for further gene discovery in ALS. Given this, it is particularly laudable that the authors have made individual-level exome sequence data rapidly available to the broader scientific community.

References

1. Cirulli ET, et al. Science. 2015; 347:1436. [PubMed: 25700176]
2. Singleton A, Hardy J. Hum Mol Genet. 2011; 20:R158. [PubMed: 21875901]
3. Nalls MA, et al. Nat Genet. 2014; 46:989. [PubMed: 25064009]
4. Lambert JC, et al. Nat Genet. 2013; 45:1452. [PubMed: 24162737]

5. exac.broadinstitute.org

6. www.1000genomes.org

7. Renton AE, Chiò A, Traynor BJ. Nat Neurosci. 2014; 17:17. [PubMed: 24369373]

Author Manuscript

Author Manuscript

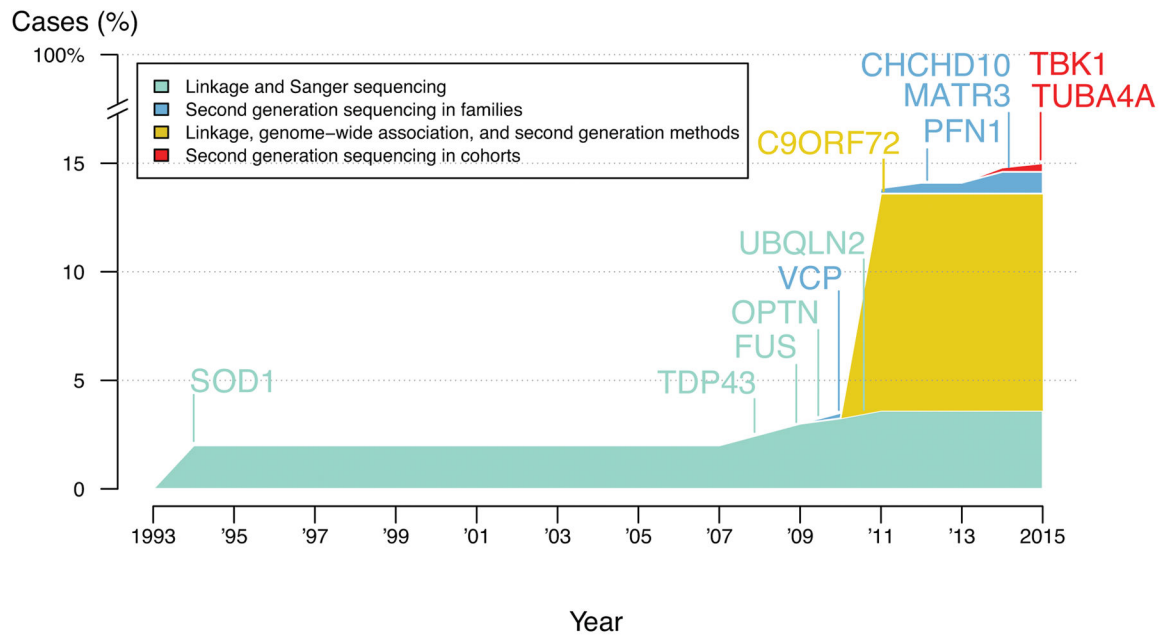Author Manuscript

Author Manuscript

Figure.