RESEARCH ARTICLES

ALS GENES

Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways

Elizabeth T. Cirulli,[1]* Brittany N. Lasseigne,[2]* Slavé Petrovski,[3] Peter C. Sapp,[4]
Patrick A. Dion,[5] Claire S. Leblond,[5] Julien Couthouis,[6] Yi-Fan Lu,[3] Quanli Wang,[3]
Brian J. Krueger,[3] Zhong Ren,[3] Jonathan Keebler,[7] Yujun Han,[7] Shawn E. Levy,[2]
Braden E. Boone,[2] Jack R. Wimbish,[2] Lindsay L. Waite,[2] Angela L. Jones,[2]
John P. Carulli,[8] Aaron G. Day-Williams,[8] John F. Staropoli,[8] Winnie W. Xin,[9]
Alessandra Chesi,[6] Alya R. Raphael,[6] Diane McKenna-Yasek,[4] Janet Cady,[10]
J. M. B. Vianney de Jong,[11] Kevin P. Kenna,[12] Bradley N. Smith,[13] Simon Topp,[13]
Jack Miller,[13] Athina Gkazi,[13] FALS Sequencing Consortium,† Ammar Al-Chalabi,[13]
Leonard H. van den Berg,[14] Jan Veldink,[14] Vincenzo Silani,[15] Nicola Ticozzi,[15]
Christopher E. Shaw,[13] Robert H. Baloh,[16] Stanley Appel,[17] Ericka Simpson,[17]
Clotilde Lagier-Tourenne,[18] Stefan M. Pulst,[19] Summer Gibson,[19] John Q. Trojanowski,[20]
Lauren Elman,[21] Leo McCluskey,[21] Murray Grossman,[22] Neil A. Shneider,[23]
Wendy K. Chung,[24] John M. Ravits,[25] Jonathan D. Glass,[26] Katherine B. Sims,[9]
Vivianna M. Van Deerlin,[20] Tom Maniatis,[27] Sebastian D. Hayes,[8,28] Alban Ordureau,[28]
Sharan Swarup,[28] John Landers,[4] Frank Baas,[11] Andrew S. Allen,[29] Richard S. Bedlack,[30]
J. Wade Harper,[28] Aaron D. Gitler,[6] Guy A. Rouleau,[5] Robert Brown,[4] Matthew B. Harms,[10]
Gregory M. Cooper,[2] Tim Harris,[8]‡ Richard M. Myers,[2]§ David B. Goldstein[3]§

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disease with no effective treatment. We report the results of a moderate-scale sequencing study aimed at increasing the number of genes known to contribute to predisposition for ALS. We performed whole-exome sequencing of 2869 ALS patients and 6405 controls. Several known ALS genes were found to be associated, and *TBK1* (the gene encoding TANK-binding kinase 1) was identified as an ALS gene. TBK1 is known to bind to and phosphorylate a number of proteins involved in innate immunity and autophagy, including optineurin (OPTN) and p62 (SQSTM1/sequestosome), both of which have also been implicated in ALS. These observations reveal a key role of the autophagic pathway in ALS and suggest specific targets for therapeutic intervention.

Amyotrophic lateral sclerosis (ALS) is a fatal, progressive neurodegenerative disease characterized by loss of motor neuron function for which there is no effective treatment or definitive diagnostic test (most cases are diagnosed clinically) (*1*). Approximately 10% of ALS cases are familial and inherited in an autosomal dominant, autosomal recessive, or X-linked mode; the remaining cases are apparently sporadic (*2*, *3*). Approximately 20 genes collectively explain a majority of familial cases, but these genes can explain only a minority (about 10%) of sporadic cases (*2*, *3*) (Table 1).

Protein and protein-RNA aggregates are a common feature of ALS pathology. These aggregates often include proteins encoded by genes that cause ALS when mutated, including those encoding SOD1, TARDBP (TDP-43), and FUS (*4*). Multiple genes (e.g., *C9orf72*, *GRN*, *VCP*, *UBQLN2*, *OPTN*, *NIPA1*, *SQSTM1*) in addition to *TARDBP* harbor variants pathogenic for TARDBP proteinopathy manifesting as ALS. This pathological TARDBP is part of a common pathway linked to neurodegeneration caused by diverse genetic abnormalities (*5*). Although murine models of ALS are limited, silencing certain ALS genes can cause regression of the disease phenotypes and clearance of the protein aggregates (*6*).

Identifying ALS genes

To identify genetic variants associated with ALS, we sequenced the exomes of 2869 patients with ALS and 6405 controls. We ran a standard collapsing analysis in which the gene was the unit of analysis, and we coded individuals according to the presence or absence of "qualifying" variants in each sequenced gene, where qualifying was defined according to one of six different genetic models: dominant coding, recessive coding, dominant not benign, recessive not benign, dominant loss of function (LoF), and recessive LoF (*7*). A total of 17,249 genes had more than one case or control sample with a genetic variant meeting the inclusion criteria for at least one of the genetic models tested (Fig. 1 and figs. S1 and S2). After correcting for multiple tests, the known

ALS gene *SOD1* ($P = 7.05 \times 10^{-8}$; dominant coding model) was found to have a study-wide significant enrichment of rare variants in ALS cases relative to controls, with qualifying variants in 0.871% of cases and 0.078% of controls. The genes *HLA-B*, *ZNF729*, *SIRPA*, and *TP53* were found to have a significant enrichment of variants in controls; however, these associations appear to be due to sequencing differences and to subsets of the controls having been ascertained on the basis of relevant phenotypes.

On the basis of their associations with ALS in a preliminary discovery-phase analysis that used 2843 cases and 4310 controls, we chose 51 genes (table S4) for analysis in an additional 1318 cases

[1]Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, NC 27708, USA. [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA. [3]Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA. [4]Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01655, USA. [5]Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3A 2B4, Canada. [6]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. [7]Duke University School of Medicine, Durham, NC 27708, USA. [8]Biogen Idec, Cambridge, MA 02142, USA. [9]Neurogenetics DNA Diagnostic Laboratory, Center for Human Genetics Research, Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA. [10]Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA. [11]Department of Genome Analysis, Academic Medical Center, Meibergdreef 9, 1105AZ Amsterdam, Netherlands. [12]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Republic of Ireland. [13]Department of Basic and Clinical Neuroscience, King's College London, Institute of Psychiatry, Psychology and Neuroscience, London SE5 8AF, UK. [14]Department of Neurology, Brain Center Rudolf Magnus, University Medical Centre Utrecht, 3508 GA Utrecht, Netherlands. [15]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan 20149, Italy, and Department of Pathophysiology and Transplantation, Dino Ferrari Center, Università degli Studi di Milano, Milan 20122, Italy. [16]Cedars Sinai Medical Center, Los Angeles, CA 90048, USA. [17]Houston Methodist Hospital, Houston, TX 77030, USA, and Weill Cornell Medical College of Cornell University, New York, NY 10065, USA. [18]Ludwig Institute for Cancer Research and Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA. [19]Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA. [20]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [21]Department of Neurology, Penn ALS Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [22]Department of Neurology, Penn Frontotemporal Degeneration Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA. [23]Department of Neurology, Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA. [24]Department of Pediatrics and Medicine, Columbia University, New York, NY 10032, USA. [25]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA. [26]Department of Neurology, Emory University, Atlanta, GA 30322, USA. [27]Department of Biochemistry & Molecular Biophysics, Columbia University, New York, NY 10027, USA. [28]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. [29]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27708, USA. [30]Duke ALS Clinic and Durham VA Medical Center, Durham, NC 27708, USA.
*These authors contributed equally to this work. †The full author list is included at the end of the manuscript. ‡Corresponding author. E-mail: tim.harris@biogenidec.com §These authors contributed equally to this work.

Table 1. Variants in previously described and currently reported ALS genes. Entries for reported inheritance model, reported FALS explained, and reported SALS explained are adapted from (*3, 4, 51*) with additional information from (*17–21, 52–54*). AD, autosomal dominant; AR, autosomal recessive; XD, X-linked. Best-model data are based on discovery data set for genes not included in the replication data set, and otherwise D = discovery, R = replication, and C = combined. Potential ALS cases explained are calculated as [(cases with variant in best model) − (controls with variant in best model)]; as case variants are risk factors for disease and may not be causal, this represents the potential percentage of cases for which this gene plays a role in disease.

Gene	Reported inheritance model	Reported FALS explained	Reported SALS explained	Best model with case enrichment in present study (*P* value)	Cases with variant in best model	Controls with variant in best model	Potential ALS cases explained
TBK1	N/A	N/A	N/A	Dom not benign (D = 1.12×10^{-5}; R = 5.78×10^{-7}; C = 3.60×10^{-11})	D = 23 (0.802%); R = 23 (1.745%); C = 46 (1.099%)	D = 12 (0.187%); R = 5 (0.211%); C = 17 (0.194%)	0.905%
NEK1	N/A	N/A	N/A	Dom LoF (D = 1.06×10^{-6}; R = 0.001; C = 3.15×10^{-9})	D = 25 (0.871%); R = 10 (0.759%); C = 35 (0.836%)	D = 6 (0.094%); R = 2 (0.084%); C = 8 (0.091%)	0.745%
SOD1	AR/AD	12%	1.50%	Dom coding (7.05×10^{-8})	25 (0.871%)	5 (0.078%)	0.793%
TARDBP	AD	4%	1%	Dom coding (2.93×10^{-6})	19 (0.662%)	6 (0.094%)	0.569%
OPTN	AR/AD	<1%	<1%	Dom not benign (D = 0.023; R = 0.002; C = 0.002)	D = 18 (0.627%); R = 8 (0.607%); C = 26 (0.621%)	D = 16 (0.25%); R = 4 (0.169%); C = 20 (0.228%)	0.393%
SPG11	AR	<1%	<1%	Dom LoF (D = 0.022; R = 0.183; C = 0.023)	D = 20 (0.697%); R = 5 (0.379%); C = 25 (0.597%)	D = 20 (0.312%); R = 7 (0.295%); C = 27 (0.308%)	0.289%
VCP	AD	1%	1%	Dom coding (0.022)	8 (0.279%)	4 (0.062%)	0.216%
HNRNPA1	AD	<1%	<1%	Dom coding (0.103)	6 (0.209%)	5 (0.078%)	0.131%
ATXN2*	AD	<1%	<1%	Rec coding (0.205)	4 (0.139%)	2 (0.031%)	0.108%
ANG	AD	<1%	<1%	Dom LoF (0.217)	2 (0.070%)	1 (0.016%)	0.054%
CHCHD10	AD	<1%	<1%	Dom coding (0.225)	2 (0.070%)	0 (0%)	0.070%
SIGMAR1	AR	<1%	<1%	Dom LoF (0.226)	1 (0.035%)	0 (0%)	0.035%
FIG4	AR/AD	<1%	<1%	Dom LoF (0.232)	9 (0.314%)	12 (0.187%)	0.126%
SS18L1	AD	<1%	<1%	Dom LoF (0.241)	1 (0.035%)	0 (0%)	0.035%
GRN	AD	<1%	<1%	Dom not benign (0.357)	14 (0.488%)	24 (0.375%)	0.113%
SETX	AD	<1%	<1%	Rec not benign (0.379)	3 (0.105%)	4 (0.062%)	0.042%
HNRNPA2B1	AD	<1%	<1%	Dom not benign (0.423)	3 (0.105%)	4 (0.062%)	0.042%
SQSTM1	AD	1%	<1%	Dom LoF (0.546)	1 (0.035%)	2 (0.031%)	0.004%
TAF15	AR/AD	<1%	<1%	Rec not benign (0.555)	2 (0.070%)	1 (0.016%)	0.054%
FUS	AR/AD	4%	1%	Dom LoF (0.612)	2 (0.070%)	3 (0.047%)	0.023%
ALS2	AR	<1%	<1%	Rec coding (0.655)	2 (0.070%)	4 (0.062%)	0.007%
VAPB	AD	<1%	<1%	Dom not benign (0.688)	3 (0.105%)	5 (0.078%)	0.027%
NEFH	AD	<1%	<1%	Dom coding (0.673)	22 (0.767%)	37 (0.578%)	0.189%
C9orf72*	AD	40%	7%	Dom not benign (1.000)	4 (0.139%)	7 (0.109%)	0.030%
CHMP2B	AD	<1%	<1%	Rec coding (1.000)	1 (0.035%)	1 (0.016%)	0.019%
MATR3	AD	<1%	<1%	Dom coding (1.000)	19 (0.662%)	35 (0.546%)	0.116%
PFN1	AD	<1%	<1%	Rec coding (1.000)	9 (0.314%)	15 (0.234%)	0.080%
PRPH	AD	<1%	<1%	Dom LoF (1.000)	1 (0.035%)	2 (0.031%)	0.004%
SPAST	AD	<1%	<1%	Dom coding (1.000)	6 (0.209%)	12 (0.187%)	0.022%
TUBA4A	AD	1%	<1%	Dom not benign (1.000)	2 (0.070%)	3 (0.047%)	0.023%
ELP3†	Allelic	<1%	<1%	Rec coding (1.000)	0 (0%)	0 (0%)	0%
DAO†	AD	<1%	<1%	Rec coding (1.000)	0 (0%)	0 (0%)	0%
DCTN1†	AD	<1%	<1%	Dom coding (0.668)	32 (1.115%)	76 (1.187%)	0%
EWSR1†	AD	<1%	<1%	Dom coding (0.375)	10 (0.349%)	28 (0.437%)	0%
GLE1†	AD	<1%	<1%	Rec LoF (1.000)	0 (0%)	0 (0%)	0%
UBQLN2†	XD	<1%	<1%	Dom LoF (1.000)	0 (0%)	0 (0%)	0%

*Because the known causal variants are repeat expansions that are not generally captured by next-generation sequencing, no case enrichment is expected. †No model showed case enrichment.

and 2371 controls (sequenced using either whole exome or custom capture) (7). This analysis definitively identified TANK-binding kinase 1 (*TBK1*) as an ALS gene, with a discovery association $P = 1.12 \times 10^{-5}$, a replication $P = 5.78 \times 10^{-7}$, and a combined $P = 3.60 \times 10^{-11}$ (dominant not benign model). In the combined data set, dominant not benign variants in this gene were found in 1.099% of cases and 0.194% of controls, with LoF variants occurring in 0.382% of cases and 0.034% of controls.

Analysis of clinical features

We also performed gene-based collapsing analyses to identify genes associated with patients' age of onset, site of onset, and survival time. No genes showed genome-wide significant association with any of these features. When applying multiple-test correction to only known ALS predisposition genes and *TBK1*, we found that D-amino acid oxidase (*DAO*) significantly correlated with survival times, with variant carriers showing shorter survival times ($P = 5.5 \times 10^{-7}$, dominant coding model). In mice, DAO is required for the clearance of D-serine. Indeed, D-serine levels are increased in SOD1 mutant mice and in spinal cords from people with familial ALS (FALS) or sporadic ALS (SALS) (8, 9). Known FALS mutations seem to reduce DAO activity, leading to neurotoxicity (10).

ALS patients with mutations in more than one known ALS gene are reported to have a younger age of onset (11). We did not replicate this finding in our data set. Without sequence data for known *C9orf72* carriers (by far the most common ALS variant) and without information about *ATXN2* expansions, we cannot adequately assess such an association.

Associations with other ALS genes

Although *SOD1* was the only previously known ALS gene to attain a genome-wide significant association in our data, many other known ALS genes showed strong associations. For example, rare coding variants in *TARDBP* occurred in 0.662% of the ALS cases and 0.094% of controls in our study, ranking this gene second to *SOD1* genome-wide under the dominant coding model (discovery data set, $P = 2.93 \times 10^{-6}$; Fig. 1). Consistent with previous reports and the ALS pathogenic *TARDBP* "DM" variants in the Human Genome Mutation Database (HGMD) (3, 12), we observed that the implicated nonsynonymous variants were generally predicted to have a benign effect on protein structure and function by PolyPhen-2 (13) and were clearly concentrated in the 3′ protein-coding portion of the gene in the ALS cases relative to controls (Fig. 2).

In the case of *OPTN*, we observed rare damaging variants in 0.621% of ALS cases and 0.228% of controls (combined dominant not benign model, $P = 0.002$). The greatest enrichment was for LoF variants, which occur in 0.334% of cases and 0.114% of controls (combined dominant LoF model, $P = 0.013$). Whereas the initial studies of *OPTN* in ALS found a role in only a few families with a recessive genetic model, subsequent studies iden-

tified dominant mutations (14, 15). Here, dominant variants appeared to make a substantial contribution to sporadic disease.

Finally, we also observed a modest excess of qualifying variants in *VCP* (discovery dominant coding model, $P = 0.022$) and of LoF variants in *SPG11* (combined dominant LoF model, $P = 0.023$). The former was driven by variants near the cell division protein 48 domain 2 region, where variants were found in 71% of case variants as compared to 25% of control variants (Fig. 2). Similar to *OPTN*, *SPG11* has previously been reported as a cause of recessive juvenile ALS, but our data indicate that it could play a broader role because these cases did not have early onset (16).

We did not identify even a nominal association with other previously reported ALS genes in our data set, including the recently reported *TUBA4A*, *MATR3*, *GLE1*, *SS18L1*, and *CHCHD10* (Table 1) (17–21). A fraction of our samples were genetically screened for some of the known genes and positive cases had been removed before sequencing, which may partially explain the lack of signal (7). Additionally, a comparison with genes implicated in a recent assessment of the role of 169 previously reported and candidate ALS genes in 242 sporadic ALS cases and 129 controls showed no overlap beyond signals for *SOD1* and *SPG11* (22). Some of these previously studied genes are mutated so rarely that even the sample size

presented here is not sufficient to detect causal variant enrichment, while others simply show comparable proportions of rare variants among cases and controls. Finally, certain genes did not show associations owing to the nature of the causal variation: Most known pathogenic variants in *ATXN2* and *C9orf72* are repeats that cannot be identified in our sequence data.

TBK1, autophagy, and neuroinflammation

Previous studies have implicated both *OPTN* (optineurin) (23) and *SQSTM1* (p62) (24) in ALS. The current study implicates *TBK1* and suggests that *OPTN* is a more important disease gene than previously recognized. These genes play important and interconnected roles in both autophagy and inflammation, emerging areas of interest in ALS research (Fig. 3) (25–27). Mutations in *SOD1*, *TARDBP*, and *FUS* result in the formation of protein aggregates that stain with antibodies to SQSTM1 and OPTN (28). These aggregates are thought to lead to a cargo-specific subtype of autophagy involved in the degradation of ubiquitinated proteins through the lysosome (29). The SQSTM1 and OPTN proteins function as cargo receptors, recruiting ubiquitinated proteins to the autophagosome via their LC3 interaction region (LIR) motifs. TBK1 binds and phosphorylates both OPTN and SQSTM1 (30–32) and enhances the binding of OPTN to the essential autophagosome
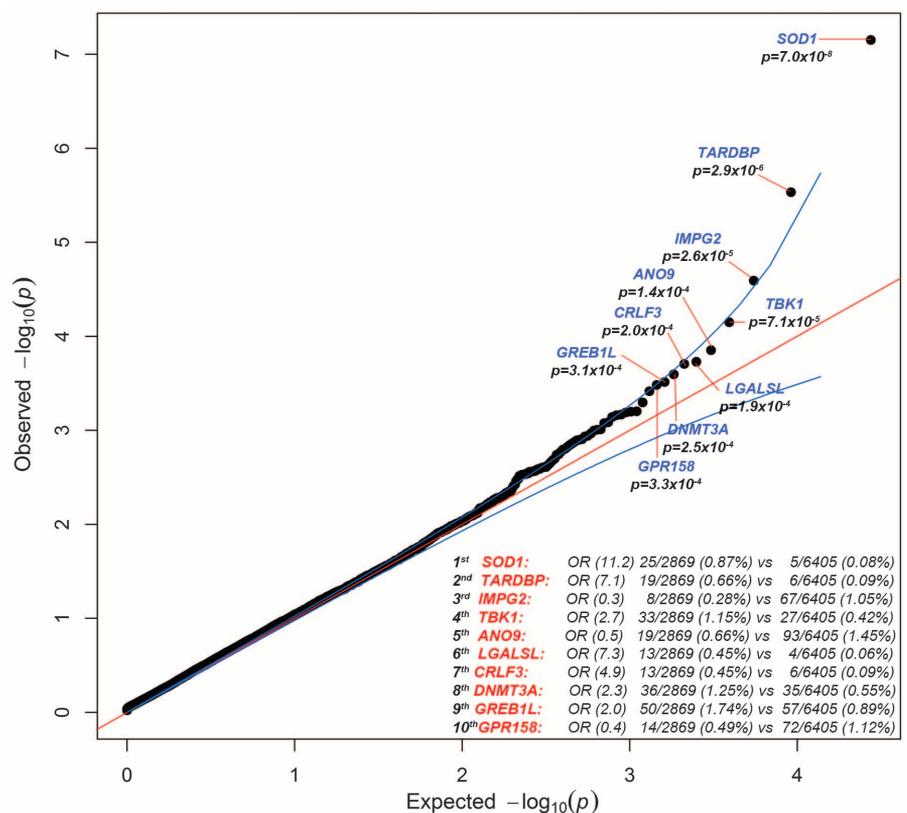

Fig. 1. Quantile-quantile plot of discovery results for dominant coding model. Results for the analysis of 2869 case and 6405 control exomes are shown; 16,491 covered genes passed quality control with more than one case or control carrier for this test. The genes with the top 10 associations are labeled. The genomic inflation factor λ is 1.060. The association with *SOD1* passed correction for multiple tests.

protein LC3, thereby facilitating the autophagic turnover of infectious bacteria coated with ubiquitinated proteins, a specific cargo of the OPTN adaptor (33). Considering that TBK1 colocalizes with OPTN and SQSTM1 in autophago-somes, it is possible that all three proteins associate with protein aggregates in ALS (33). Indeed, TBK1 appears to play a role in the degradation of protein aggregates by autophagy (34). Additionally, OPTN also functions in the autophagic turnover of damaged mitochondria via the Parkin ubiquitin ligase pathway (35). Finally, VCP, encoded by another gene with mutations that cause ALS, also binds to ubiquitinated protein aggregates. VCP and autophagy are required for the removal of stress granules (dense cytoplasmic protein-RNA aggregates), which are a common feature of ALS pathology (36). Thus, OPTN, SQSTM1, VCP, and TBK1 may be critical components of the aggresome pathway required for the removal of pathological ribonucleoprotein inclusions (37). It appears that defects in this pathway can be selective for motor neuron death, in some cases apparently sparing other neuronal cell types.

In addition to their roles in autophagy, OPTN, SQSTM1, and TBK1 all function in the NF-κB pathway (Fig. 3) (27, 38). For example, IκB kinases (IKKα and IKKβ) phosphorylate the IKK-related kinase TBK1, which in turn phosphorylates the IκB kinases, suppressing their activity in a negative autoregulatory feedback loop (39). TBK1 also phosphorylates and activates the transcription factor IRF3 (40–42) and the critical innate immunity signaling components MAVS and STING (43). The coordinate activation of NF-κB and IRF3 turns on the transcription of many inflammatory genes, including interferon-β (44). The innate immune pathway and neuroinflammation in general are thought to be important aspects of neurodegenerative disease progression (45). Thus, pathogenic variants in OPTN, SQSTM1, or TBK1 would be expected to lead to defects in autophagy and in key innate immunity signaling
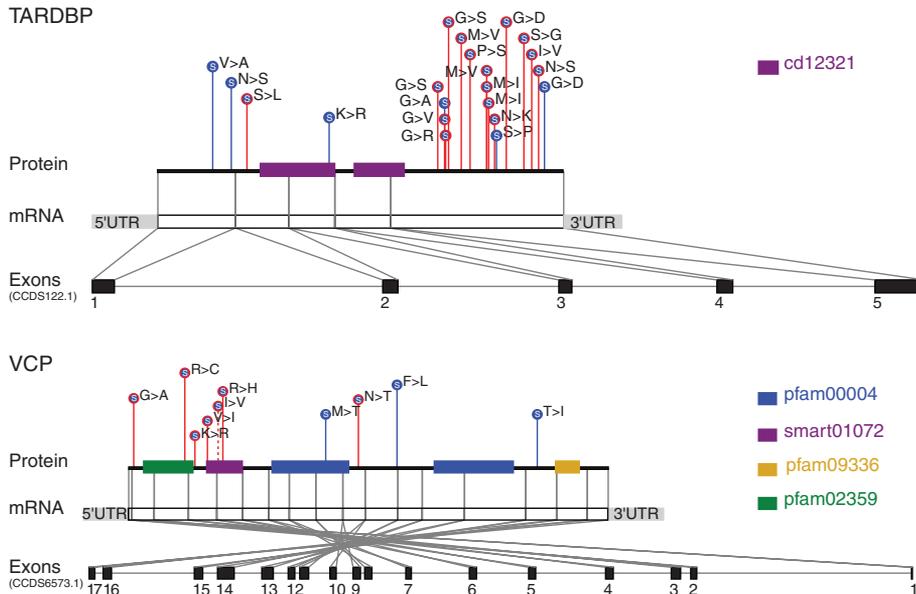
Fig. 2. Variants in TARDBP and VCP. Dominant coding variants are shown in TARDBP and VCP (discovery data set). Case variants are enriched at the 3′ end of the gene in TARDBP and near the cell division protein 48 domain 2 region in VCP. LoF variants are filled in red, and nonsynonymous variants are filled in blue. Case variants are shown with red lines, control variants are shown with blue lines, and variants found in both cases and controls are shown with dashed lines.

Fig. 3. Genes and pathways implicated in ALS disease progression. Genes known to have sequence variants that cause or are associated with ALS are indicated in red. These mutations can lead to the formation of protein or protein-RNA aggregates that appear as inclusion bodies in post mortem samples from both familial and sporadic ALS patients. Some of the mutant proteins adopt "prion-like" structures (see text for more detail). The misfolded proteins activate the ubiquitin-proteasome autophagy pathways to remove the misfolded proteins. Ubiquilin2 (UBQLN2) functions in both the ubiquitin-proteasome and autophagy pathways. TBK-1 (boxed) lies at the interface between autophagy and inflammation and associates with and phosphorylates both optineurin and p62, which can in turn enhance inflammation. ISG15 is induced by type I interferons (α and β) and interacts with p62 and HDAC6 in the autophagosome.
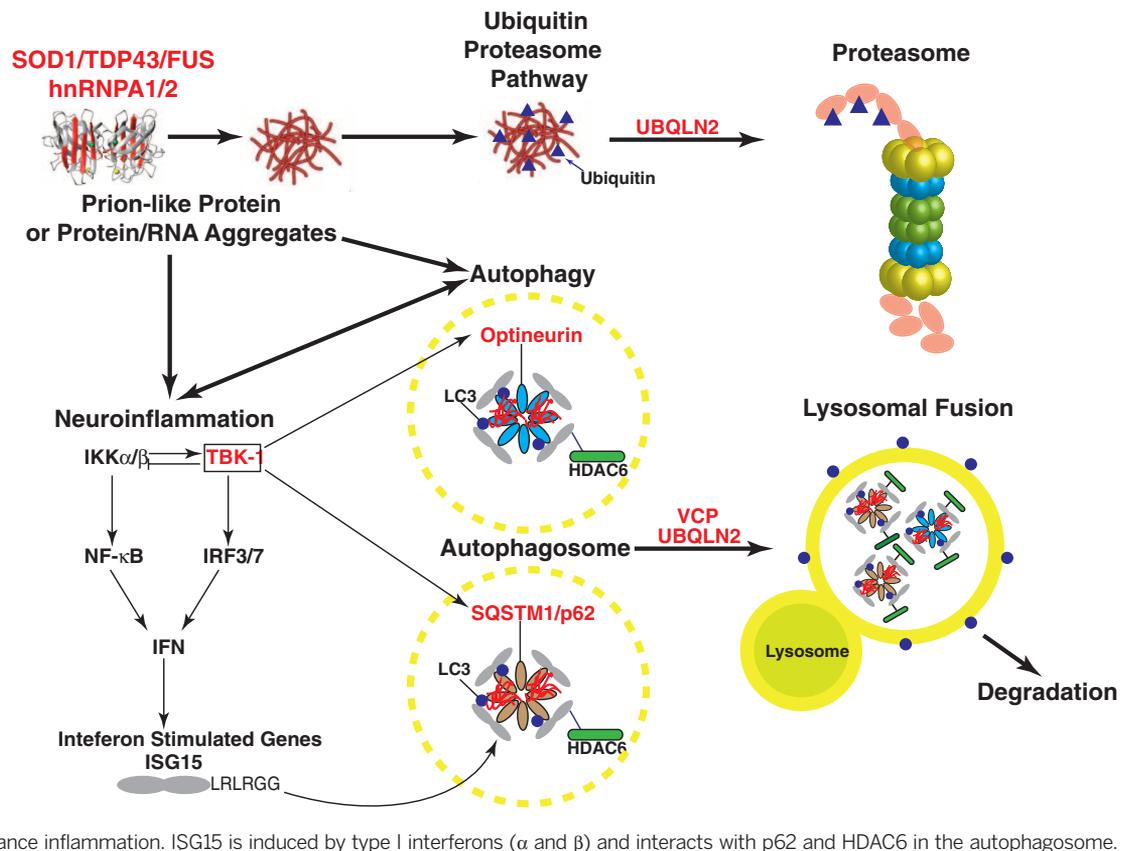
pathways. Mutations in these genes might therefore interfere with the normal function of these pathways in maintaining normal cellular riboproteostasis (*37*).

The simple observation of enrichment of qualifying variants in patients shows that some of the variants we have identified influence risk of disease. We cannot determine, however, the extent to which they may interact with any other variants or other risk factors in determining risk. We therefore focus on estimating the proportion of patients in which variants in the relevant genes either cause or contribute to disease by subtracting the proportion of controls with qualifying variants in a gene from the proportion of cases with such variants. Although we saw no enrichment of case variants in *SQSTM1*, variants in *OPTN* and *TBK1* were estimated to explain or contribute to 1.30% of cases in our data set when taken together (combined data set), suggesting an important subgroup of patients that may have a common biological etiology. No individual ALS cases had qualifying variants in more than one of these three genes.

The case variants found in *OPTN* and *TBK1* were largely heterozygous and LoF, which suggests that a reduction in trafficking of cargo through the autophagosomal pathway or disruption of autophagosomal maturation may promote disease. Although the most obvious enrichment of case variants in *TBK1* was seen for LoF, there was also a signal for nonsynonymous variants, which were found in 1.027% of cases and 0.365% of controls (combined data set). If any of these nonsynonymous variants are selective LoF for specific TBK1 functions as opposed to complete LoF variants, they may help elucidate which TBK1 function is most relevant to disease. We

did not observe any clear concentration of qualifying variants in any part of the *TBK1* gene (Fig. 4).

NEK1 associates with ALS2 and VAPB

Although no additional genes showed sufficiently strong evidence to be definitively declared disease genes at this point, some of the strongly associated genes identified here may be securely implicated as sample sizes increase. One gene of particular interest is *NEK1* (NIMA-related kinase 1). This gene just reached experiment-wide significance in the combined discovery and replication data sets (discovery $P = 1.06 \times 10^{-6}$, replication $P = 0.001$, combined $P = 3.15 \times 10^{-9}$; dominant LoF model). In the combined data set, dominant LoF variants in this gene were found in 0.836% of cases and 0.091% of controls (fig. S3). Additional studies are needed to confirm this suggestive association. Even if LoF variants in this gene do predispose to ALS, their relatively high prevalence in our controls and in public databases indicates that such variants have quite low penetrance, given that the lifetime prevalence of ALS is approximately 0.2%.

NEK1 is a widely expressed multifunctional kinase linked to multiple cellular processes, but it has not been linked to ALS. In an unbiased proteomic search for NEK1-interacting proteins in human embryonic kidney (HEK) 293T cells, we discovered an interaction between NEK1 and two widely expressed proteins previously found to be mutated in familial ALS: (i) the RAB guanine nucleotide exchange factor ALS2 (also called Alsin) involved in endosomal trafficking, and (ii) the endoplasmic reticulum protein VAPB involved in lipid trafficking to the plasma membrane (fig. S4, A and B, and table S5) (*46*). ALS2 reciprocally associated with NEK1 in HEK293T

cells, and both ALS2 and VAPB associated with NEK1 reciprocally in mouse neuronal cell line NSC-34 (fig. S4, C to E).

Other top genes showing interesting association patterns but not obtaining genome-wide significance included *ENAH*, with variants in 0.263% of cases and 0.011% of controls (combined data set) (discovery $P = 1.82 \times 10^{-5}$, replication $P = 0.133$, combined $P = 9.58 \times 10^{-6}$; recessive not benign model); *CRLF3*, with variants in 0.453% of cases and 0.094% of controls (discovery $P = 0.0002$; dominant coding model); *DNMT3A*, with variants in 1.003% of cases and 0.456% of controls (combined data set) (discovery $P = 0.0002$, replication $P = 0.261$, combined $P = 0.0002$; dominant not benign model); and *LGALSL*, with variants in 0.382% of cases and 0.068% of controls (combined data set) (discovery $P = 0.0002$, replication $P = 0.356$, combined $P = 0.0002$; dominant coding model).

Conclusions

Our results implicate *TBK1* as an ALS gene, thereby providing insight into disease biology and suggesting possible directions for drug screening programs. We have also provided evidence that *OPTN* plays a broader role in ALS than previously recognized. Both TBK1 and OPTN are involved in autophagy, with TBK1 possibly playing a crucial role in autophagosome maturation as well as the clearance of pathological aggregates (*31*, *34*). These observations highlight a critical role of autophagy and/or inflammation in disease predisposition. It is also noteworthy that many drugs have been developed that act on TBK1-mediated pathways owing to their role in tumor cell survival (*47*) and can therefore be used to investigate the effects of drug-dependent loss of function of the kinase.

We also provide a large genetic data set for ALS, which suggests other possible ALS genes and provides a substantial collection of pathogenic variants across ALS genes (for genotype counts for all genes for all cases from this study, see alsdb.org). After removing the number of variants expected to be seen on the basis of frequencies of rare variants in controls, we identify more than 70 distinct pathogenic mutations across *SOD1*, *OPTN*, *TARDBP*, *VCP*, *SPG11*, and *TBK1* that can be used in future efforts to functionally characterize the role of these ALS genes. The identification of *TBK1* and the expanded role for *OPTN* as ALS genes reinforce the growing recognition of the central role of autophagy and neuroinflammation in the pathophysiology of ALS (Fig. 3). These pathways appear to be activated in response to the formation of various types of cellular inclusions, the most prominent of which appear to be ribonucleoprotein complexes; this has led to the proposal that the control of protein misfolding (proteostasis) or ribonucleoprotein/RNA misfolding ("ribostasis") plays a key role in neurodegenerative diseases (*37*). Cellular ribonucleoprotein inclusions can be caused by mutations in low-complexity sequence domains or "prion" domains of RNA binding proteins (*37*, *48*) and can be exacerbated by
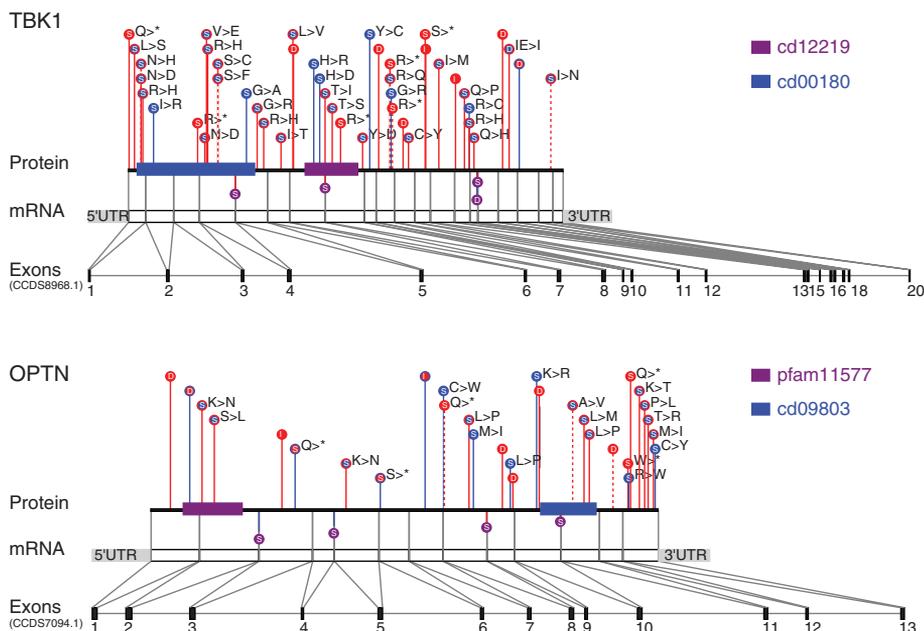

Fig. 4. Variants in *TBK1* and *OPTN*. Dominant not benign variants are shown in *TBK1* and *OPTN* (combined data sets). LoF variants are filled in red, nonsynonymous variants are filled in blue, and splice variants are filled in purple and shown below the protein line. Case variants are shown with red lines, control variants are shown with blue lines, and variants found in both cases and controls are shown with dashed lines.

mutations that diminish the autophagy pathway. Remarkably, a hallmark of motor neuron pathology in >95% of sporadic and familial ALS patients is the formation of TARDBP inclusions, which suggests that defects in ribostasis are a common feature of the disease (5, 49). The prominence of this disease mechanism in ALS has been proposed to be the consequence of the normal function of low-complexity domains in RNA binding proteins in the assembly of functional "RNA granules" such as P-bodies and stress granules [see (37) for detailed discussion].

Our exome sequencing study has identified variants that definitively predispose humans to a sporadic, complex human disease. Larger exome sequencing studies may reveal identifiable roles for genes that have not yet achieved significant associations. There is reason for optimism that such studies will begin to fill in an increasingly complete picture of the key genes implicated in ALS, providing multiple entry points for therapeutic intervention (Fig. 3). It is also likely that whole-genome sequencing (especially with longer reads) will prove of particular value in ALS, given that there are many causal variants refractory to identification by contemporary exome sequencing. Finally, we note that effective studies will depend critically on the control samples available. For example, we used the recently released ExAC data set of >60,000 samples to focus on extremely rare variants in our samples (50). Well-characterized, publicly available control sample sets will be of great importance for further discovery of variants associated with complex traits, in particular for whole-genome sequencing studies.

REFERENCES AND NOTES

1. L. Poppe, L. Rué, W. Robberecht, L. Van Den Bosch, *Exp. Neurol.* **262**, 138–151 (2014).
2. S. Chen, P. Sayana, X. Zhang, W. Le, *Mol. Neurodegener.* **8**, 28 (2013).
3. A. E. Renton, A. Chiò, B. J. Traynor, *Nat. Neurosci.* **17**, 17–23 (2014).
4. A. Al-Chalabi *et al.*, *Acta Neuropathol.* **124**, 339–352 (2012).
5. M. Neumann, *Rev. Neurol.* **169**, 793–798 (2013).
6. R. A. Smith *et al.*, *J. Clin. Invest.* **116**, 2290–2296 (2006).
7. See supplementary materials on *Science* Online.
8. J. Sasabe *et al.*, *EMBO J.* **26**, 4149–4159 (2007).
9. M. Thompson *et al.*, *J. Neurochem.* **120**, 598–610 (2012).
10. P. Paul, J. de Belleroche, *Amino Acids* **43**, 1823–1831 (2012).
11. J. Cady *et al.*, *Ann. Neurol.* **77**, 100–113 (2015).
12. P. D. Stenson *et al.*, *Hum. Mutat.* **21**, 577–581 (2003).
13. I. A. Adzhubei *et al.*, *Nat. Methods* **7**, 248–249 (2010).
14. A. Iida *et al.*, *Neurobiol. Aging* **33**, 1843.e19–1843.e24 (2012).
15. M. van Blitterswijk *et al.*, *Neurobiol. Aging* **33**, 1016.e1–1016.e7 (2012).
16. H. Daoud *et al.*, *Neurobiol. Aging* **33**, 839.e5–839.e9 (2012).
17. B. N. Smith *et al.*, *Neuron* **84**, 324–331 (2014).
18. J. O. Johnson *et al.*, *Nat. Neurosci.* **17**, 664–666 (2014).
19. H. M. Kaneb *et al.*, *Hum. Mol. Genet.* **24**, 1363–1373 (2015).
20. A. Chaussenot *et al.*, *Neurobiol. Aging* **35**, 2884.e1–2884.e4 (2014).
21. S. Bannwarth *et al.*, *Brain* **137**, 2329–2345 (2014).
22. J. Couthouis, A. R. Raphael, R. Daneshjou, A. D. Gitler, *PLOS Genet.* **10**, e1004704 (2014).
23. H. Maruyama *et al.*, *Nature* **465**, 223–226 (2010).
24. S. L. Rea, V. Majcher, M. S. Searle, R. Layfield, *Exp. Cell Res.* **325**, 27–37 (2014).
25. H. Maruyama, H. Kawakami, *Geriatr. Gerontol. Int.* **13**, 528–532 (2013).
26. M. Thomas, J. Alegre-Abarrategui, R. Wade-Martins, *Brain* **136**, 1345–1360 (2013).
27. D. Kachaner, P. Génin, E. Laplantine, R. Weil, *Cell Cycle* **11**, 2808–2818 (2012).
28. B. A. Keller *et al.*, *Acta Neuropathol.* **124**, 733–747 (2012).
29. E. L. Scotter *et al.*, *J. Cell Sci.* **127**, 1263–1278 (2014).
30. S. Morton, L. Hesson, M. Peggie, P. Cohen, *FEBS Lett.* **582**, 997–1002 (2008).
31. M. Pilli *et al.*, *Immunity* **37**, 223–234 (2012).
32. C. E. Gleason, A. Ordureau, R. Gourlay, J. S. Arthur, P. Cohen, *J. Biol. Chem.* **286**, 35663–35674 (2011).
33. P. Wild *et al.*, *Science* **333**, 228–233 (2011).
34. J. Korac *et al.*, *J. Cell Sci.* **126**, 580–592 (2013).
35. Y. C. Wong, E. L. Holzbaur, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4439–E4448 (2014).
36. J. R. Buchan, R. M. Kolaitis, J. P. Taylor, R. Parker, *Cell* **153**, 1461–1474 (2013).
37. M. Ramaswami, J. P. Taylor, R. Parker, *Cell* **154**, 727–736 (2013).
38. M. Komatsu, S. Kageyama, Y. Ichimura, *Pharmacol. Res.* **66**, 457–462 (2012).
39. K. Clark *et al.*, *Biochem. J.* **434**, 93–104 (2011).
40. S. Sharma *et al.*, *Science* **300**, 1148–1151 (2003).
41. K. A. Fitzgerald *et al.*, *Nat. Immunol.* **4**, 491–496 (2003).
42. H. Hemmi *et al.*, *J. Exp. Med.* **199**, 1641–1650 (2004).
43. S. Liu *et al.*, *Science* 10.1126/science.aaa2630 (2015).
44. M. G. Wathelet *et al.*, *Mol. Cell* **1**, 507–518 (1998).
45. C. K. Glass, K. Saijo, B. Winner, M. C. Marchetto, F. H. Gage, *Cell* **140**, 918–934 (2010).
46. Y. Yang *et al.*, *Nat. Genet.* **29**, 160–165 (2001).
47. J. Li *et al.*, *Int. J. Cancer* **134**, 1972–1980 (2014).
48. M. Kato *et al.*, *Cell* **149**, 753–767 (2012).
49. E. B. Lee, V. M. Lee, J. Q. Trojanowski, *Nat. Rev. Neurosci.* **13**, 38–50 (2012).
50. Exome Aggregation Consortium (ExAC); http://exac.broadinstitute.org.
51. P. Van Damme, W. Robberecht, *Curr. Opin. Neurol.* **26**, 466–472 (2013).
52. C. Münch, A. Rolfs, T. Meyer, *Amyotroph. Lateral Scler.* **9**, 251–253 (2008).
53. T. Meyer *et al.*, *Neurology* **65**, 141–143 (2005).
54. A. Chesi *et al.*, *Nat. Neurosci.* **16**, 851–855 (2013).

MOLECULAR MOTORS

The structure of the dynactin complex and its interaction with dynein

Linas Urnavicius,[1]* Kai Zhang,[1]* Aristides G. Diamant,[1]* Carina Motz,[1] Max A. Schlager,[1] Minmin Yu,[1] Nisha A. Patel,[2] Carol V. Robinson,[2] Andrew P. Carter[1]†

Dynactin is an essential cofactor for the microtubule motor cytoplasmic dynein-1. We report the structure of the 23-subunit dynactin complex by cryo–electron microscopy to 4.0 angstroms. Our reconstruction reveals how dynactin is built around a filament containing eight copies of the actin-related protein Arp1 and one of β-actin. The filament is capped at each end by distinct protein complexes, and its length is defined by elongated peptides that emerge from the α-helical shoulder domain. A further 8.2 angstrom structure of the complex between dynein, dynactin, and the motility-inducing cargo adaptor Bicaudal-D2 shows how the translational symmetry of the dynein tail matches that of the dynactin filament. The Bicaudal-D2 coiled coil runs between dynein and dynactin to stabilize the mutually dependent interactions between all three components.

Dynactin works with the cytoplasmic dynein-1 motor (dynein) to transport cargos along the microtubule cytoskeleton (1–3). Together, these protein complexes maintain the cell's spatial organization, return components from the cell's periphery, and assist with cellular division (4). Mutations in either complex cause neurodegeneration (5), and both can be co-opted by viruses that travel to the nucleus (6). Dynein and dynactin are similar in size and

Science

AAAS

Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways
Elizabeth T. Cirulli *et al.*
Science **347**, 1436 (2015);
DOI: 10.1126/science.aaa3650

This copy is for your personal, non-commercial use only.