High-throughput targeted sequencing to assess the role of RNA-binding proteins in ALS

Julien Couthouis, Alya R. Raphael, Roxana Daneshjou and Aaron D. Gitler

Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

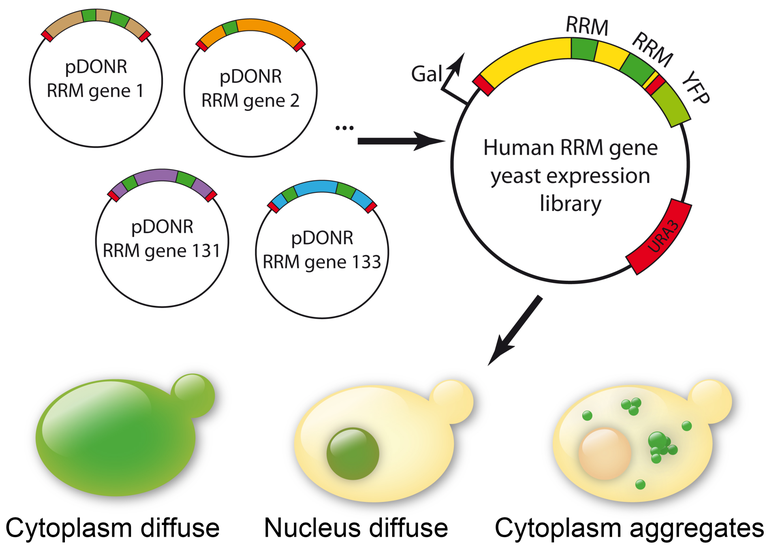
Background

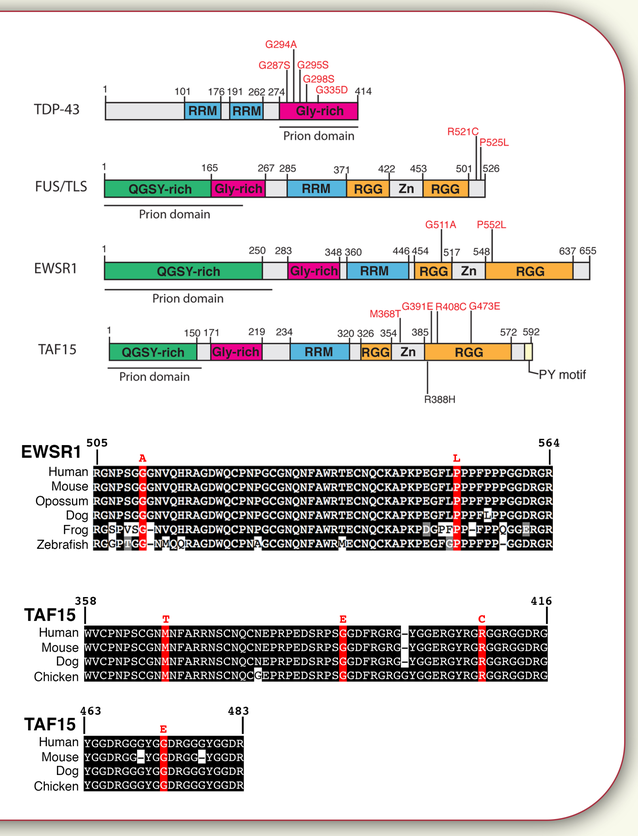
ALS is a devastating adult-onset neurodegenerative disease that attacks upper and lower motor neurons [1] leading to a fatal muscle paralysis, causing death within 2 to 5 years of disease onset.

Mutations in several genes have been linked to both sporadic or familial forms of ALS, including SOD1, TARDBP, FUS/TLS, C9orf72 and others [2]. Two of these genes, TARDBP (TDP-43) and FUS/TLS (FUS) both encode similar types of RNA-binding proteins [3] and have been identified as components of pathological inclusions in neurons of ALS patients [4-6]. An emerging concept suggested by the discoveries of FUS and TDP-43 in ALS is that defects in RNA metabolism might contribute to the pathogenesis.

TDP-43 and FUS both contain RNA recognition motifs (RRM) [3] and they form cytoplasmic aggregates and are toxic when expressed in yeast [7, 8]. Including FUS and TDP-43, there are at least 226 RRM-containing proteins (PFAM ID PF00076) present in the human proteome. We previously focused on TAF15 and EWSR1, and identified new variants in ALS cases [9-10]. Functional studies showed cellular mislocalization and toxicity of these new variants, promoting TAF15 and EWSR1 to the list of ALS risk factor genes.

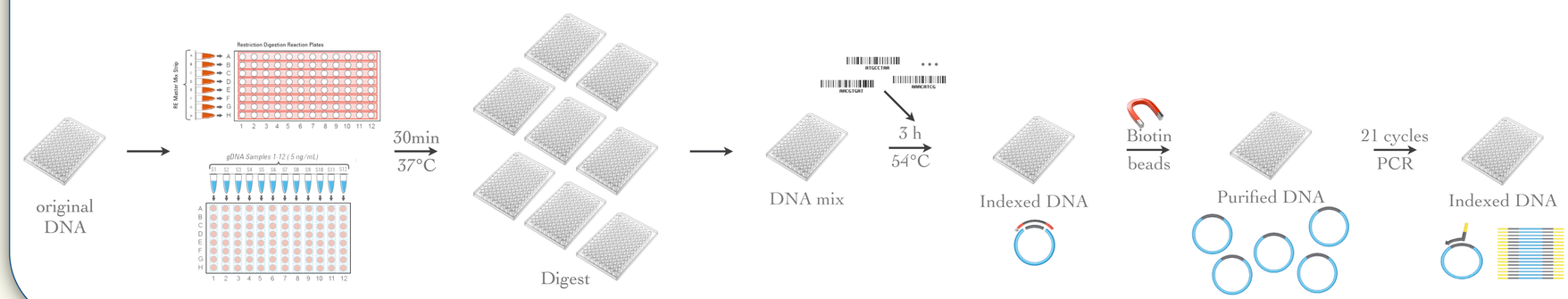
Using new targeted high throughput sequencing techniques we decided to investigate other RRM genes, but also on genes previously linked to ALS and candidates from a previous whole exome sequencing project using ALS trios [11].

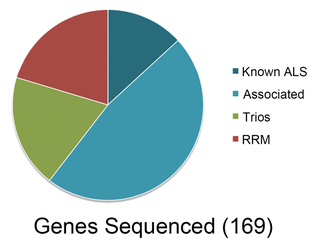


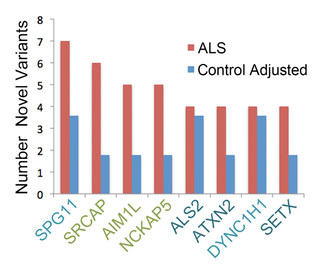


Methods

A custom capture library of 169 genes was designed to include all genes previously linked to ALS, candidates from a previous ALS trios study and RRM genes [9-11].
The Haloplex technology was preferred, as it only requires a small quantity of DNA and uses a fast enzymatic shearing of the genomic DNA.
242 and 129 DNA samples, respectively from ALS patients and age-matched controls, were processed. Highly multiplexed samples were sequenced using a MiSeq, and standard bioinformatics tools (bowtie2, samtools, Picard, GATK) were used to call variants.

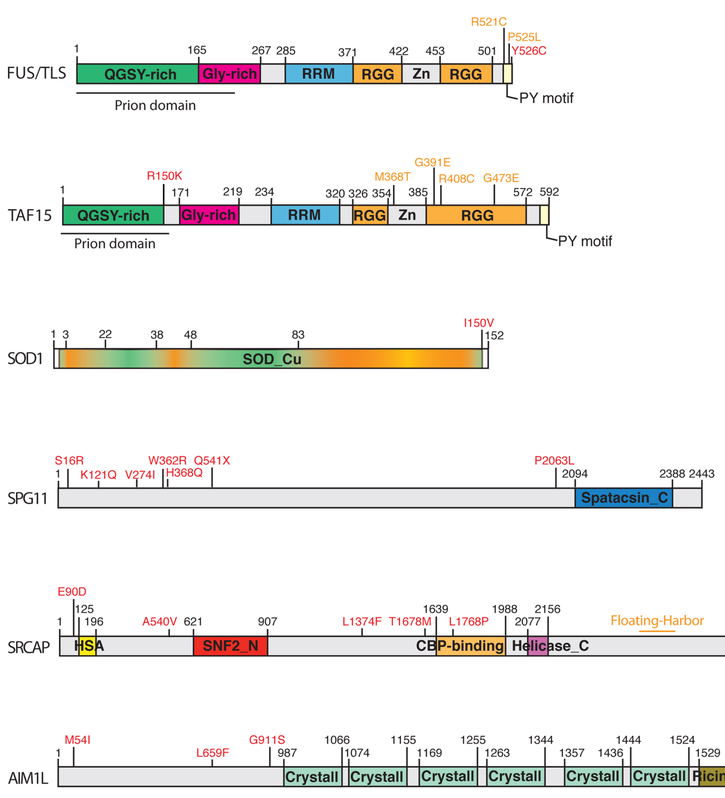


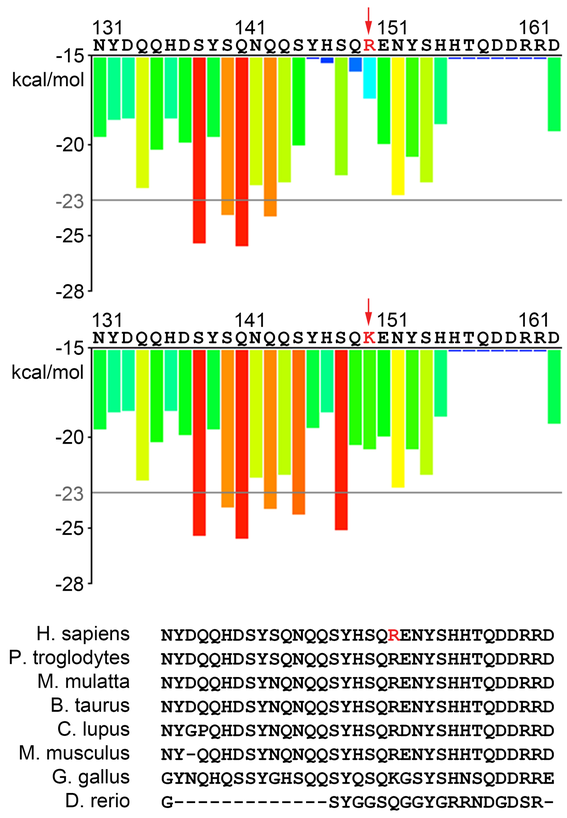


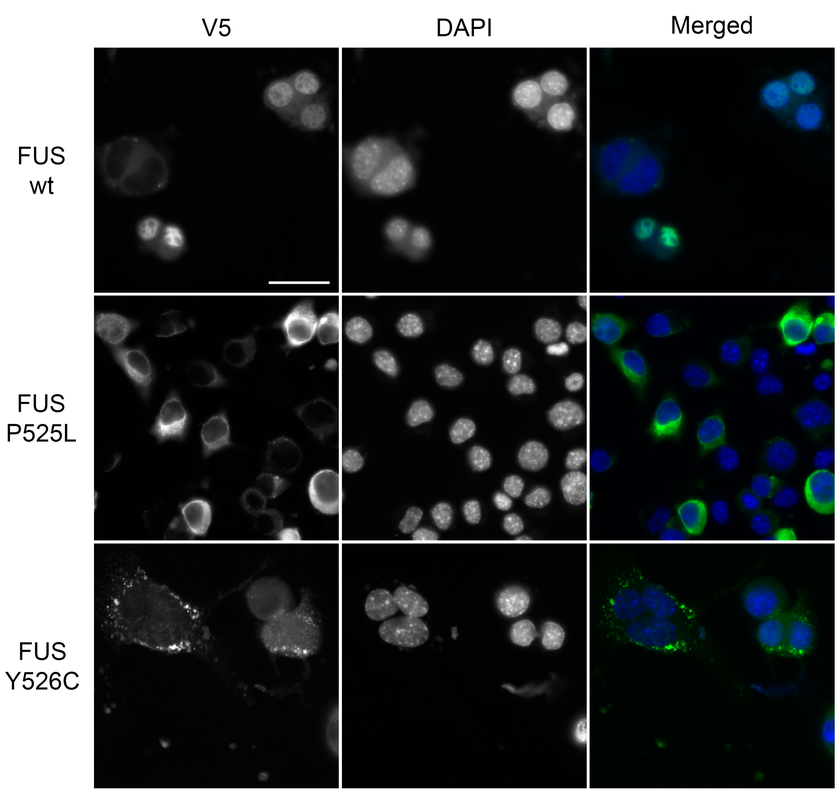


New and rare ALS variants identified

Many new variants were found, both in ALS cases and controls. Thus a significant enrichment of mutation in ALS cases was found in a few genes: SPG11, SRCAP, AIM1L, NCKAP5, ALS2, ATXN2, DYNC1H1 and SETX.
When compared to previously reported mutations linked to ALS, our new variants do not always localize in the same region, nor the same functional domain.
TAF15's new mutation is found on the last residue of its predicted prion domain. This R150K mutation is predicted by zipperDB to greatly increase its fibrilization propensity. This amino acid is also extremly conserved among other mammals, which may indicate a deleterious mutation.
The Y526C mutation in FUS is localized in its NLS domain, a region already know to be critically involved in ALS, although this exact residue has never been previously linked to ALS.







Conclusion

We used a candidate gene approach to discover new mutations in biologically relevant ALS genes using sporadic ALS cases. Additionally, we used a large array of tools to study newly identified variants and assess relevance to disease. We used bioinformatics tools to assess variant pathogenicity and, when possible, functional studies to compare newly identified variants with other ALS variants [12]. This supports the idea that novel or rare mutations in a set of key genes may contribute to sporadic ALS.

The abundance and relative affordability of next generation sequencing techniques has launched a new era in the study of human disease. Suddenly, the capacity to discover the causative mutation behind disease seems almost at one's fingertips. This approach is more straightforward in homogenous diseases, or in cases where large families are available.

References

1. D. W. Cleveland, J. D. Rothstein, Nat Rev Neurosci 2, 806 (Nov, 2001).
2. P. Van Damme, W. Robberecht, Curr Opin Neurol 22, 486 (Oct, 2009).
3. C. Lagier-Tourenne, D. W. Cleveland, Cell 136, 1001 (Mar 20, 2009).
4. M. Neumann et al., Science 314, 130 (Oct 6, 2006).
5. T. J. Kwiatkowski, Jr. et al., Science 323, 1205 (Feb 27, 2009).
6. C. Vance et al., Science 323, 1208 (Feb 27, 2009).
7. B. S. Johnson et al., Proc Natl Acad Sci USA 105, 6439 (Apr 29, 2008).
8. B. S. Johnson et al., J Biol Chem 284, 20329 (Jul 24, 2009).
9. J. Couthouis et al., Proc Natl Acad Sci USA. 108(52):20881-90 (Dec 27, 2011)
10. J. Couthouis et al., Hum Mol Genet. 21(13):2899-911 (Jul 1, 2012)
11. A. Chesi et al., Nat Neurosci. 16(7):851-5 (Jul ,2013).
12. J. Couthouis et al., PLoS Genet. 10(10):e1004704 (Oct 9, 2014)

Acknowledgments

This work was supported by NIH Director's New Innovator Award 1DP2OD004417, 1R01NS065317 and a grant from the Packard Center for ALS Research at Johns Hopkins (A.D.G.). A.D.G. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts. A.R. is a Bright Focus Postdoctoral Fellow. RD is a Stanford: Medical Scientist Training Program Fellow

http://goo.gl/WOzyLU