Exome sequencing in ALS identifies new risk genes and pathways

Brittany N. Lasseigne, Ph.D.
January 2015
Postdoctoral Fellow, Rick Myers Lab
Amyotrophic lateral sclerosis
ā’-mi-ə-trō’fik lāt’ər-əl sklē-rō’sēs

What is ALS?

Ice Bucket Challenge Boosts ALS Awareness Globally
Monthly page views of Wikipedia articles explaining ALS, by language

- **English**: 2,875,160
  - January ’13 - July ’14 (Average): 163,300
  - Aug ’14: 579,404
- **German**: 526,122
  - Jan - Jul ’14: 45,792
  - Aug ’14: 526,122
- **Spanish**: 38,969
  - Jan - Jul ’14: 189,847
  - Aug ’14: 14,929
- **Russian**: 178,471
  - Jan - Jul ’14: 3,016
  - Aug ’14: 172,066
- **French**: 32,084
  - Jan - Jul ’14: 172,066
  - Aug ’14: 32,084

- Only 50% of Americans had even heard of ALS pre-Ice Bucket Challenge
- Within weeks, there were over 2.4 million YouTube videos with over 1 billion views
- Over $115 million raised (>40x more than previous year)

Image: Forbes. alsa.org
Outline

• Background
• Approach and Methods
• Known ALS Genes
• Novel ALS Genes
• Conclusions
Amyotrophic Lateral Sclerosis (ALS)

- Also known as Lou Gehrig’s disease
- Progressive neurodegenerative disease causing muscle weakness and atrophy due to degeneration of motor neurons
- ~5,600 new cases in the US annually
- Median survival time from onset to death is 39 months

Photo: phillysportshistory.com
ALS Symptoms and Diagnosis

Initial symptoms and progression rate can vary

No single test or procedure for definitive ALS diagnosis:

- Clinical examination and tests to rule out other diseases that mimic ALS:
  - electrodiagnostics
  - blood/urine tests
  - spinal tap
  - x-rays/MRIs
  - myelogram of cervical spine
  - muscle/nerve biopsy
  - neurological examination, etc.
ALS Management

- Riluzole (blocks sodium channels associated with damaged neurons) improves survival by months

- Physical therapy can help delay loss of strength and maintain endurance

- Breathing support when the muscles that assist breathing weaken

Genetic Discoveries in ALS

Percentage ALS explained by genetic mutation since 1992

- Familial ALS: 10% of cases
- Sporadic ALS: 90% of cases

- Almost all known ALS-causing mutations are dominant
- Exact molecular mechanism causing disease unknown
- Superoxide dismutase 1

Adapted from Renton, et al. 2014.
89% of sporadic ALS cases are not explained by known genetic alterations

Adapted from Renton, et al. 2014.
### Heterogeneous symptoms, progression, and genetic mutations

20+ Distinct ALS Subtypes

<table>
<thead>
<tr>
<th>Genetic subtype</th>
<th>Chromosomal locus</th>
<th>Gene</th>
<th>Protein</th>
<th>Onset</th>
<th>Inheritance</th>
<th>Clinical feature</th>
<th>Other diseases caused by the gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS1</td>
<td>21q22.1</td>
<td>SOD1</td>
<td>Cu/Zn SOD-1</td>
<td>Adult</td>
<td>AD/AR</td>
<td>Typical ALS</td>
<td>NA</td>
</tr>
<tr>
<td>ALS2</td>
<td>2q33-2q35</td>
<td>Alsin</td>
<td>Alsin</td>
<td>Juv</td>
<td>AR</td>
<td>Slowly progressive, predominantly UMN signs like limb, &amp; facial spasticity</td>
<td>PLS IAHSP</td>
</tr>
<tr>
<td>ALS3</td>
<td>18q21</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Adu</td>
<td>AD</td>
<td>Typical ALS with limb onset especially lower limb</td>
<td>NA</td>
</tr>
<tr>
<td>ALS4</td>
<td>9q34</td>
<td>SETX</td>
<td>Senataxin</td>
<td>Juv</td>
<td>AD</td>
<td>Slowly progressive, distal hereditary motor neuropathy with pyramidal signs</td>
<td>SCAR 1 and AOA2</td>
</tr>
<tr>
<td>ALS5</td>
<td>15q15-21</td>
<td>SPG 11</td>
<td>Spatacsin</td>
<td>Juv</td>
<td>AR</td>
<td>Slowly progressive</td>
<td>HSP</td>
</tr>
<tr>
<td>ALS6</td>
<td>16p11.2</td>
<td>FUS</td>
<td>Fused in Sarcoma</td>
<td>Juv/Adu</td>
<td>AD/AR</td>
<td>Typical ALS</td>
<td>NA</td>
</tr>
<tr>
<td>ALS8</td>
<td>20q13.3</td>
<td>VAPB</td>
<td>VAPB</td>
<td>Adu</td>
<td>AD</td>
<td>Typical and atypical ALS</td>
<td>SMA</td>
</tr>
<tr>
<td>ALS9</td>
<td>14q11.2</td>
<td>ANG</td>
<td>Angiogenin</td>
<td>Adu</td>
<td>AD</td>
<td>Typical ALS, FTD and Parkinsonism</td>
<td>NA</td>
</tr>
<tr>
<td>ALS10</td>
<td>1p36.2</td>
<td>TARDBP</td>
<td>DNA-binding protein</td>
<td>Adu</td>
<td>AD</td>
<td>Typical ALS</td>
<td>NA</td>
</tr>
<tr>
<td>ALS11</td>
<td>6q21</td>
<td>FIG 4</td>
<td>Phosphoinositide-sphosphatease</td>
<td>Adu</td>
<td>AD</td>
<td>Rapid progressive with prominent corticospinal tract signs</td>
<td>CMT 4 J</td>
</tr>
<tr>
<td>ALS12</td>
<td>10p13</td>
<td>OPTN</td>
<td>Optineurin</td>
<td>Adu</td>
<td>AD/AR</td>
<td>Slowly progressive with limb onset and predominant UMN signs</td>
<td>Primary Open Angle Glaucoma</td>
</tr>
<tr>
<td>ALS14</td>
<td>9p13.3</td>
<td>VCP</td>
<td>VCP</td>
<td>Adu</td>
<td>AD</td>
<td>Adult onset, with or without FTD</td>
<td>IBMPFD</td>
</tr>
<tr>
<td>ALS15/ ALSX</td>
<td>Xp11</td>
<td>UBQLN2</td>
<td>Ubiquilin 2</td>
<td>Adu/Juv</td>
<td>XD</td>
<td>UMN signs proceeding LMN signs</td>
<td>NA</td>
</tr>
<tr>
<td>ALS16</td>
<td>9q13.2-21.3</td>
<td>SIGMAR1</td>
<td>SIGMAR1</td>
<td>Juv</td>
<td>AR</td>
<td>Juvenile onset typical ALS</td>
<td>FTD</td>
</tr>
<tr>
<td>ALS-FTD1</td>
<td>9q21-22</td>
<td>unknown</td>
<td>unknown</td>
<td>Adu</td>
<td>AD</td>
<td>ALS with FTD</td>
<td>FTD</td>
</tr>
<tr>
<td>ALS-FTD2</td>
<td>9p21</td>
<td>C9ORF72</td>
<td>C9ORF72</td>
<td>Adu</td>
<td>AD</td>
<td>ALS with FTD</td>
<td>FTD</td>
</tr>
<tr>
<td>NA</td>
<td>2p13</td>
<td>DCTN1</td>
<td>Dynactin</td>
<td>Adu</td>
<td>AD</td>
<td>Distal hereditary motor neuropathy with vocal paresis</td>
<td>NA</td>
</tr>
<tr>
<td>Other rare-occurring ALS genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALS3</td>
<td>18q21</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Adu</td>
<td>AD</td>
<td>Typical ALS with limb onset especially lower limb</td>
<td>NA</td>
</tr>
<tr>
<td>ALS7</td>
<td>20pter-p13</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Adu</td>
<td>AD/AR</td>
<td>Typical ALS</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>12q22.23</td>
<td>DAO</td>
<td>DAO</td>
<td>Adu</td>
<td>AD</td>
<td>Typical ALS</td>
<td>NA</td>
</tr>
</tbody>
</table>

Figure: Chen, et al. 2013.
Various molecular pathways are implicated in ALS

Figure: Chen, et al. 2013.
Neurotoxic Protein Aggregates in >95% of ALS Patients

- Alzheimer’s plaques
- Parkinson’s Lewy bodies
- Huntington’s intranuclear inclusions
- Prion amyloid plaques
- Amyotrophic lateral sclerosis aggregates

Image: QR Pharma.
ALS Genome Sequencing Consortium

Biogen Idec (Tim Harris)
HudsonAlpha (Rick Myers)
Duke University (David Goldstein)

Project Goals
Identify rare coding variants and new genes/pathways associated with sporadic ALS
Identifying Variants with Exome Sequencing

**Exome Sequencing**
- Identify variation in coding regions (genes)
- Advantage: Interpretability and lower cost compared to whole genome sequencing

**Compare Variants**
CTACGATTCGA Control Group (n=~6500)
CTAGGATTCGA Affected Patient Group (n=~3000)
HudsonAlpha Project Workflow

- Sample preparation, library construction, and sequencing
  - Base calling, quality filtering, and mapping
    - Variant calling
      - Variant annotation
        - Gene burden testing
          - Analyze results

---

**Sample genotyping**

**Greg Cooper and Rick Myers**

**Shawn Levy, Braden Boone, Angela Jones, and the GSL**

**Gary Beard, Scott Newberry and the Computational Services Team**

**Lindsay Jones, Kevin Roberts, and the Absher Lab**

**Jack Wimbish**
HudsonAlpha Project Workflow

Greg Cooper and Rick Myers

Sample preparation, library construction, and sequencing

Base calling, quality filtering, and mapping

Variant calling

Variant annotation

Gene burden testing

Analyze results

Sample genotyping

Shawn Levy, Braden Boone, Angela Jones, and the GSL

Gary Beard, Scott Newberry and the Computational Services Team

Lindsay Jones, Kevin Roberts, and the Absher Lab

Jack Wimbish
HudsonAlpha Project Workflow

Shawn Levy, Braden Boone, Angela Jones, and the GSL

Gary Beard, Scott Newberry and the Computational Services Team

Lindsay Jones, Kevin Roberts, and the Absher Lab

Jack Wimbish

Sample preparation, library construction, and sequencing

Base calling, quality filtering, and mapping

Variant calling

Variant annotation

Gene burden testing

Analyze results

ATCGCGATTTAACCACGCGCTA
ATCGCGAATTAAACCACGCGCTA
ATCGCGAATTAAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
ATCGCGATTTAACCACGCGCTA
HudsonAlpha Project Workflow

Sample preparation, library construction, and sequencing

Base calling, quality filtering, and mapping

Variant calling

Variant annotation

Gene burden testing

Analyze results

Sample genotyping

Shawn Levy, Braden Boone, Angela Jones, and the GSL

Gary Beard, Scott Newberry and the Computational Services Team

Lindsay Jones, Kevin Roberts, and the Absher Lab

Jack Wimbish

Greg Cooper and Rick Myers

Enrichment of pathogenic variants by gene

2/6,405=0.031% Controls

19/2,870=0.662% Cases
HudsonAlpha Project Workflow

Sample genotyping

- Shawn Levy, Braden Boone, Angela Jones, and the GSL
- Gary Beard, Scott Newberry and the Computational Services Team
- Lindsay Jones, Kevin Roberts, and the Absher Lab
- Jack Wimbish

Gene burden testing

- Base calling, quality filtering, and mapping
- Variant calling
- Variant annotation

Analyze results

- Sample preparation, library construction, and sequencing

- Biological interpretation
- Validation cohort
- Follow-up experiments
Identifying and Filtering Variants

Gene Model
CCDS +10 bp into each intron

Search Area: Exon + 10bp
Identifying and Filtering Variants

Gene Model
CCDS +10 bp into each intron

MAF Mode: Rare Variants
- Dominant (MAF: ExAC Global <0.005%, 1000G <1%, Internal <0.05%)
- Recessive (homozygous and compound heterozygous, ExAC Global/1000G/Internal < 1% MAF)

<table>
<thead>
<tr>
<th>Position</th>
<th>Ref</th>
<th>Alt</th>
<th>Allele #/Count</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>21:33038865</td>
<td>A</td>
<td>C</td>
<td>3984/103,280</td>
<td>0.03857 (3.857%)</td>
</tr>
<tr>
<td>21:33040890</td>
<td>A</td>
<td>C</td>
<td>1/121,216</td>
<td>0.000008250 (0.0008250%)</td>
</tr>
</tbody>
</table>
Identifying and Filtering Variants

Gene Model
CCDS +10 bp into each intron

MAF Mode: Rare Variants
• Dominant (MAF: ExAC Global <0.005%, 1000G <1%, Internal <0.05%)
• Recessive (homozygous and compound heterozygous, ExAC Global/1000G/Internal < 1% MAF)
Identifying and Filtering Variants

**Gene Model**
CCDS +10 bp into each intron

**MAF Mode: Rare Variants**
- Dominant (MAF: ExAC Global <0.005%, 1000G <1%, Internal <0.05%)
- Recessive (homozygous and compound heterozygous, ExAC Global/1000G/Internal < 1% MAF)

**Qualifying Filters**
- MAF: minor allele frequency
- CADD: score deleteriousness of variants (scores >=10)
- PolyPhen: predict possible impact of AA substitution on protein
- Protein altering annotation (splice, nonsense, frameshift, missense mutations)
- Protein loss of function annotation (splice, nonsense, frameshift mutations)
Identifying and Filtering Variants

Gene Model
CCDS +10 bp into each intron

MAF Mode: Rare Variants
- Dominant (MAF: ExAC Global <0.005%, 1000G <1%, Internal <0.05%)
- Recessive (homozygous and compound heterozygous, ExAC Global/1000G/Internal < 1% MAF)

Qualifying Filters
- MAF: minor allele frequency
- CADD: score deleteriousness of variants (scores >=10)
- PolyPhen: predict possible impact of AA substitution on protein
- Protein altering annotation (splice, nonsense, frameshift, missense mutations)
- Protein loss of function annotation (splice, nonsense, frameshift mutations)
Qualifying Filters

- MAF: minor allele frequency
- CADD: score deleteriousness of variants (scores >=10)
- PolyPhen: predict possible impact of AA substitution on protein
- Protein altering annotation (splice, nonsense, frameshift, missense mutations)
- Protein loss of function annotation (splice, nonsense, frameshift mutations)

**Gene Model**
CCDS +10 bp into each intron

**MAF Mode: Rare Variants**
- Dominant (MAF: ExAC Global <0.005%, 1000G <1%, Internal <0.05%)
- Recessive (homozygous and compound heterozygous, ExAC Global/1000G/Internal < 1% MAF)

**Identifying and Filtering Variants**

1 model x 2 modes x 5 filters = 10 conditions
Gene Burden Testing of Rare Variants

**Count Qualifying Variants:**
- Count qualifying variants in a gene-based collapsing analysis including exons meeting coverage benchmarks in both cases and controls
  - Example: Loss of Function (splice, nonsense, or frameshift)

**Gene X: Variant Enrichment**

- 2/6,405 = 0.031% Controls
- 19/2,870 = 0.662% Cases

**Compare Frequency Distributions**
- Significant enrichment of qualifying variants between groups
- 97% of genes had at least one sample with a qualifying variant
Project Information

- 2,870 ALS cases and 6,405 controls passed QC (HAIB + Duke) and were included in analysis (discovery)

- Follow-up custom capture sequencing of 1,318 additional cases and 2,371 additional controls

Analyzing rare variant case burden for:
- Predisposition
- Site of Onset
- Age of Onset
- Gender
- Survival

- Analysis utilizes combined p-values (Cochran-Mantel-Haenszel) corrected for number of genes tested over all models and assessed for homogeneity of effects across different groups (Breslow-Day)
Significant Genes

Q-Q (quantile-quantile) plot:

- Examine distribution of p-values compared to expected
- If the distributions are similar, plot approximately lies on line $y=x$
Known ALS Genes: SOD1

Significant after multiple testing correction across all models:

- **SOD1**: 0.870% of cases and 0.078% of controls (dominant coding model)

- **SOD1**: First gene associated with fALS, enzyme that destroys free superoxide radicals

- Component of protein aggregates in ALS motor neurons
Known ALS Genes: TARDBP

Many other known ALS genes show strong association:

• **TARDBP**: 0.661% of cases and 0.094% of controls (dominant coding variants)
  – Major protein found in protein aggregates

Cirulli & Lasseigne, et al. 2015. (under review)
Known ALS Genes: TARDBP

Many other known ALS genes show strong association:
• *TARDBP*: 0.661% of cases and 0.094% of controls (dominant coding variants)
  – Major protein found in protein aggregates

Cirulli & Lasseigne, et al. 2015. (under review)
Known ALS Genes: TARDDBP

Many other known ALS genes show strong association:
- **TARDDBP**: 0.661% of cases and 0.094% of controls (dominant coding variants)
  - Major protein found in protein aggregates

Qualifying Variants

![Diagram showing various types of variants](image)

- **LOF variant**
- **Missense variant**
- **Splice variant**
- **Case variant**
- **Control variant**
- **Case/control variant**
Known ALS Genes: TARDBP

Many other known ALS genes show strong association:

- **TARDBP**: 0.661% of cases and 0.094% of controls (dominant coding variants)
  - Major protein found in protein aggregates

3’ UTR: involved in ribonucleoprotein binding and splicing

Cirulli & Lasseigne, et al. 2015. (under review)
Known ALS Genes: OPTN

Many other known ALS genes show strong dominant association:
• **OPTN**: 0.620% of cases (non-benign variants), enrichment 0.334% of cases (LoF variants)
  – First associated with FALS; we show association between SALS and **OPTN**
  – Involved in autophagosomal trafficking

Cirulli & Lasseigne, et al. 2015. (under review)
Known ALS Genes: VCP and SPG11

Known ALS associated genes with modest excess of mutations (dominant coding):
• **VCP**: 78% of mutations in same domain
• **SPG11**: Previously reported cause of recessive juvenile ALS (Daoud, et al. 2012.), broader role in SALS?

Other reported genes had slight/no association in our data:
– Some cases pre-screened for some known genes
– Certain genes are not expected to show association due to the nature of the causal variation (ex. *C9orf72*)

## Known ALS Genes:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reported FALS</th>
<th>Reported SALS</th>
<th>Potential ALS in Our Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOD1</td>
<td>12%</td>
<td>1.50%</td>
<td>0.79%</td>
</tr>
<tr>
<td>TARDBP</td>
<td>4%</td>
<td>1%</td>
<td>0.57%</td>
</tr>
<tr>
<td>OPTN</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.39%</td>
</tr>
<tr>
<td>SPG11</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.31%</td>
</tr>
<tr>
<td>VCP</td>
<td>1%</td>
<td>1%</td>
<td>0.22%</td>
</tr>
<tr>
<td>NEFH</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.19%</td>
</tr>
<tr>
<td>HNRNPA1</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.13%</td>
</tr>
<tr>
<td>FIG4</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.13%</td>
</tr>
<tr>
<td>MATR3</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.12%</td>
</tr>
<tr>
<td>ATXN2**</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.11%</td>
</tr>
<tr>
<td>GRN</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.11%</td>
</tr>
<tr>
<td>PFN1</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.08%</td>
</tr>
<tr>
<td>CHCHD10</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.07%</td>
</tr>
<tr>
<td>ANG</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.05%</td>
</tr>
<tr>
<td>TAF15</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.05%</td>
</tr>
<tr>
<td>SIGMAR1</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.04%</td>
</tr>
<tr>
<td>SS18L1</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.04%</td>
</tr>
<tr>
<td>SETX</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.04%</td>
</tr>
<tr>
<td>VAPB</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C9orf72**</td>
<td>40%</td>
<td>7%</td>
<td>0.03%</td>
</tr>
<tr>
<td>FUS</td>
<td>4%</td>
<td>1%</td>
<td>0.02%</td>
</tr>
<tr>
<td>CHMP2B</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.02%</td>
</tr>
<tr>
<td>SPAST</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.02%</td>
</tr>
<tr>
<td>ALS2</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.01%</td>
</tr>
<tr>
<td>SQSTM1</td>
<td>1%</td>
<td>&lt;1%</td>
<td>0.00%</td>
</tr>
<tr>
<td>PRPH</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0.00%</td>
</tr>
<tr>
<td>TUBA4A*</td>
<td>1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>ELP3*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>DAO*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>DCTN1*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>EWSR1*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>GLE1*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>UBQLN2*</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Other reported genes had slight/no association in our data:

- Some cases pre-screened for some known genes
- Certain genes are not expected to show association due to the nature of the causal variation (ex. C9orf72)
- Other genes are mutated so rarely that even for this sample size, we did not detect statistically significant causal variants
Identifying Novel ALS Genes: \textit{TBK1}

- \textit{TBK1} interacts with \textit{OPTN} and \textit{SQSTM1}, both ALS-associated genes

- These genes play important roles in autophagy and inflammation:
  - facilitating autophagic turnover of ubiquitylated bacteria
  - regulating autophagosome maturation
  - regulating the NF-κB pathway

- \textit{TBK1}’s role suggests \textit{OPTN} is more important than previously recognized

- Non-benign variants: 1.097\% of cases
- LoF variants: 0.382\% of cases
Autophagy in ALS

- TBK1 co-localizes with OPTN and SQSTM1 in autophagosomes -- it is possible all three associate with protein aggregates in ALS and may be critical components of the aggresome pathway for protein degradation.

- Mutations in OPTN and TBK1 (1.30% of cases in our dataset) may be an important subgroup of patients with a common biological etiology.

- Case mutations in OPTN and TBK1 were largely heterozygous and LoF, suggesting a reduction in trafficking of cargo through augophagosomal pathway or disruption of maturation may promote disease.

1. Mutation in SOD1, TARDBP/TDP-43 or FUS

2. Protein aggregates stain with anti-SQSTM1 and OPTN antibodies

3. SQSTM1 and OPTN proteins function as the cargo receptors, recruiting aggregated proteins to the autophagosome

4. This leads to autophagy: degradation of aggregated proteins

Identifying Novel ALS Genes: **NEK1**

- **NEK1**: multi-functional kinase, role in cilia formation and centrosome function, never previously linked to ALS

- Follow-up custom capture sequencing (1,318 additional cases and 2,371 additional controls) further supports **NEK1**’s role in ALS predisposition

Cirulli & Lasseigne, et al. 2015. (under review)
NEK1 associates with ALS2 and VAPB

- To investigate binding partners, we performed an unbiased screen of NEK1-interacting proteins in human kidney epithelial cells via AP-MS

- Interactions validated by immunoprecipitation followed by western blotting of co-expressed proteins in neuronal NSC-34 cells

- Suggests NEK1 may contribute to ALS through multiple mechanisms:
  - ALS2 and VAPB control cytoplasmic trafficking of endosomes and lipids in diverse cell lineages, respectively, both biological functions that are now appreciated as important in other neurodegenerative diseases

Recessive causes of ALS when mutated:
ALS2: RAB guanine nucleotide exchange factor
VAPB/VAPA: transmembrane proteins that transfer lipids from the ER to the plasma membrane

Cirulli & Lasseigne, et al. 2015. (under review)
Analysis of Clinical Features

Analyzing rare variant case burden for:

• Predisposition
• Site of Onset
• Age of Onset
• Gender
• Survival

No genome-wide significant associations with these features
DAO Mutation Carriers

- Focusing on known ALS predisposition genes:
  - DAO mutation carriers are significantly associated with shorter survival times
  - Known FALS mutations reduce DAO activity (required for D-serine clearance) leading to neurotoxicity (Paul, et al. 2012) and D-serine levels are increased in SOD1 mice and ALS patient spinal cords (Sasabe, et al. 2007; Thompson, et al. 2012)

Conclusions

• This study provides one of the few examples of WES successfully identifying variants that predispose humans to a sporadic, complex disease

• We implicate two novel genes (TBK1 and NEK1), suggest OPTN plays a broader role in ALS than previously recognized, and propose possible new directions for drug screening programs

• Both NEK1 and TBK1 are protein kinases with binding partners implicated in ALS (recessive mutations), suggesting a potential regulatory relationship among risk factors and that signaling systems linked with autophagy and vesicle trafficking may be prominently affected pathways in ALS

Finding the 87%

• **WGS/Longer reads:**
Future studies will likely benefit from WGS and the ability to identify regulatory mutations and small structural variants (ex. repeat expansions like *C9orf72* and *ATXN2*)

• **Larger sample sizes:**
Very large sample sizes are required for a comprehensive picture of ALS genetics:
  - in our data set, *SOD1*, *NEK1*, and *TBK1* account for 2.439% of cases

• **More, well-characterized controls:**
Well-characterized, publically available control sample sets are crucial for discovering additional variants associated with complex traits
Thank-you!

University of Massachusetts Medical School
Diane McKenna-Yasek, John Landers, Robert Brown

McGill University
Patrick A. Dion, Claire S. Leblond, Guy A. Rouleau

Stanford University School of Medicine
Julien Couthouis, Alessandra Chesi, Alya Raphael, Aaron Gitler

Columbia University
Neil Shneider, Wendy Chung, Tom Maniatis

Massachusetts General Hospital
Winnie W. Xin, Katherine B. Sims

Perelman School of Medicine at the Univ. of Penn.
John Q. Trojanowski, Vivianna Van Deerlin, Lauren Elman, Leo McCluskey, Murray Grossman

Emory University
Jonathan D. Glass

Duke ALS Clinic and Durham VA Medical Center
Richard S. Bedlack

Washington University School of Medicine
Janet Ca, Matthew B. Harms

University of California, San Diego
John M. Ravits

Cedars Sinai Medical Center
Robert H. Baloh

Houston Methodist Hospital/
Weill Cornell Medical College of Cornell University
Stanley Appel, Ericka Simpson

University of Utah School of Medicine
Stefan M. Pulst, Summer Gibson

Harvard Medical School
Sebastian D. Hayes, Alban Ordureau, Sharan Swarup, J. Wade Harper

---

Rick Myers
Genomic Services Lab (Shawn Levy, Braden Boone, Jack Wimbish, Angela Jones)
Absher Lab (Lindsay Jones)
Computational Services (Scott Newberry, Gary Beard)
Myers Lab and G. Cooper Lab

David Goldstein
Slavé Petrovski, Yi-Fan Lu, Quanli Wang,
Jonathan Keebler, Brian J. Krueger,
Zhong Ren, Andrew S. Allen

Liz Cirulli

Tim Harris
John P. Carulli, Aaron Day-Williams, John Staropoli, Sebastian D. Hayes
C9orf72 nucleotide repeat structures initiate molecular cascades of disease

Aaron R. Haeusler¹,², Christopher J. Donnelly³,⁴, Goran Periz¹,², Eric A. J. Simko¹,², Patrick G. Shaw⁵, Min-Sik Kim⁵, Nicholas J. Maragakis³, Juan C. Troncoso⁶, Akhilesh Pandey⁵, Rita Sattler³,⁴, Jeffrey D. Rothstein²,³,⁴ & Jiou Wang¹,²

- DNA and RNA of the C9orf72 HRE (GGGGCC)_n form G-quadruplexes
- This impedes transcription and results in abortive RNAs
- Structure-dependent HRE-binding proteins cause nucleolar stress (sequestration of proteins and competition for RNA-processing machinery)
- Repeat-Associated Non-ATG dependent translation (RAN-translation) causes aggregative polypeptides

Loop 1

Loop 2

Loop 3
RUNNing a RAN-Translatable Gene

Adapted from Pearson, PLoS Gen, 2011.
CCDS:

- Consensus Coding Sequence
- Collaborative effort to maintain a dataset of protein-coding regions that are identically annotated on the human and mouse reference genome assemblies by the NCBI and Ensembl genome annotation pipelines.