Science Webinar Series

PLEASE STAND BY... the webinar

A Traveler’s Guide to Next Generation Sequencing: Navigating the Sea of Genetic Variants

will begin shortly...

Instructions for Viewers

› Change the size of any window by dragging the lower right corner. Use controls in top right corner to close or maximize each window.

› What each widget does:

   - shows the audio media player
   - opens the Ask a Question box
   - download slides and more info
   - Facebook login
   - Twitter login (#ScienceWebinar)

   - shows slide window
   - shows speaker bios
   - search Wikipedia
   - LinkedIn login

   - if you need help
Science Webinar Series

A Traveler’s Guide to Next Generation Sequencing:
Navigating the Sea of Genetic Variants

26 June 2013

Brought to you by the Science/AAAS Custom Publishing Office

Participating Experts:

Scott D. Kahn, Ph.D.
Illumina
San Diego, CA

Ruthild Weber M.D.
Hannover Medical School
Hannover, Germany

Sponsored by: BIOPBASE BIOLOGICAL DATABASES
Status of Next Generation Sequencing Information for Use in a Clinical Setting

Scott D. Kahn, PhD
Chief Information Officer and VP
Capacity, Annotation, and Growing Clinical Need

Sequencing Progress vs Compute and Storage
Moore's and Kryder's Laws fall far behind

% Annotated SNPS
% of clinical genomes

2008 2009 2010 2011 2012 2013 2014

0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.001 0.0015 0.002 0.0025 0.003 0.0035 0.004 0.005 0.01 0.015 0.02 0.025 0.03 0.035 0.04 0.045 0.05 0.055 0.06 0.065 0.07 0.075 0.08 0.085 0.09 0.095 0.1 0.105 0.11 0.115 0.12 0.125 0.13 0.135 0.14 0.145 0.15 0.155 0.16 0.165 0.17 0.175 0.18 0.185 0.19 0.195 0.2 0.205 0.21 0.215 0.22 0.225 0.23 0.235 0.24 0.245 0.25
Sequencing Technology
Illumina Sequencing

DNA (<1 ug)

Sample preparation

Cluster growth (0.1 – 0.5 billion)

Sequencing (2 x 35-100 bases)

Image acquisition

Base calling & Q scoring
Next Generation Sequencing: How it Compares

A. Capillary sequencing

Gene → Amplicon → Sequence → Variant

B. Illumina IGS sequencing

CTCATCAGCAGCAAACGT
CTCATCAGGAGCAAACGT
CTCATCAGGAGCAAACGTC
CTCATCAGGAGCAAACGTCTGC
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
CTCATCAGGAGCAAACGTCTGCTTT
97 64 55 76 62 59 62 91 90 90 56 26 65 61 69 58 77 49 51

AG

AG

95

80
Analytical Sensitivity and Specificity
## Summary of Increased Accuracy (Specificity)

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Mendelian Conflicts</th>
<th>Accuracy</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.62</td>
<td>13,032</td>
<td>99.9995%</td>
<td>unfiltered</td>
</tr>
<tr>
<td>96.10</td>
<td>8,383</td>
<td>99.9997%</td>
<td>+ gVCF filters</td>
</tr>
<tr>
<td>95.25</td>
<td>5,309</td>
<td>99.9998%</td>
<td>+ score:coverage</td>
</tr>
</tbody>
</table>

↑ 1.43% loss in sensitivity  
↑ 59.26% loss in conflicts

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Conflicts</th>
<th>Accuracy</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.90</td>
<td>4,928</td>
<td>99.9998%</td>
<td>BWA+MPG*</td>
</tr>
</tbody>
</table>

**NB: Accuracy is expressed here as % total filtered calls that are Mendelian concordant**

* Accurate and comprehensive sequencing of personal genomes  
*Genome Res.* 2011 21: 1498-1505  

**Eland+CASAVA**
Platinum Genome Project

► Where are we doing well?
► What parts of the genome are still inaccessible or less accurately called – and most importantly, why?
► Maximum utility for use in research and medical applications
► Determine key areas for improvement and assess progress
► Assess performance in real-life situations
## Initial Dataset

<table>
<thead>
<tr>
<th>Sample</th>
<th>Depth</th>
<th>=&gt;Q30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA12877</td>
<td>219.63</td>
<td>91.3</td>
</tr>
<tr>
<td>NA12878</td>
<td>211.88</td>
<td>93.6</td>
</tr>
<tr>
<td>NA12882</td>
<td>217.95</td>
<td>93.2</td>
</tr>
<tr>
<td>NA12881</td>
<td>46.67</td>
<td>91.7</td>
</tr>
<tr>
<td>NA12880</td>
<td>48.37</td>
<td>91.4</td>
</tr>
<tr>
<td>NA12879</td>
<td>48.01</td>
<td>92</td>
</tr>
<tr>
<td>NA12883</td>
<td>54.73</td>
<td>94</td>
</tr>
<tr>
<td>NA12884</td>
<td>43.76</td>
<td>93.2</td>
</tr>
<tr>
<td>NA12885</td>
<td>54.56</td>
<td>94</td>
</tr>
<tr>
<td>NA12886</td>
<td>64.98</td>
<td>91</td>
</tr>
<tr>
<td>NA12887</td>
<td>48.33</td>
<td>92.4</td>
</tr>
<tr>
<td>NA12888</td>
<td>47.61</td>
<td>92.2</td>
</tr>
<tr>
<td>NA12889</td>
<td>49.99</td>
<td>91</td>
</tr>
<tr>
<td>NA12890</td>
<td>59.34</td>
<td>88</td>
</tr>
<tr>
<td>NA12891</td>
<td>45.49</td>
<td>93</td>
</tr>
<tr>
<td>NA12892</td>
<td>50.32</td>
<td>93.4</td>
</tr>
<tr>
<td>NA12893</td>
<td>47.69</td>
<td>92.7</td>
</tr>
</tbody>
</table>

SNP Chip Coverage: 99.8%
Genotype Concordance: 99.28%
Q30 accuracy equals 99.9%
Consistency across all the replicates

- How many replicates were able to be called at a given position?
- How many different genotypes were present at that position?
### Consistency Among Technical Replicates

#### Number of different genotypes

<table>
<thead>
<tr>
<th>Number of replicates</th>
<th>Number of different genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.96</td>
</tr>
<tr>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>2</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
</tr>
<tr>
<td>7</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td>0.17</td>
</tr>
<tr>
<td>10</td>
<td>0.20</td>
</tr>
<tr>
<td>11</td>
<td>0.24</td>
</tr>
<tr>
<td>12</td>
<td>0.32</td>
</tr>
<tr>
<td>13</td>
<td>0.61</td>
</tr>
<tr>
<td>14</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

#### “Metal” Genome

<table>
<thead>
<tr>
<th>“Metal”</th>
<th>Genome</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold</td>
<td>95.1%</td>
<td></td>
</tr>
<tr>
<td>Silver</td>
<td>2.95%</td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>0.01%</td>
<td></td>
</tr>
<tr>
<td>Lead</td>
<td>1.96%</td>
<td></td>
</tr>
</tbody>
</table>

*Note: The image includes a graph showing the consistency among technical replicates, with the X-axis representing the number of replicates passing the genotype quality filter and the Y-axis representing the number of different genotypes. The graph also includes a legend showing the percentage of different genotypes for each “Metal” genome.*
### Practical/Clinical/Medical Relevance

200x build comparison in medically-relevant CDS regions

<table>
<thead>
<tr>
<th>Metal</th>
<th>ALL</th>
<th>Same</th>
<th>Different</th>
<th>Percent the Same</th>
<th>Percent in Metal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>1,187</td>
<td>1,182</td>
<td>5</td>
<td>99.58%</td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>1,151</td>
<td>1,151</td>
<td>0</td>
<td>100.00%</td>
<td>96.97%</td>
</tr>
<tr>
<td>Silver</td>
<td>29</td>
<td>26</td>
<td>3</td>
<td>89.66%</td>
<td>2.44%</td>
</tr>
<tr>
<td>Copper</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>100.00%</td>
<td>0.17%</td>
</tr>
<tr>
<td>Lead</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>60.00%</td>
<td>0.42%</td>
</tr>
</tbody>
</table>
Analytics of Sequencing in a Clinical Laboratory
Individual Genome Sequencing in a Clinical Laboratory

► We established a protocol for sequencing whole genomes
  – Using Professional and Regulatory agency guidelines
  – Under advisement from an external Ethics Advisory Board
  – Geneticists, Clinicians, Ethicists and Attorneys participated in design of protocol

► Our protocol and validations were evaluated externally
  – CLIA certified for high complexity testing
  – CAP accredited
Data Flow and Quality Filters

120 bp paired end reads from a 400 bp library
Accuracy Across the Genome

- >94% of reference covered, >94% of exome (same regions reported)
- concordance to genotyping calls (Human 2.5M-duo = 2,500,000 SNP sites)

Concordance

average: 99.92%
Accuracy: at Single Base Pair

► >40 samples of known genotype sequenced at >6 million-fold depth over 50 loci
► 100% accuracy in genotype calls
► Sub-sampled 10,000 times at various folds of coverage:

Sensitivity
At average 30 fold coverage:
>99.99%

* \( P(x, p, N) = \sum \frac{N!}{(X!)(N-X)!} p^x q^{(N-X)} \)

Specificity
At average 30 fold coverage:
>99.99%
Precision

- **Repeatability:**
  - 13 sequencing runs over 2-month period
  - Three libraries prepared from the same sample
  - Mitochondrial genome sequences compared
  - >99.999% concordance of base calls (detected a heteroplasmy)

- **Reproducibility:**
  - Two Clinical Laboratory Scientists
  - 4 HiSeq 2000 & 2 HiSeq 2500 Sequencers
  - 100% concordance of Phi X genomes between different runs
    - carried out by same or different laboratory scientists
What Have We Learned?

► The technology is only as good as the quality thresholds imposed (informatics!)
  – We can calibrate sensitivity and specificity thresholds
    ▪ As we increase thresholds, we report less data, or test becomes more expensive

► Improve the way we call data
  – Bayesian method will save more good data
  – Insertions/deletions

► We need to improve on the deliverable:
  – “Will you fax me the results?”
  – Annotation of variants to genes, codons
  – Programs for navigating the data,
    ▪ Genes of interest
    ▪ Quality thresholds
    ▪ Types of variants
  – Include clinical information
    ▪ Standard variants? Tailored to clinical indication?
Science Webinar Series

A Traveler’s Guide to Next Generation Sequencing: Navigating the Sea of Genetic Variants

26 June 2013

Brought to you by the Science/AAAS Custom Publishing Office

Participating Experts:

Scott D. Kahn, Ph.D.
Illumina
San Diego, CA

Ruthild Weber M.D.
Hannover Medical School
Hannover, Germany
Our first trip on the Sea of Variants and the lessons learned
Our aim: discover a new island, i.e. syndromic intellectual disability gene

- Only one affected family member
- Intellectual disability is frequently caused by a *de novo* mutation
- Cancer predisposition is frequently autosomal dominant and could be caused by a *de novo* event if no other family members are affected

Exome sequencing in mother, father and patient to filter for *de novo* variants

Weber Group
Department of Human Genetics
Our first trip on the Sea of Variants and the lessons learned

• Our finding: the *de novo* strategy did not yield a relevant *de novo* variant in the patient with syndromic intellectual disability.

• Our thinking: perhaps our assumptions were wrong and the genetic cause of the patient’s phenotype is no *de novo* event.

• Our conclusion: we need a different approach to find the causative variant in this patient. Why do not we try a data mining tool?
Our first trip on the Sea of Variants and the lessons learned
Data mining using Genome Trax™

Genome Trax™ 2013.1
Our first trip on the Sea of Variants and the lessons learned
Results of the databased strategy

Total number of variants in patient’s exome from peripheral blood: 26,974 variants
After removing bad quality and known variants from in-house exomes: 1,795 variants
Input Genome Trax™ (setting: HGMD® inherited disease mutations): 1,795 variants
Output Genome Trax™: 24 variants
Variants related to syndromal phenotype or cancer: 3 variants

Mutations and variations
- HGMD® inherited disease mutations
- HGMD® imputed
- Pharmacogenomics Variants BETA
- ClinVar Variants
- GWAS Catalogue
- COSMIC somatic disease mutations
- EVS Exome Variants
- dbNSFP Nonsynonymous functional predictions
- dbSNP regulatory SNPs
- Ensembl regulatory SNPs

Regulatory features
- TRANSFAC® experimentally verified TFBS
- Predicted ChIP-Seq TFBS
- Predicted TFBS in DNAse hyper sensitivity regions
- CpG islands
- Microsatellites
- Virtual Transcription Start Sites (TSSs)
- Post translational modifications
- miRNA

Gene functional assignments
- Disease associations
- Drug targets
- Pathway membership
- HGMD® disease genes

Novel variations
- Mutation effect prediction using snpEff

Uploaded Track

Annotate variations with the following features:

- **Mutations and variations**
  - HGMD® inherited disease mutations
  - HGMD® imputed
  - Pharmacogenomics Variants BETA
  - ClinVar Variants
  - GWAS Catalogue
  - COSMIC somatic disease mutations
  - EVS Exome Variants
  - dbNSFP Nonsynonymous functional predictions
  - dbSNP regulatory SNPs
  - Ensembl regulatory SNPs

- **Regulatory features**
  - TRANSFAC® experimentally verified TFBS
  - Predicted ChIP-Seq TFBS
  - Predicted TFBS in DNAse hyper sensitivity regions
  - CpG islands
  - Microsatellites
  - Virtual Transcription Start Sites (TSSs)
  - Post translational modifications
  - miRNA

- **Gene functional assignments**
  - Disease associations
  - Drug targets
  - Pathway membership
  - HGMD® disease genes

- **Novel variations**
  - Mutation effect prediction using snpEff

- **Uploaded Track**
The databased strategy identified known compound heterozygous stop mutations in the *BLM* gene inherited from the parents.

---


Patient  Mother  Father

---

*BLM*, c.2695C>T; p.Arg899X

Patient  Mother  Father
The databased strategy identified a heterozygous missense mutation in the \textit{CHEK2} gene that was \textit{de novo} according to Sanger sequencing.

\textit{CHEK2}, c.1427C>T; p.Thr476Met
The identified heterozygous missense variant in the CHEK2 gene is a known and functionally relevant mutation according to HGMD® professional

**CHEK2, c.1427C>T; p.Thr476Met**

---

<table>
<thead>
<tr>
<th>HGMD accession</th>
<th>Disease/phenotype</th>
<th>Gene symbol</th>
<th>Codon change</th>
<th>Amino acid change</th>
<th>Codon number</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM119709</td>
<td>Breast cancer</td>
<td>CHEK2</td>
<td>ACG-ATG</td>
<td>Thr-Met</td>
<td>476</td>
</tr>
</tbody>
</table>

The T476M substitution exhibits a shift from polar to non-polar and displays an increase in Kyte-Doolittle hydrophobicity from -0.7 to 1.9. Approximately 1.01% of missense mutations in HGMD are Thr-Met. The mutation occurs 56 amino acids from the end of the protein.

<table>
<thead>
<tr>
<th>Literature citation</th>
<th>Citation type</th>
<th>Notes</th>
</tr>
</thead>
</table>

---

**Genomic sequence (GRCh37.3):** TGGTGGTAGTGGATCCAAGGACGTATAAGATACCCCGTTGCT

**Genomic coordinate (GRCh37.3):** Chr 22: 29090054

**HGVS nomenclature:** NM_007194.3 c. 1427C>T, NP_009125.1 p.T476M

**dbSNP number:** rs142763740

**Variant class:** DM Disease causing mutation

**Comments:** Descri in Suppl. Table 1 (online).

**CpG:** Yes
Our first trip on the Sea of Variants and the lessons learned

- Our findings 1: the patient had a known syndrome (Bloom syndrome explaining the cancer predisposition, skin hyper- and hypopigmentation, and short stature) caused by known mutations in a known gene ($BLM$). The syndrome had presumably not been diagnosed because it is rare (incidence: 2:100,000) and the patient’s phenotype was not typical (the patient also had intellectual disability due to an additional 22q11.2 microduplication).

- Our findings 2: the patient also carried a known $de$ $novo$ mutation in another cancer predisposition gene ($CHEK2$) that was not found by the $de$ $novo$ strategy because the father’s WES result was false positive.
Our first trip on the Sea of Variants and the lessons learned

- Our thinking: perhaps we should include the databased strategy into our initial filtering approach, and go on to the *de novo* strategy only if no relevant variants are found. This would decrease costs and allow analysis in patients with unavailable parental blood samples.

- Our conclusion: unfortunately, we had not discovered a new island, i.e. no new syndromic intellectual disability gene, but rather a successful strategy to identify causative variants.
Our filtering strategy (part 1)

Total variants in exome from peripheral blood

- Quality filtering
  - E.g. coverage

- Filtering based on variant seriousness
  - Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELS) are retained

Filtering based on deleteriousness prediction

- For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained

Candidate strategy

- Variants in known or suspected disease genes or gene groups (own lists) based on own literature search are retained

Databased strategy

- Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®

Filtering based on clinical features

- Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
Our filtering strategy (part 2)

Population filtering
The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration

Filtering based on inheritance model
E.g. based on information provided by e.g. OMIM, POSSUMweb

Comparison with in house exomes

Disease causing variants
No disease causing variants

Trio-based analysis for de novo strategy or no diagnosis
Applying our filtering strategy to the analysis of patients with severe bilateral renal hypodysplasia (RHD)

- Congenital malformation
- Onset of symptoms usually in childhood
- Genetically heterogeneous
- Detection rate using targeted mutation analysis by Sanger: <20%
- Inheritance is mostly dominant, but can be recessive
- Mutations described as pathogenic can be *de novo* or inherited

Renal hypodysplasia:
small kidneys containing
immature undifferentiated
nephrons and cysts
The databased strategy can be modified by selecting different data sources

3. Annotate variations with the following features:

**Mutations and variations**
- ✔ HGMD® inherited disease mutations (115,600)
- ✔ HGMD® imputed
- ✔ Pharmacogenomics Variants BETA
- ✔ ClinVar Variants
- ✔ GWAS Catalogue (8,700)
- ✔ COSMIC somatic disease mutations (166,000)
- ✔ EVS Exome Variants
- ✔ dbNSFP Nonsynonymous functional predictions
- ✔ dbSNP regulatory SNPs
- ✔ Ensembl regulatory SNPs

**Regulatory features**
- ✔ TRANSFAC® experimentally verified TFBS
- ✔ Predicted ChIP-Seq TFBS
- ✔ Predicted TFBS in DNAse hyper sensitivity regions
- ✔ CpG islands
- ✔ Microsatellites
- ✔ Virtual Transcription Start Sites: TSSs
- ✔ Post translational modifications (15,800)
- ✔ miRNA

**Gene functional assignments**
- ✔ Disease associations (46,700)
- ✔ Drug targets
- ✔ Pathway membership
- ✔ HGMD® disease genes

**Novel variations**
- ✔ Mutation effect prediction using snpEff

**Uploaded Track**
Results using our filtering strategy in RHD patients (part 1)

Total variants in exome from peripheral blood: 42,673 variants

Quality filtering
E.g. coverage
34,786 variants

Filtering based on variant seriousness
Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELS) are retained

Filtering based on deleteriousness prediction
For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained
2,297 variants

Databased strategy
Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®

Candidate strategy
Variants in known or suspected RHD genes or gene groups (own lists) based on own literature search are retained

Filtering based on clinical features
Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
83 variants
Results using our filtering strategy in RHD patients (part 2)

- Population filtering
  The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration

- Filtering based on inheritance model
  E.g. based on information provided by e.g. OMIM, POSSUMweb

- Comparison with in house exomes

- Disease causing variants
- No disease causing variants

- Trio-based analysis for \textit{de novo} strategy or no diagnosis
Examples of read out in bilateral renal hypodysplasia patients

Patient 1

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_14_96703484</td>
<td>+</td>
<td>C&gt;T</td>
<td>BDKRB2</td>
<td>Disease</td>
<td>Kidney failure, chronic</td>
<td>1/1</td>
<td>rs1046248</td>
<td>0.084</td>
<td>T/T: 1.3%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td></td>
<td>Probably damaging</td>
</tr>
<tr>
<td>Var_15_33445438</td>
<td>-</td>
<td>G&gt;A</td>
<td>FMN1</td>
<td>Disease</td>
<td>Oligosyndactyly, radio ulnar synostosis, hearing loss, renal defect</td>
<td>0/1</td>
<td>rs149624435</td>
<td>0.01</td>
<td>G/A: 2.6%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Tolerated</td>
<td></td>
<td>Probably damaging</td>
</tr>
<tr>
<td>Var_2_110959026</td>
<td>-</td>
<td>G&gt;T</td>
<td>NPHP1</td>
<td>Disease</td>
<td>Nephronophthisis 1</td>
<td>0/1</td>
<td>rs33958626</td>
<td>0.017</td>
<td>G/T: 9.2%</td>
<td>Non-synonymous coding</td>
<td>Pro/Thr</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td>Var_10_102509529</td>
<td>+</td>
<td>G&gt;-'</td>
<td>PAX2</td>
<td>Disease</td>
<td>Renal hypodysplasia, renal-coloboma syndrome, macular abnormalities, bilateral optic nerve anomalies</td>
<td>0/1</td>
<td></td>
<td></td>
<td>Frameshift coding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dominant</td>
</tr>
</tbody>
</table>
### Examples of read out in bilateral renal hypodysplasia patients

#### Patient 1

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_14_96703484</td>
<td>+</td>
<td>C&gt;T</td>
<td>BDKRB2</td>
<td>Disease</td>
<td>Kidney failure, chronic</td>
<td>1/1</td>
<td>rs1046248</td>
<td>0.084</td>
<td>T/T: 1.3%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Dominating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_15_33445438</td>
<td>-</td>
<td>G&gt;A</td>
<td>FMN1</td>
<td>HGMD Disease Genes</td>
<td>Oligosyndactyly, radio ulnar synostosis, hearing loss, renal defect</td>
<td>0/1</td>
<td>rs14962435</td>
<td>0.01</td>
<td>G/A: 2.6%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_2_110959026</td>
<td>-</td>
<td>G&gt;T</td>
<td>NPHP1</td>
<td>HGMD Disease Genes</td>
<td>Nephronophthisis 1</td>
<td>0/1</td>
<td>rs33958626</td>
<td>0.017</td>
<td>G/T: 9.2%</td>
<td>Non-synonymous coding</td>
<td>Pro/Thr</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Recessive</td>
</tr>
<tr>
<td><strong>Var_10_102509529</strong></td>
<td>+</td>
<td>G&gt;−</td>
<td>PAX2</td>
<td>HGMD Disease Genes</td>
<td>Renal hypodyplasia, renal-coloboma syndrome, macular abnormalities, bilateral optic nerve anomalies</td>
<td>0/1</td>
<td></td>
<td></td>
<td>Frameshift coding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dominant</td>
</tr>
</tbody>
</table>

#### Patient 2

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_2_45233463</td>
<td>-</td>
<td>G&gt;A</td>
<td>SIX2</td>
<td>HGMD mutation</td>
<td>Renal hypodysplasia: 722C&gt;T</td>
<td>0/1</td>
<td>rs147806994</td>
<td>&lt;0.01</td>
<td>G/A: 0.5%</td>
<td>Non-synonymous coding</td>
<td>Pro/Leu</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Var_9_14819370</td>
<td>-</td>
<td>G&gt;T</td>
<td>FREM1</td>
<td>HGMD Disease Genes</td>
<td>Bifid nose, renal agenesis and anorectal malformations syndrome</td>
<td>1/1</td>
<td>rs7023244</td>
<td>0.134</td>
<td>T/T: 6.6%</td>
<td>Non-synonymous coding</td>
<td>Ser/Tyr</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
</tr>
</tbody>
</table>
## Examples of read out in bilateral renal hypodysplasia patients

### Patient 1

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_14_96703484</td>
<td>+</td>
<td>C&gt;T</td>
<td>BDKRB2</td>
<td>Disease</td>
<td>Kidney failure, chronic</td>
<td>1/1</td>
<td>0.084</td>
<td>T/T: 1.3%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Probable damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_15_33445438</td>
<td>-</td>
<td>G&gt;A</td>
<td>FMN1</td>
<td>HGMD Disease Genes</td>
<td>Oligosyndactyl, radio ulnar synostosis, hearing loss, renal defect</td>
<td>0/1</td>
<td>0.01</td>
<td>G/A: 2.6%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Tolerated</td>
<td>Probable damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_2_110959026</td>
<td>-</td>
<td>G&gt;T</td>
<td>NPHP1</td>
<td>HGMD Disease Genes</td>
<td>Nephronophthisis 1</td>
<td>0/1</td>
<td>0.017</td>
<td>G/T: 9.2%</td>
<td>Non-synonymous coding</td>
<td>Pro/Thr</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Var_10_102509529</td>
<td>+</td>
<td>G&gt;</td>
<td>PAX2</td>
<td>HGMD Disease Genes</td>
<td>Renal hypodysplasia, renal-coloboma syndrome, macular abnormalities,</td>
<td>0/1</td>
<td></td>
<td></td>
<td>Frameshift coding</td>
<td></td>
<td></td>
<td>Dominant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Patient 2

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_2_45233463</td>
<td>-</td>
<td>G&gt;A</td>
<td>SIX2</td>
<td>HGMD mutation</td>
<td>Renal hypodysplasia: 722C&gt;T</td>
<td>0/1</td>
<td>&lt;0.01</td>
<td>G/A: 0.5%</td>
<td>Non-synonymous coding</td>
<td>Pro/Leu</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Var_9_14819370</td>
<td>-</td>
<td>G&gt;T</td>
<td>FREM1</td>
<td>HGMD Disease Genes</td>
<td>Bifid nose, renal agenesis and anorectal malformations syndrome</td>
<td>1/1</td>
<td>0.134</td>
<td>T/T: 6.6%</td>
<td>Non-synonymous coding</td>
<td>Ser/Tyr</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
</tr>
</tbody>
</table>

Detection rate using WES plus our filtering strategy in severe bilateral renal hypodysplasia patients: 64%
Modifying the filtering strategy to increase the detection rate (part 1)

**Total variants in exome from peripheral blood**: 42,673 variants
- Quality filtering
  - E.g. coverage
  - 34,786 variants
- Filtering based on variant seriousness
  - Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELs) are retained
  - 2,297 variants
- Filtering based on deleteriousness prediction
  - For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained
  - 83 variants
- Databased strategy
  - Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®
- Filtering based on clinical features
  - Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
  - 5 variants
- Population filtering
  - The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration
- Filtering based on inheritance model
  - E.g. based on information provided by e.g. OMIM, POSSUMweb
  - 3 variants
- Comparison with in house exomes
- Disease causing variants
- No disease causing variants
- Trio-based analysis for de novo strategy or no diagnosis

---

Weber Group
Department of Human Genetics

MH Hannover Medical School
Modifying the filtering strategy to increase the detection rate (part 2)

- **Total variants in exome from peripheral blood**: 42,673 variants
  - Quality filtering: E.g. coverage
    - Filtering based on variant seriousness: Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELS) are retained
      - Filtering based on deleteriousness prediction: For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained
        - Candidate strategy: Variants in known or suspected RHD genes or gene groups (own lists) based on own literature search are retained
          - Databased strategy: Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD
            - Filtering based on clinical features: Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
              - Population filtering: The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration
                - Filtering based on inheritance model: E.g. based on information provided by e.g. OMIM, POSSUMweb
                  - Comparison with in house exomes: 3 variants
                    - Disease causing variants: No disease causing variants
                      - Trio-based analysis for de novo strategy or no diagnosis

Weber Group
Department of Human Genetics

MH Hannover Medical School
Modifying the filtering strategy to increase the detection rate (part 3)

- Total variants in exome from peripheral blood: 42,673 variants
  - Quality filtering
    - E.g. coverage: 34,786 variants
  - Filtering based on variant seriousness
    - Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELs) are retained: 2,297 variants
- Candidate strategy
  - Variants in known or suspected RHD genes or gene groups (own lists) based on own literature search are retained: 83 variants
- Databased strategy
  - Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®: filtering based on clinical features
    - Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
- Population filtering
  - The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration
  - Filtering based on inheritance model
    - E.g. based on information provided by e.g. OMIM, POSSUMweb: 5 variants
  - Comparison with in house exomes: 3 variants
- Disease causing variants: No disease causing variants

Trio-based analysis for de novo strategy or no diagnosis
Applying our filtering strategy to the analysis of patients with amyotrophic lateral sclerosis (ALS)

- Neurodegenerative disease
- Onset of symptoms usually in late adulthood
- Most cases are sporadic
- Genetically heterogeneous
- Inheritance is mostly dominant, but can be recessive
- Mutations in the genes *SOD1* and *C9ORF72* (mutated in 20% and 40-50% of familial ALS patients, respectively) were excluded in the patients analyzed
Results using our filtering strategy in ALS patients (part 1)

Total variants in exome from peripheral blood:
- Quality filtering (e.g. coverage) result: 45,287 variants
- Filtering based on variant seriousness: serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELs) are retained, resulting in 38,393 variants

Filtering based on deleteriousness prediction:
- For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained, resulting in 2,523 variants

Candidate strategy:
- Variants in known or suspected ALS genes or gene groups (own lists) based on own literature search are retained

Databased strategy:
- Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®

Filtering based on clinical features:
- Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained, resulting in 45 variants
Results using our filtering strategy in ALS patients (part 2)

- Population filtering
  The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration
  - 45 variants
  - 11 variants

- Filtering based on inheritance model
  E.g. based on information provided by e.g. OMIM, POSSUMweb
  - 9 variants
  - 6 variants

- Comparison with in house exomes
  - Disease causing variants
  - No disease causing variants

  Trio-based analysis for *de novo* strategy or no diagnosis
# Examples of read out in amyotrophic lateral sclerosis patients

## Patient 1

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs Number</th>
<th>MAF</th>
<th>Genotype frequency</th>
<th>European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_6_44272479</td>
<td>C&gt;T</td>
<td>AARS2</td>
<td>HGMD Disease Genes</td>
<td>Mitochondrial cardiomyopathy, infantile, combined oxidative phosphorylation deficiency</td>
<td>0/1</td>
<td>rs2286963</td>
<td>0.203</td>
<td>G/G: 9.0%</td>
<td>Non-synonymous coding</td>
<td>Gly/Asp</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_2_211060050</td>
<td>T&gt;G</td>
<td>ACADL</td>
<td>Disease</td>
<td>Muscle weakness</td>
<td>1/1</td>
<td>rs2286963</td>
<td>0.203</td>
<td>G/G: 9.0%</td>
<td>Non-synonymous coding</td>
<td>Lys/Gln</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_2_74588717</td>
<td>G&gt;A</td>
<td>DCTN1</td>
<td>HGMD mutation</td>
<td>Amyotrophic lateral sclerosis:3746C&gt;T</td>
<td>0/1</td>
<td>rs72466496</td>
<td>&lt;0.01</td>
<td>G/A: 0%</td>
<td>Non-synonymous coding</td>
<td>Thr/Ile</td>
<td>deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Dominant</td>
<td></td>
</tr>
<tr>
<td>Var_19_51857615</td>
<td>-</td>
<td>ETFB</td>
<td>HGMD Disease Genes</td>
<td>Electron transfer flavoprotein deficiency</td>
<td>0/1</td>
<td>rs2288750</td>
<td>0.378</td>
<td>A/A: 4.7%</td>
<td>Non-synonymous coding</td>
<td>Arg/Trp</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_2_241722445</td>
<td>G&gt;A</td>
<td>KIF1A</td>
<td>HGMD Disease Genes</td>
<td>Spastic paraparesis</td>
<td>1/1</td>
<td>rs2288750</td>
<td>0.378</td>
<td>A/A: 4.7%</td>
<td>Non-synonymous coding</td>
<td>Arg/Trp</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_17_44060859</td>
<td>+</td>
<td>MAPT</td>
<td>HGMD Disease Genes</td>
<td>FTDP/corticobasal degeneration</td>
<td>0/1</td>
<td>rs6375007</td>
<td>0.083</td>
<td>G/A: 8.4%</td>
<td>Non-synonymous coding</td>
<td>Gin/Arg</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Dominant</td>
<td></td>
</tr>
<tr>
<td>Var_10_30629226</td>
<td>-</td>
<td>MTPAP</td>
<td>HGMD Disease Genes</td>
<td>Spastic ataxia</td>
<td>1/1</td>
<td>rs1047991</td>
<td>0.248</td>
<td>A/A: 5.0%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_5_70308262</td>
<td>C&gt;T</td>
<td>NAIP</td>
<td>HGMD Disease Genes</td>
<td>Amyotrophic lateral sclerosis</td>
<td>0/1</td>
<td>rs61757629</td>
<td>0.01</td>
<td>C/T: 3.4%</td>
<td>Non-synonymous coding</td>
<td>Ala/Thr</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_13_23911234</td>
<td>G&gt;T</td>
<td>SACS</td>
<td>HGMD Disease Genes</td>
<td>Spastic ataxia, Charlevoix-Saguenay</td>
<td>0/1</td>
<td>rs146722795</td>
<td>0.005</td>
<td>G/T: 1.0%</td>
<td>Non-synonymous coding</td>
<td>Leu/Ile</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
</tbody>
</table>
### Examples of read out in amyotrophic lateral sclerosis patients

#### Patient 1

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs Number</th>
<th>MAF</th>
<th>Genotype frequency</th>
<th>European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_6_44272479 - C&gt;T</td>
<td>AARS2</td>
<td>HGMD Disease Genes</td>
<td>Mitochondrial cardiomyopathy, infantile, combined oxidative phosphorylation deficiency</td>
<td>0/1</td>
<td>rs2286963</td>
<td>0.203</td>
<td>G/G: 9.0%</td>
<td>Non-synonymous coding</td>
<td>Gly/Asp</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_2_211060050 - T&gt;G</td>
<td>ACADL</td>
<td>Disease</td>
<td>Muscle weakness</td>
<td>1/1</td>
<td>rs2286963</td>
<td>0.203</td>
<td>G/G: 9.0%</td>
<td>Non-synonymous coding</td>
<td>Lys/Gln</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_2_74588717 - G&gt;A</td>
<td>DCTN1</td>
<td>HGMD mutation</td>
<td>Amyotrophic lateral sclerosis:3746C&gt;T</td>
<td>0/1</td>
<td>rs72466496</td>
<td>&lt;0.01</td>
<td>G/A: 0%</td>
<td>Non-synonymous coding</td>
<td>Thr/Ile</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Dominant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_19_51857615 - &gt;G</td>
<td>ETFB</td>
<td>HGMD Disease Genes</td>
<td>Electron transfer flavoprotein deficiency</td>
<td>0/1</td>
<td>rs2288750</td>
<td>0.378</td>
<td>A/A: 4.7%</td>
<td>Non-synonymous coding</td>
<td>Arg/Trp</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_2_241722445 - G&gt;A</td>
<td>KIF1A</td>
<td>HGMD Disease Genes</td>
<td>Spastic paraparesis</td>
<td>1/1</td>
<td>rs2288750</td>
<td>0.378</td>
<td>A/A: 4.7%</td>
<td>Non-synonymous coding</td>
<td>Arg/Trp</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_17_44060859 + A&gt;G</td>
<td>MAPT</td>
<td>HGMD Disease Genes</td>
<td>FTDP/corticobasal degeneration</td>
<td>0/1</td>
<td>rs63750072</td>
<td>0.083</td>
<td>G/A: 8.4%</td>
<td>Non-synonymous coding</td>
<td>Gin/Arg</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Dominant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_10_30629226 - G&gt;A</td>
<td>MTPAP</td>
<td>HGMD Disease Genes</td>
<td>Spastic ataxia</td>
<td>1/1</td>
<td>rs1047991</td>
<td>0.248</td>
<td>A/A: 5.0%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_5_70308262 - C&gt;T</td>
<td>NAIP</td>
<td>HGMD Disease Genes</td>
<td>Amyotrophic lateral sclerosis</td>
<td>0/1</td>
<td>rs61757629</td>
<td>0.01</td>
<td>C/T: 3.4%</td>
<td>Non-synonymous coding</td>
<td>Ala/Thr</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_13_23911234 - G&gt;T</td>
<td>SACS</td>
<td>HGMD Disease Genes</td>
<td>Spastic ataxia, Charlevoix-Saguenay</td>
<td>0/1</td>
<td>rs146722795</td>
<td>0.005</td>
<td>G/T: 1.0%</td>
<td>Non-synonymous coding</td>
<td>Leu/Leu</td>
<td>Deleterious</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Examples of read out in amyotrophic lateral sclerosis patients

### Patient 2

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/ Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_20_57597970</td>
<td>+</td>
<td>A&gt;C</td>
<td>ATP5E</td>
<td>DNAse</td>
<td>Mitochondrial complex V deficiency</td>
<td>0/1</td>
<td>CP053743</td>
<td>0.034</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Gln/Pro</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_1_6509148</td>
<td>+</td>
<td>G&gt;C</td>
<td>ESPN</td>
<td>HGMD</td>
<td>Disease Genes</td>
<td>0/1</td>
<td></td>
<td></td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Gly/Arg</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_1_67833643</td>
<td>+</td>
<td>G&gt;A</td>
<td>IL12RB2</td>
<td>Disease</td>
<td>Deafness and vestibular areflexia</td>
<td>0/1</td>
<td>rs2307153</td>
<td>0.05</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Gly/Asp</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Var_8_19805708</td>
<td>+</td>
<td>G&gt;A</td>
<td>LPL</td>
<td>Disease</td>
<td>Paraparesis, tropical spastic</td>
<td>0/1</td>
<td>rs1801177</td>
<td>0.027</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Asp/Asn</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_9_135203530</td>
<td>.</td>
<td>A&gt;C</td>
<td>SETX</td>
<td>EVS</td>
<td>rs3739922:SETX, ALS4, juvenile</td>
<td>0/1</td>
<td>rs3739922</td>
<td>0.125</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Phe/Cys</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Var_9_135203756</td>
<td>.</td>
<td>C&gt;T</td>
<td>SETX</td>
<td>EVS</td>
<td>rs145097270:SETX, ALS4, juvenile</td>
<td>0/1</td>
<td>rs145097270</td>
<td>&lt;0.01</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Asp/Asn</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_16_89613169</td>
<td>+</td>
<td>G&gt;T</td>
<td>SPG7</td>
<td>HGMD</td>
<td>mutation:1552+1G&gt;T</td>
<td>0/1</td>
<td>rs141644720</td>
<td>&lt;0.01</td>
<td>/</td>
<td>Essential splice site</td>
<td>G/T: 0.03%</td>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_21_45820196</td>
<td>+</td>
<td>C&gt;T</td>
<td>TRPM2</td>
<td>HGMD</td>
<td>ALS and Parkinson disease</td>
<td>0/1</td>
<td>rs35288229</td>
<td>0.033</td>
<td>/</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
</tbody>
</table>
# Examples of read out in amyotrophic lateral sclerosis patients

**Patient 2**

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/ Disease</th>
<th>Genotype frequency</th>
<th>Consequence</th>
<th>AA change</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_20_57597970</td>
<td>+</td>
<td>A&gt;C</td>
<td>ATP5E</td>
<td>DNAse</td>
<td>Mitochondrial complex V deficiency</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Gln/Pro</td>
<td>Deleterious</td>
<td></td>
<td>Recessive</td>
</tr>
<tr>
<td>Var_1_6509148</td>
<td>+</td>
<td>G&gt;C</td>
<td>ESPN</td>
<td>HGMD</td>
<td>Disease Genes</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Gly/Arg</td>
<td>Deleterious</td>
<td></td>
<td>Recessive/Dominant</td>
</tr>
<tr>
<td>Var_1_67833643</td>
<td>+</td>
<td>G&gt;A</td>
<td>IL12RB2</td>
<td>Disease</td>
<td>Paraparesis, tropical spastic</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Gly/Asp</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td>Var_8_19805708</td>
<td>+</td>
<td>G&gt;A</td>
<td>LPL</td>
<td>Disease</td>
<td>Sleep apnea, obstructive</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Asp/Asn</td>
<td>Deleterious</td>
<td></td>
<td>Recessive/Dominant</td>
</tr>
<tr>
<td>Var_9_135203530</td>
<td>.</td>
<td>A&gt;C</td>
<td>SETX</td>
<td>Disease</td>
<td>rs3739922:SETX, ALS4, juvenile</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Phe/Cys</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td>Var_9_135203756</td>
<td>.</td>
<td>C&gt;T</td>
<td>SETX</td>
<td>Disease</td>
<td>rs145097270:SETX, ALS4, juvenile</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Asp/Asn</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
<td>Dominant</td>
</tr>
<tr>
<td>Var_16_89613169</td>
<td>+</td>
<td>G&gt;T</td>
<td>SPG7</td>
<td>HGMD</td>
<td>mutation Spastic paraplegia:1552+1G&gt;T</td>
<td>0/1</td>
<td>Essential splice site</td>
<td></td>
<td>Recessive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_21_45820196</td>
<td>+</td>
<td>C&gt;T</td>
<td>TRPM2</td>
<td>HGMD</td>
<td>Disease Genes</td>
<td>0/1</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Detection rate using WES plus our filtering strategy in amyotrophic lateral sclerosis patients: 50%
Modifying the filtering strategy for the analysis of two or more patients or siblings affected by the same rare disease or tumor (part 1)

- **Total variants in exome from peripheral blood**
  - 44,711 variants
  - 45,709 variants

- **Overlapping strategy**
  - Variants shared by the two affected siblings including heterogenous, homozygous and compound heterozygous variants are retained
  - 29,721 variants

- **Quality filtering**
  - E.g. coverage

- **Filtering based on variant seriousness**
  - Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELs) are retained

- **Filtering based on deleteriousness prediction**
  - For non-synonymous coding variants: variants indicated to be deleterious in at least one prediction program (e.g. SIFT, PolyPhen2, Condel, MutationTaster) are retained
  - 1,556 variants + 316 variants = 1,872 variants
Modifying the filtering strategy for the analysis of two or more patients or siblings affected by the same rare disease or tumor (part 2)

Candidate strategy
Variants in known or suspected disease genes or gene groups (own lists) based on own literature search are retained

1,556 variants + 316 variants = 1,872 variants

Databased strategy
Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®

Filtering based on clinical features
Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained

Population filtering
The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration

167 variants

33 variants

Comparison with in house exomes

32 variants

Comparison with variants of parents and other siblings

5 variants

Disease causing variants
**Read out for two siblings affected by the same rare tumor**

**Overlapping heterozygous variants**

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF actual</th>
<th>MAF European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_5_112176431</td>
<td>+</td>
<td>G&gt;A</td>
<td>APC</td>
<td>HGMD Disease Genes</td>
<td>Colorectal cancer</td>
<td>0/1</td>
<td>rs148275069</td>
<td>&lt;0.01</td>
<td>0.003</td>
<td>Non synonymous coding</td>
<td>Asp/Asn</td>
<td>Deleterious</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
</tr>
<tr>
<td>Var_3_142281353</td>
<td>-</td>
<td>C&gt;G</td>
<td>ATR</td>
<td>HGMD Disease Genes</td>
<td>Oropharyngeal cancer syndrome</td>
<td>0/1</td>
<td>rs2229033</td>
<td>0.01</td>
<td>0.016</td>
<td>Non synonymous coding</td>
<td>Lys/Asn</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td>Var_3_119526203</td>
<td>+</td>
<td>G&gt;A</td>
<td>NR112</td>
<td>HGMD Disease Genes</td>
<td>Head and neck cancer, second primary tumour, association</td>
<td>0/1</td>
<td>rs12721607</td>
<td>0.01</td>
<td>0.021</td>
<td>Non synonymous coding</td>
<td>Gly/Arg</td>
<td>Tolerated</td>
<td>Possibly damaging</td>
<td>Deleterious</td>
</tr>
<tr>
<td>Var_1_46714242</td>
<td>+</td>
<td>A&gt;G</td>
<td>RAD54L</td>
<td>HGMD Disease Genes</td>
<td>Breast cancer</td>
<td>0/1</td>
<td>rs28363192</td>
<td>0.033</td>
<td>0.016</td>
<td>Non synonymous coding</td>
<td>Asp/Gly</td>
<td>Deleterious</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td>Var_19_44057574</td>
<td>-</td>
<td>G&gt;A</td>
<td>XRCC1</td>
<td>HGMD Disease Genes, Disease</td>
<td>Increased lung cancer risk, association with</td>
<td>0/1</td>
<td>rs1799782</td>
<td>0.153</td>
<td>0.09</td>
<td>Non synonymous coding</td>
<td>Arg/Trp</td>
<td>Deleterious</td>
<td>Probably damaging</td>
<td>Deleterious</td>
</tr>
</tbody>
</table>
Modifying the filtering strategy for presumed recessive cases: Two-hit strategy

- **Total variants in exome from peripheral blood**: 42,617 variants

  - **Filtering based on homozygosity**: Homozygous variants are retained
    - 16,518 variants
      - Quality filtering
        - E.g. coverage
      - 15,185 variants
        - Filtering based on variant seriousness
          - Serious variants (non-synonymous coding, frameshift coding, stop gained/lost, essential splice site, complex INDELs) are retained
          - 752 variants
            - Databased strategy
              - Processing of variants using mining tools such as Genome Trax™ providing e.g. association with disease features based on e.g. annotations on inherited disease genes from HGMD®
              - 663 variants
                - Filtering based on clinical features
                  - Variants in genes associated with clinical disease features (own list) based on mining tool information e.g. Genome Trax™ are retained
                  - 46 variants
                    - Population filtering
                      - The frequency of variants (MAF and genotype frequency in population analyzed provided by e.g. ENSEMBL, exome variant server) is taken into consideration
                      - 4 variants
                        - 745 variants
                          - 699 variants
                            - 44 variants
                              - 9 variants

Weber Group
Department of Human Genetics

MH Hannover Medical School
### Read out for patient with presumed recessive disease

**Homozygous variants**

<table>
<thead>
<tr>
<th>Variant</th>
<th>Strand</th>
<th>Variation</th>
<th>Gene</th>
<th>Type</th>
<th>Feature/ Disease</th>
<th>Genotype</th>
<th>rs number</th>
<th>MAF</th>
<th>Genotype frequency European population</th>
<th>Consequence</th>
<th>AA change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>Condel</th>
<th>Inheritance model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var_3_158367837</td>
<td>+</td>
<td>C&gt;T</td>
<td>GFM1</td>
<td>Disease</td>
<td>Mitochondrial diseases, liver diseases, brain damage, chronic, Combined oxidative phosphorylation deficiency</td>
<td>1/1</td>
<td>rs56167308</td>
<td>0.148</td>
<td>T/T: 4.2%</td>
<td>Non-synonymous coding</td>
<td>Arg/Cys</td>
<td>Deleterious</td>
<td>Unknown</td>
<td>Recessive</td>
<td></td>
</tr>
<tr>
<td>Var_6_29911198</td>
<td>+</td>
<td>T&gt;C</td>
<td>HLA-A</td>
<td>Disease</td>
<td>e.g. Hepatitis C, chronic, liver cirrhosis, biliary, Hepatitis B</td>
<td>1/1</td>
<td>rs1059516</td>
<td>0.148</td>
<td></td>
<td>Non-synonymous coding</td>
<td>Ser/Pro</td>
<td>Deleterious</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var_1_66036441</td>
<td>+</td>
<td>A&gt;G</td>
<td>LEPR</td>
<td>Disease</td>
<td>HGMD Disease Genes Altered carbohydrate metabolism</td>
<td>1/1</td>
<td>CM032948</td>
<td>0.423</td>
<td>G/G: 9.5%</td>
<td>Non-synonymous coding</td>
<td>Lys/Arg</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td></td>
</tr>
<tr>
<td>Var_2_231077676</td>
<td>-</td>
<td>G&gt;A</td>
<td>SP110</td>
<td>HGMD Disease Genes</td>
<td>Hepatic veno-occlusive disease with immunodeficiency</td>
<td>1/1</td>
<td>rs11556887</td>
<td>0.058</td>
<td>A/A: 0.5%</td>
<td>Non-synonymous coding</td>
<td>Ala/Val</td>
<td>Tolerated</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Recessive</td>
</tr>
</tbody>
</table>
Evaluating WES findings: using online resources, literature search, and the patient’s clinical data

OMIM entry

#235550

HEPATIC VENOOCCLUSIVE DISEASE WITH IMMUNODEFICIENCY; VODI

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac</td>
<td>Endocardial fibrosis</td>
</tr>
<tr>
<td>HEENT</td>
<td>Microcephaly [EoM image]</td>
</tr>
<tr>
<td>Immunology</td>
<td>Immune deficidency</td>
</tr>
<tr>
<td></td>
<td>Hypogammaglobulinemia</td>
</tr>
<tr>
<td></td>
<td>Multiple infections</td>
</tr>
<tr>
<td></td>
<td>Lymphoid germinal center defect</td>
</tr>
<tr>
<td></td>
<td>Mature plasma cell deficiency</td>
</tr>
<tr>
<td>Liver</td>
<td>Venoocclusive disease</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Autosomal recessive</td>
</tr>
</tbody>
</table>

Clinical Immunology (2012) 145, 102–107

Hepatic veno-occlusive disease with immunodeficiency (VODI): First reported case in the U.S. and identification of a unique mutation in Sp110

Tiffany Wang a, Peck Ong b, Tony Roscioli c, d, e, Simon T. Cliffe c, Joseph A. Church b, *

Abstract  Familial hepatic veno-occlusive disease with immunodeficiency (VODI, OMIM: 235550), a rare form of severe combined immune deficiency, was first described in Australian Lebanese patients as being associated with homozygous mutations in SP110, a gene encoding a PML nuclear body-associated protein. We present the first case of confirmed VODI in the United States, and identify the first novel missense mutation in SP110.

The 3-year-old daughter of Hispanic parents without known consanguinity presented at age 5 months with fever, hepatomegaly, and pancytopenia. Her brother aged at age 3 months from hepatic failure of undetermined etiology. Initial T- and B-cell counts were low, but eventually normalized. Serum IgG and IgM levels were low for age. Lymphoproliferation to mitogens and allogenic B-cells was normal, but absent to tetanus and candida antigens. Serum antibody levels against pneumococcal, HIB and tetanus antigens were low. Liver biopsies at ages 5 and 9 months were consistent with hepatic veno-occlusive disease or hVOD (also known as sinusoidal obstruction syndrome or SOS) and broncho-aveolar lavage detected Pneumocystis jiroveci. The patient recovered from her acute disease and has been clinically stable on immunoglobulin replacement therapy and trimethoprim-sulfamethoxazole prophylaxis. T-Cell receptor excision circle (TREC) analysis suggests that VODI will not be detected by newborn screening for severe combined immunodeficiency that relies on this assay.
Navigating the Sea of Variants with precision and ease: Putting together your own tool kit for WES data interpretation
Acknowledgments

Collaborators in projects presented:
Dieter Haffner, MHH
Lars Pape, MHH
Ulrich Baumann, MHH
Susanne Petri, MHH
Christian Hartmann, MHH
Carl Friedrich Classen, Rostock

Sponsors:

ELSE KRÖNER-FRESENIUS-STIFTUNG

Deutsche Krebshilfe
Helfen. Forschen. Informieren.

DFG Deutsche Forschungsgemeinschaft

Weber Group
Department of Human Genetics

MHH Hannover Medical School
Science Webinar Series

A Traveler’s Guide to Next Generation Sequencing: Navigating the Sea of Genetic Variants

26 June 2013

Brought to you by the Science/AAAS Custom Publishing Office

Participating Experts:

Scott D. Kahn, Ph.D.
Illumina
San Diego, CA

Ruthild Weber M.D.
Hannover Medical School
Hannover, Germany

Q&A
To submit your questions, type them into the text box and click Submit.

Sponsored by:

BIOBASE
BIOLOGICAL DATABASES
Look out for more webinars in the series at:

webinar.sciencemag.org

To provide feedback on this webinar, please e-mail your comments to webinar@aaas.org

For related information on this webinar topic, go to:

www.biobase-international.com/product/genome-trax