Challenges of Clinical Implementation of Genomic Medicine

Gholson J. Lyon, M.D. Ph.D.
INFORMED CONSENT AUTHORIZATION TO PARTICIPATE IN A CLINICAL INVESTIGATION

Family Name: ____________________________
Title: (Protocol #: 100) Study of the Genetic Causes of Complex Neurologic Psychiatric Disorders

Version: 14-Apr-2011
Protocol: 100

APPROVED BY
Independent IRB

Signature 14-Apr-2011
Date
Penetrance and Expressivity

• We do not really know the penetrance or expressivity of pretty much ALL mutations in humans, as we have not systematically sequenced or karyotyped any genetic alteration in MILLIONS of well-phenotyped people.

• Do single mutations drive outcome predominately, or are the results modified substantially by other mutations and/or environment? Is there really such a thing as genetic determinism for MANY mutations?
Ancestry Matters! - Ogden Syndrome

The mutation is **necessary**, but we do not know if it is **sufficient** to cause this phenotype in ANY genetic background. It simply “contributes to” the phenotype.
The mutation disrupts the N-terminal acetylation machinery (NatA) in human cells.
Simulated structure of S37P mutant

Max Doerfel

Yiyang Wu
hNaa10p-S37P is functionally impaired *in vivo* using a yeast model.

Unpublished data from Thomas Arnesen, do not further distribute.
Proteomics Analysis of EBV-transformed cell lines from family members

- **I.**
  - B.
  - WT
  - +

- **II.**
  - A.
  - +
  - MUT
  - mut
  - +/mut
  - carrier

- **III.**
  - 1.
  - WT
  - +/-
  - proband
  - WT
  - (+/mut)
  - carrier
  - WT
  - (+/mut)
  - carrier
  - MUT

- **SB**
  - = stillborn

- **Proband**
  - = patient samples analyzed by N-terminal COFRADIC analyses (#1 to #5)

- **Patient samples prepared for N-terminal COFRADIC analyses (but still to be analyzed)**
  - (#8 and #9)

- **Patient samples analyzed by N-terminal COFRADIC analyses**
  - (#1 to #5)

- **Legend:**
  - □ = male 
  - ○ = female
  - □ = FFPE DNA (for patient III.7.) or DNA from blood available (and for some of them: EBV transformed cell lines available + skin fibroblast of patient III.6.)
  - SB = stillborn
  - = proband

**Family Members:**

- **III.4.** proband hemizygous, mutant (89323) (#1a) (#1b)
- **II.2.** mother of proband, carrier (89324) (#2)
- **II.A.** married-in father of proband, WT(89325)
- **III.2.** brother of proband, WT(90526) (#3)
- **III.1.** sister of proband, WT (90527) (#4)
- **I.2.** grandmother of proband, carrier (90528)
- **I.1.** married-in grandfather of proband, WT(90529) (#5)
- **II.7.** aunt of proband, WT (90530) (#6)
- **II.3.** aunt of proband, carrier (90531)
- **II.B.** married-in uncle of proband, WT(90532) (#8)
- **II.8.** uncle of proband, WT(90688) (#9)
- **II.5.** aunt of proband, carrier with deceased boy (90797) (#7)
Scatterplots displaying the correlation of the degrees of Nα-acetylation when comparing a control (brother WT)(in this case Y-axis) and the proband or mother (carrier) (Y-axis) N-terminome datasets. The N-termini displaying a significant variation in the degree of Nα-acetylation (see above) are highlighted in orange.
Results from EBV-transformed lymphocytes

<table>
<thead>
<tr>
<th>Peptide</th>
<th>AA</th>
<th>BB</th>
<th>CC</th>
<th>DD</th>
<th>EE</th>
<th>FF</th>
<th>GG</th>
<th>HH</th>
<th>II</th>
<th>JJ</th>
<th>KK</th>
<th>LL</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAEDGGPGEQNR</td>
<td>66</td>
<td>86</td>
<td>92</td>
<td>91</td>
<td>87</td>
<td>84</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AADTQVSETLKR</td>
<td>52</td>
<td>80</td>
<td>84</td>
<td>84</td>
<td>80</td>
<td>81</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAESALQVEKLOAR</td>
<td>58</td>
<td>89</td>
<td>92</td>
<td>92</td>
<td>87</td>
<td>87</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVFAADLDLR</td>
<td>66</td>
<td>95</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>95</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVEKEAAGGGISEEAAQYDR</td>
<td>69</td>
<td>90</td>
<td>91</td>
<td>94</td>
<td>91</td>
<td>95</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLGAPDESSVR</td>
<td>51</td>
<td>79</td>
<td>74</td>
<td>79</td>
<td>70</td>
<td>72</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLSPEAER</td>
<td>74</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>96</td>
<td>97</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAGGGGGSSKASSASSAGALESSLDR</td>
<td>72</td>
<td>85</td>
<td>84</td>
<td>85</td>
<td>82</td>
<td>84</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEEANDKKPTTKFELER</td>
<td>79</td>
<td>91</td>
<td>93</td>
<td>93</td>
<td>94</td>
<td>93</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>AC</th>
<th>BC</th>
<th>CC</th>
<th>DD</th>
<th>EE</th>
<th>FF</th>
<th>GG</th>
<th>HH</th>
<th>II</th>
<th>JJ</th>
<th>KK</th>
<th>LL</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q92616</td>
<td>GCN1L</td>
<td>Translational activator GCN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14241</td>
<td>ELOA1</td>
<td>Transcription elongation factor B polypeptide 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P78346</td>
<td>RPP30</td>
<td>Ribonuclease P protein subunit p30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9UBE0</td>
<td>SAE1</td>
<td>SUMO-activating enzyme subunit 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7Z456</td>
<td>KI21A</td>
<td>Kinesin-like protein KIF21A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9NUG6</td>
<td>PDRG1</td>
<td>p53 and DNA damage-regulated protein 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5VT52</td>
<td>RPRD2</td>
<td>Regulation of nuclear pre-mRNA domain-containing protein 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q92989</td>
<td>CLP1</td>
<td>Polyrribonucleotide 5'-hydroxyl-kinase Clp1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

New Syndrome with Dysmorphology, Mental Retardation, “Autism”, “ADHD”

Likely X-linked or Autosomal Recessive, with X-linked being supported by extreme X-skewing in the mother
<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5 years</td>
</tr>
<tr>
<td>3.5 years</td>
</tr>
<tr>
<td>7 years</td>
</tr>
<tr>
<td>3 years</td>
</tr>
<tr>
<td>5 years</td>
</tr>
<tr>
<td>9 years</td>
</tr>
</tbody>
</table>
Workup Ongoing for past 10 years

• Numerous genetic tests negative, including negative for Fragile X and many candidate genes.

• No obvious pathogenic CNVs – microarrays normal.

• Sequenced whole genomes of Mother, Father and Two Boys, using Complete Genomics, obtained data in June of this year, i.e. version 2.0 CG pipeline.
Complete Genomics chemistry - combinatorial probe anchor ligation (cPAL)

Diagram showing the process of reading bases 1-5 and 6-10, with different probes and anchor ligation steps.

Legend:
- Ad1, Ad2, Ad3, Ad4, Ad5, Ad6
- Insertion of DNA
- Bulk of genomic DNA
- DNB adaptor/anchor binding site
- Genomic sequence
- Common probes
- Degenerate anchor

Diagram highlights the ligation process with colored arrows and annotations.
22,174
Located within a coding region

272
Located on the X chromosome

56
X-linked model of inheritance (shared between boys + mother, not in father)

7
< 1% frequency in dbSNP135

6
< 1% frequency in 1k Genomes Phase 1 data

5
< 1% frequency in NHLBI6500 exomes

3
Protein change
### Variant classification

<table>
<thead>
<tr>
<th>Variant</th>
<th>Reference</th>
<th>Alternate</th>
<th>Classification</th>
<th>Gene</th>
<th>Transcript 1</th>
<th>Exon 1</th>
<th>HGVS Coding 1</th>
<th>HGVS Protein 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X:47307978-SNV</td>
<td>G</td>
<td>T</td>
<td>Nonsyn SNV</td>
<td>ZNF41</td>
<td>NM_007130</td>
<td>5</td>
<td>c.1191C&gt;A</td>
<td>p.As397Glu</td>
</tr>
<tr>
<td>X:63444792-SNV</td>
<td>C</td>
<td>A</td>
<td>Nonsyn SNV</td>
<td>ASB12</td>
<td>NM_130388</td>
<td>2</td>
<td>c.739G&gt;T</td>
<td>p.Gly247Cys</td>
</tr>
<tr>
<td>X:70621541-SNV</td>
<td>T</td>
<td>C</td>
<td>Nonsyn SNV</td>
<td>TAF1</td>
<td>NM_004606</td>
<td>25</td>
<td>c.4010T&gt;C</td>
<td>p.lle1337Thr</td>
</tr>
</tbody>
</table>

### SIFT classification

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Reference</th>
<th>Coding?</th>
<th>SIFT Score</th>
<th>Score &lt;= 0.05</th>
<th>Ref/Alt Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>47307978</td>
<td>G</td>
<td>YES</td>
<td>0.649999976</td>
<td>0</td>
<td>G/T</td>
</tr>
<tr>
<td>X</td>
<td>63444792</td>
<td>C</td>
<td>YES</td>
<td>0</td>
<td>1</td>
<td>C/A</td>
</tr>
<tr>
<td>X</td>
<td>70621541</td>
<td>T</td>
<td>YES</td>
<td>0.00999999776</td>
<td>1</td>
<td>T/C</td>
</tr>
</tbody>
</table>

### VAAST score

<table>
<thead>
<tr>
<th>RANK</th>
<th>Gene</th>
<th>p-value</th>
<th>p-value-ci</th>
<th>Score</th>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASB12</td>
<td>1.56E-11</td>
<td>1.55557809307134e-11,0.000290464582480396</td>
<td>38.63056297</td>
<td>chrX:63444792;38.63;C-&gt;A;G-&gt;C;0,3</td>
</tr>
<tr>
<td>2</td>
<td>TAF1</td>
<td>1.56E-11</td>
<td>1.55557809307134e-11,0.000290464582480396</td>
<td>34.51696816</td>
<td>chrX:70621541;34.52;T-&gt;C;l-&gt;T;0,3</td>
</tr>
<tr>
<td>3</td>
<td>ZNF41</td>
<td>1.56E-11</td>
<td>1.55557809307134e-11,0.000290464582480396</td>
<td>32.83011803</td>
<td>chrX:47307978;32.83;G&gt;T;D-&gt;E;0,3</td>
</tr>
</tbody>
</table>
Mutations in the ZNF41 Gene Are Associated with Cognitive Deficits: Identification of a New Candidate for X-Linked Mental Retardation

Sarah A. Shoichet, Kirsten Hoffmann, Corinna Menzel, Udo Trautmann, Bettina Moser, Maria Hoeltzenbein, Bernard Echenne, Michael Partington, Hans van Bokhoven, Claude Moraine, Jean-Pierre Fryns, Jamel Chelly, Hans-Dieter Rott, Hans-Hilger Ropers, and Vera M. Kalscheuer

1Max-Planck-Institute for Molecular Genetics, Berlin; 2Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen-Nuremberg; 3Centre Hospitalier Universitaire de Montpellier, Hôpital Saint-Eloi, Montpellier, France; 4Hunter Genetics and University of Newcastle, Waratah, Australia; 5Department of Human Genetics, University Medical Centre, Nijmegen, The Netherlands; 6Services de Génétique–INSERM U316, CHU Bretonneau, Tours, France; 7Center for Human Genetics, Clinical Genetics Unit, Leuven, Belgium; and 8Institut Cochin de Génétique Moléculaire, Centre National de la Recherche Scientifique/INSERM, CHU Cochin, Paris

The mutation in ZNF41 may **NOT** be necessary, and it is certainly **NOT** sufficient to cause the phenotype.
So, of course we need baseline whole genome sequencing on everyone to at least understand the DNA genetic background in each pedigree or clan.

Ancestry Matters!
How do we get to “whole” genome sequencing for everyone?

• Tool Building for Human Genetics
Toward more comprehensive “personal genomes”

• Can we reliably detect a comprehensive, and accurate, set of variants using more than one pipeline, or even more than one sequencing platform?

• How much data is enough, and how reliable and reproducible are variant calls?
Moving Exome and WGS into a Clinical Setting requires both Analytic and Clinical Validity

• Analytical Validity: the test is accurate with high sensitivity and specificity.

• Clinical Validity: Given an accurate test result, what impact and/or outcome does this have on the individual person?
Understand Your Genome Symposium

During this two-day educational event, industry experts will discuss the clinical implementation of whole-genome next-generation sequencing (NGS) technology.

Individual Genome Sequence Results

Ordering Physician:
Gholson Lyon, MD
Steinmann Institute
10 West Broadway, Suite #820
Salt Lake City, UT 84101

Clinical Report

www.everygenome.com
CLIA#: 05D1092911
• ~$3000 for 30x “whole” genome as part of Illumina Genome Network on a research basis only, but ~$5,000 for whole genome performed in a CLIA lab at Illumina.
2-3 rounds of sequencing at BGI to attain goal of >80% of target region at >20 reads per base pair

<table>
<thead>
<tr>
<th>Exome Capture Statistics</th>
<th>K24510-84060</th>
<th>K24510-92157-a</th>
<th>K24510-84615</th>
<th>K24510-88962</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target region (bp)</td>
<td>46,401,121</td>
<td>46,401,121</td>
<td>46,401,121</td>
<td>46,257,379</td>
</tr>
<tr>
<td>Raw data yield (Mb)</td>
<td>12,490</td>
<td>14,571</td>
<td>14,129</td>
<td>9,398</td>
</tr>
<tr>
<td>Reads mapped to genome</td>
<td>110,160,277</td>
<td>135,603,094</td>
<td>135,087,576</td>
<td>83,942,646</td>
</tr>
<tr>
<td>Reads mapped to target region</td>
<td>68,042,793</td>
<td>84,379,239</td>
<td>80,347,146</td>
<td>61,207,116</td>
</tr>
<tr>
<td>Data mapped to target region (Mb)</td>
<td>5,337.69</td>
<td>6,647.18</td>
<td>6,280.01</td>
<td>4,614.47</td>
</tr>
<tr>
<td><strong>Mean depth of target region</strong></td>
<td><strong>115.03</strong></td>
<td><strong>143.25</strong></td>
<td><strong>135.34</strong></td>
<td><strong>99.76</strong></td>
</tr>
<tr>
<td>Coverage of target region (%)</td>
<td>0.9948</td>
<td>0.9947</td>
<td>0.9954</td>
<td>0.9828</td>
</tr>
<tr>
<td>Average read length (bp)</td>
<td>89.91</td>
<td>89.92</td>
<td>89.95</td>
<td>89.75</td>
</tr>
<tr>
<td>Fraction of target covered &gt;=4X</td>
<td>98.17</td>
<td>98.38</td>
<td>98.47</td>
<td>94.25</td>
</tr>
<tr>
<td>Fraction of target covered &gt;=10X</td>
<td>95.18</td>
<td>95.90</td>
<td>95.97</td>
<td>87.90</td>
</tr>
<tr>
<td><strong>Fraction of target covered &gt;=20X</strong></td>
<td><strong>90.12</strong></td>
<td><strong>91.62</strong></td>
<td><strong>91.75</strong></td>
<td><strong>80.70</strong></td>
</tr>
<tr>
<td>Fraction of target covered &gt;=30X</td>
<td>84.98</td>
<td>87.42</td>
<td>87.67</td>
<td>74.69</td>
</tr>
<tr>
<td>Capture specificity (%)</td>
<td>61.52</td>
<td>62.12</td>
<td>59.25</td>
<td>73.16</td>
</tr>
<tr>
<td>Fraction of unique mapped bases on or near target</td>
<td>65.59</td>
<td>65.98</td>
<td>63.69</td>
<td>85.46</td>
</tr>
<tr>
<td>Gender test result</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>F</td>
</tr>
</tbody>
</table>
Depth of Coverage in 15 exomes > 20 reads per bp in target region

![Bar chart showing coverage depth and fraction of target covered.](chart.png)
Pipelines Used on Same Set of Seq Data by Different Analysts, using Hg19 Reference Genome

1) BWA - **GATK** (version 1.5) with recommended parameters (GATK IndelRealigner, base quality scores were re-calibrated by GATK Table Recalibration tool. Genotypes called by GATK UnifiedGenotyper. For SNVs and indels.

2) BWA - **SamTools** version 0.1.18 to generate genotype calls -- The “mpileup” command in SamTools was used for identify SNVs and indels.

3) **SOAP-Align** – SOAPsnp for SNVs– and BWA-SOAPindel (adopts local assembly based on an extended de Bruijn graph) for indels.

4) **GNUMAP-SNP** (probabilistic Pair-Hidden Markov which effectively accounts for uncertainty in the read calls as well as read mapping in an unbiased fashion), for SNVs only.

5) BWA - Sam format to Bam format - Picard to remove duplicates – **SNVer**, for SNVs only
Total mean overlap, plus or minus one standard deviation, observed between three indel calling pipelines: GATK, SOAP-indel, and SAMTools. 

a) Mean overlap when indel position was the only necessary agreement criterion. 

b) Mean overlap when indel position, base length and base composition were the necessary agreement criteria.
• How reliable are variants that are uniquely called by individual pipelines?

• Are some pipelines better at detecting rare, or novel variants than others?
Cross validation using orthogonal sequencing technology

(Complete Genomics)
### What is the “True” Personal Genome?

<table>
<thead>
<tr>
<th></th>
<th>Illumina SNVs</th>
<th>CG SNVs</th>
<th>Illumina indels</th>
<th>CG Indels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>35653</td>
<td>19407</td>
<td>4364</td>
<td>2613</td>
</tr>
<tr>
<td>SNVs</td>
<td>18331</td>
<td></td>
<td>2666</td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>48.6%</td>
<td></td>
<td>50.5%</td>
<td></td>
</tr>
<tr>
<td>Indels</td>
<td>17322</td>
<td></td>
<td>1698</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2085</td>
<td></td>
<td></td>
<td>915</td>
</tr>
<tr>
<td></td>
<td>5.5%</td>
<td></td>
<td>17.3%</td>
<td></td>
</tr>
</tbody>
</table>
Percent of Illumina SNVs validated by CG data

Personal genome “K8101”

Known

Novel

All
Higher Validation of SNVs with the BWA-GATK pipeline

- Reveals higher validation rate of unique-to-pipeline variants, as well as uniquely discovered novel variants, for the variants called by BWA-GATK, in comparison to the other 4 pipelines (including SOAP).
Much Higher Validation of the Concordantly Called Variants (by the CG data)
Validating Indels with Complete Genomics Data for the 3 pipelines
Clinical Validity?

This is SO complex that the only solid way forward is with a “networking of science” model, i.e. online database with genotype and phenotype longitudinally tracked.
REVIEW

Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress

Gholson J Lyon*1,2 and Kai Wang*2,3
Clinical Validity with Worldwide Human Genotype-Phenotype “database”?
Conclusions

- Ancestry, i.e. genetic background, matters!
- We need to sequence whole genomes of large pedigrees, and then construct super-family structures, starting in Utah.
- Collectively, we need to improve the accuracy of “whole” genomes, and also enable the sharing of genotype and phenotype data broadly, among researchers, the research participants and consumers.
Figure 4. NAT activity of recombinant hNaa10p WT or p.Ser37Pro towards synthetic N-terminal peptides. A) and B) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 µM for SESSS and 250 µM for DDDIA) and saturated levels of acetyl-CoA (400 µM). Aliquots were collected at indicated time points and the acetylation reactions were quantified using reverse phase HPLC peptide separation. Error bars indicate the standard deviation based on three independent experiments. The five first amino acids in the peptides are indicated, for further details see materials and methods. Time dependent acetylation reactions were performed to determine initial velocity conditions when comparing the WT and Ser37Pro NAT-activities towards different oligopeptides.

C) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 µM for SESSS and AVFAD, and 250 µM for DDDIA and EEEIA) and saturated levels of acetyl-CoA (400 µM) and incubated for 15 minutes (DDDIA and EEEIA) or 20 minutes (SESSS and AVFAD), at 37°C in acetylation buffer. The acetylation activity was determined as above. Error bars indicate the standard deviation based on three independent experiments. Black bars indicate the acetylation capacity of the MBP-hNaa10p wild type (WT), while white bars indicate the acetylation capacity of the MBP-hNaa10p mutant p.Ser37Pro. The five first amino acids in the peptides are indicated.