Biobanking and metabolomics

Jeffrey L. Neul M.D., Ph.D.

Division Chief of Child Neurology
Professor and Vice Chair
Department of Neurosciences
University of California, San Diego
Rady Children’s Hospital – San Diego
Disclosures

- National Institutes of Health
- Rettsyndrome.org
  - Grant support
  - Medical Advisory Board
- Rett Syndrome Research Trust
  - Grant support
- Neuren Pharmaceuticals
  - Research support
- Newron Pharmaceuticals
  - Consulting
- Eloxx Pharmaceuticals
  - Research support
Rett syndrome Natural History Study

- Multi-center longitudinal study
  - UAB – Alan Percy, PI
  - Baylor – Dan Glaze, Jeff Neul PI
  - Greenwood Genetics Center – Steve Skinner PI
  - Children’s Hospital Boston – Walter Kaufmann PI
- U54 mechanism, funded by NIH, ORD, NICHD
- Enrolled > 1200 people with clinically defined RTT and/or MECP2 mutations
- Started in 2004
Neul Biobanking Protocol

- Started in 2012
- Linked to the NHS

**Sample Isolation** → **DNA**

- XCI
- BDNF SNP

→ Assess contribution to clinical variation

- Exome sequencing

→ Identify genetic modifiers

- Plasma → Metabolic profiling

→ Identify biomarkers

- RNA → Transcriptional profiling (future)

- Fibroblast cell lines → iPSC (future)

→ Future project

- Identify biomarkers

- Functional studies (Future)

<table>
<thead>
<tr>
<th>Total</th>
<th>Proband</th>
<th>Mother</th>
<th>Father</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled</td>
<td>743</td>
<td>356</td>
<td>189</td>
<td>141</td>
</tr>
<tr>
<td>DNA</td>
<td>665</td>
<td>320</td>
<td>175</td>
<td>124</td>
</tr>
<tr>
<td>RNA</td>
<td>528</td>
<td>230</td>
<td>150</td>
<td>102</td>
</tr>
<tr>
<td>Plasma</td>
<td>515</td>
<td>226</td>
<td>150</td>
<td>102</td>
</tr>
</tbody>
</table>
Metabolomics Study

- **Goal:** Identify molecular features in plasma that correlate with disease state and severity.

- **Design:**
  - Commercial metabolomics: Metabolon
  - Samples: 226 probands, 37 siblings
  - Comparisons:
    - Severe (n=37) compared to mild (n=40)
    - Correlation with severity (n=226)
    - Affected (n=34) to unaffected sibling (n=37)
### Statistical Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Welch's Two-Sample t-Test</th>
<th>RM ANOVA</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Severe Mild</td>
<td>Proband Sib</td>
<td>Clinical Severity Score</td>
</tr>
<tr>
<td>Total biochemicals</td>
<td>59</td>
<td>94</td>
<td>82</td>
</tr>
<tr>
<td>( p \leq 0.05 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biochemicals ( \uparrow \downarrow )</td>
<td>23</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>Total biochemicals</td>
<td>28</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>( 0.05 &lt; p &lt; 0.10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biochemicals ( \uparrow \downarrow )</td>
<td>13</td>
<td>15</td>
<td>9</td>
</tr>
</tbody>
</table>

295 named compounds
q-value < 0.10
Severe versus mild

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>40</td>
</tr>
<tr>
<td>Severe</td>
<td>37</td>
</tr>
</tbody>
</table>

**MeanDecreaseAccuracy**

- creatinine
- indoleacetate
- dimethyl-sulfone
- 7-methylxanthine
- stearidonate (18:4n3)
- pregnen-diol-disulfate-
- hippurate
- 1-myristoylglycerophosphocholine*14-0-
- andro-steroid-monosulfate-2-
- 1-pentadecanoylglycerophosphocholine...
- glycerate
- trans-4-hydroxyproline
- Mannose
- 4-vinylphenol sulfate
- paraxanthine
- kynurenine
- 2-aminooctanoate
- glycohyocholate
- urea
- 4-androsten-3beta-17beta-diol-disulfate...
- 1-myristoylglycerophosphocholine*14-0-
- hippurate
- pregnen-diol-disulfate-
- stearidonate (18:4n3)
- 7-methylxanthine
- dimethyl-sulfone
- indoleacetate
- creatinine

**80% predictive accuracy**

**Random Forest Analysis**

**Increasing Importance to Group Separation**

**Principal Component Analysis**
### Metabolomics: Overlap between analyses

- 35 compounds were significant (p<0.05) between all three analyses.

<table>
<thead>
<tr>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>Biochemical Name</th>
<th>Severe/Mild</th>
<th>Proband/Sib</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>Tryptophan Metabolism</td>
<td>tryptophan betaine</td>
<td>0.3</td>
<td>0.39</td>
<td>-0.4201</td>
</tr>
<tr>
<td></td>
<td></td>
<td>indolepropionate</td>
<td>0.75</td>
<td>0.57</td>
<td>-0.2353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kynurenine</td>
<td>0.82</td>
<td>0.84</td>
<td>-0.1764</td>
</tr>
<tr>
<td></td>
<td>Urea cycle; Arginine and Proline Metabolism</td>
<td>N-methyl proline</td>
<td>0.17</td>
<td>0.39</td>
<td>-0.4261</td>
</tr>
<tr>
<td></td>
<td></td>
<td>trans-4-hydroxyproline</td>
<td>0.72</td>
<td>0.85</td>
<td>-0.2018</td>
</tr>
<tr>
<td></td>
<td>Leu, Ile and Val Metabolism</td>
<td>alpha-hydroxyisovalerate</td>
<td>0.86</td>
<td>0.72</td>
<td>-0.2305</td>
</tr>
<tr>
<td></td>
<td>Met, Cys, SAM and Taurine Metabolism</td>
<td>alpha-ketobutyrate</td>
<td>1.3</td>
<td>1.59</td>
<td>0.1356</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-hydroxybutyrate (AHB)</td>
<td>1.23</td>
<td>1.35</td>
<td>0.1908</td>
</tr>
<tr>
<td></td>
<td>Phe and Tyr Metabolism</td>
<td>phenylacetylglutamine</td>
<td>1.25</td>
<td>1.35</td>
<td>0.1505</td>
</tr>
<tr>
<td></td>
<td>Gly, Ser and Thr Metabolism</td>
<td>N-acetylglutamine</td>
<td>1.32</td>
<td>1.55</td>
<td>0.1765</td>
</tr>
<tr>
<td></td>
<td>Ala and Asp Metabolism</td>
<td>aspartate</td>
<td>1.58</td>
<td>1.74</td>
<td>0.1529</td>
</tr>
<tr>
<td></td>
<td>Glutamate Metabolism</td>
<td>glutamate</td>
<td>1.36</td>
<td>1.48</td>
<td>0.1841</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>Fructose, Mannose and Galactose Metabolism</td>
<td>fructose</td>
<td>0.64</td>
<td>0.79</td>
<td>-0.1771</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mannose</td>
<td>1.21</td>
<td>1.32</td>
<td>0.1896</td>
</tr>
<tr>
<td></td>
<td>Disaccharides</td>
<td>sucrose</td>
<td>0.6</td>
<td>0.73</td>
<td>-0.1413</td>
</tr>
<tr>
<td>Cofactors/ Vitamins</td>
<td>Tocopherol Metabolism</td>
<td>gamma-tocopherol</td>
<td>0.67</td>
<td>0.72</td>
<td>-0.1808</td>
</tr>
<tr>
<td>Energy</td>
<td>Pantothenate/CoA Metabolism</td>
<td>pantothenate</td>
<td>1.51</td>
<td>1.43</td>
<td>0.2949</td>
</tr>
<tr>
<td>Lipid</td>
<td>TCA Cycle</td>
<td>chenodeoxycholate</td>
<td>0.87</td>
<td>0.88</td>
<td>-0.1542</td>
</tr>
<tr>
<td></td>
<td>Primary Bile Acid Metabolism</td>
<td>1-myristoylglycerophosphocholine (14:0)</td>
<td>0.72</td>
<td>0.77</td>
<td>-0.2378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-linoleoylglycerophosphocholine*</td>
<td>0.79</td>
<td>0.77</td>
<td>-0.1602</td>
</tr>
<tr>
<td></td>
<td>Lysolipid</td>
<td>3-carboxy-4-methyl-5-propyl-2-furanpropanoate</td>
<td>0.24</td>
<td>1.71</td>
<td>-0.1531</td>
</tr>
<tr>
<td></td>
<td>Polyunsaturated Fatty Acid (n3 and n6)</td>
<td>dihomo-linoleate (20:2n6)</td>
<td>1.62</td>
<td>1.39</td>
<td>0.1735</td>
</tr>
<tr>
<td></td>
<td>Ketone Bodies</td>
<td>3-hydroxybutyrate (BHBA)</td>
<td>5.24</td>
<td>0.9</td>
<td>0.2008</td>
</tr>
<tr>
<td></td>
<td>Steroid</td>
<td>andro steroid monosulfate 2*</td>
<td>2.18</td>
<td>2.24</td>
<td>0.2491</td>
</tr>
<tr>
<td></td>
<td>Food Component/Plant</td>
<td>stachydrine</td>
<td>0.16</td>
<td>0.67</td>
<td>-0.4095</td>
</tr>
<tr>
<td></td>
<td></td>
<td>piperine</td>
<td>0.38</td>
<td>0.51</td>
<td>-0.3737</td>
</tr>
<tr>
<td></td>
<td></td>
<td>homostachydrine*</td>
<td>0.67</td>
<td>0.91</td>
<td>-0.3199</td>
</tr>
<tr>
<td></td>
<td>Benzoate Metabolism</td>
<td>4-vinylphenol sulfate</td>
<td>1.1</td>
<td>0.63</td>
<td>-0.2451</td>
</tr>
<tr>
<td></td>
<td>Xanthine Metabolism</td>
<td>caffeine</td>
<td>0.29</td>
<td>0.24</td>
<td>-0.2473</td>
</tr>
<tr>
<td></td>
<td></td>
<td>paraxanthine</td>
<td>0.36</td>
<td>0.27</td>
<td>-0.2454</td>
</tr>
<tr>
<td></td>
<td></td>
<td>theobromine</td>
<td>0.39</td>
<td>0.29</td>
<td>-0.3145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7-methylxanthine</td>
<td>0.66</td>
<td>0.4</td>
<td>-0.1834</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3-methylxanthine</td>
<td>0.79</td>
<td>0.47</td>
<td>-0.1728</td>
</tr>
</tbody>
</table>
Tryptophan metabolism

**Equations**

- Kynurenine: $Y = -0.0023 - 0.0076 \times X$  
  - $N=226$

- Serotonin (5HT): $Y = -0.6290 + 0.0168 \times X$  
  - $N=22$

- Tryptophan-betaine: $Y = 0.9666 - 0.0853 \times X$  
  - $N=226$

- Indolepropionate: $Y = 0.2080 - 0.0247 \times X$  
  - $N=226$

**Data**

<table>
<thead>
<tr>
<th>Sub Pathway</th>
<th>Biochemical Name</th>
<th>Welch's Two-Sample t-Test</th>
<th>RM ANOVA</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan Metabolism</td>
<td>tryptophan</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>indoleactate</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>indoleacetate</td>
<td>2.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>indolepropionate</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-indoxyl sulfate</td>
<td>1.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kynurenine</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>serotonin (5HT)</td>
<td>1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tryptophan betaine</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C-glycosyltryptophan*</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lipid and energy metabolism

Disruption of beta-oxidation
Disruption of fatty acid synthesis
Production of ketones
A Metabolic Signature of Mitochondrial Dysfunction Revealed through a Monogenic Form of Leigh Syndrome

2-hydroxybutyrate (AHB)

\[ Y = -0.1536 + 0.0128 \times X \]

N=226

Clinical Severity Score

2-hydroxybutyrate (AHB)

Mild

Severe
Insulinotropic treatments exacerbate metabolic syndrome in mice lacking MeCP2 function

Meagan R. Pitcher¹,⁵, Christopher S. Ward²,⁵, E. Melissa Arvide²,³, Christopher A. Chapleau⁶, Lucas Pozzo-Miller⁶, Andreas Hoeflich⁷, Manaswini Sivaramakrishnan⁸, Stefanie Saenger⁸, Friedrich Metzger⁸ and Jeffrey L. Neul¹,²,³,⁴,⁵,*

1. Elevation of hepatic glutathione stress (due to oxidative stress)
2. Elevation of NADH/NAD⁺ due to increased lipid oxidation
Food and caffeine metabolism

Plant phenolic compounds such as chlorogenic acid

- quinate
- benzoate
- glycine
- hippurate
- catechol sulfate

compounds with gut microbiome metabolic origin or contribution

<table>
<thead>
<tr>
<th>Sub Pathway</th>
<th>Biochemical Name</th>
<th>Severe Mild</th>
<th>Proband Sib</th>
<th>Clinical Severity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Component/Plant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-piperidinone</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,6-anhydroglucose</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gluconate</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>erythritol</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>homostachydrine*</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>piperine</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>quinate</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stachydrine</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub Pathway</th>
<th>Biochemical Name</th>
<th>Severe Mild</th>
<th>Proband Sib</th>
<th>Clinical Severity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xanthine Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>caffeine</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>paraxanthine</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>theobromine</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-methylxanthine</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-methylxanthine</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Diet and route associate with clinical severity
Dietary metabolites
Beta-oxidation and oxidative stress

- **Acetyl-carnitine**
  - Scaled conc. vs. Severity
  - Data points for PO and GT, linear trend lines

- **Alpha-ketobutyrate**
  - Scaled conc. vs. Severity
  - Data points for PO and GT, linear trend lines

- **2-Hydroxybutyrate**
  - Scaled conc. vs. Severity
  - Data points for PO and GT, linear trend lines
Ketone bodies

3 subjects on ketogenic diet
Tryptophan metabolism

**Tryptophan-betaine**

**Serotonin**

**Kynurenine**

**Indolepropionate**
Next steps

• Formal analysis of relationship between metabolites and other factors
  • Drugs
  • Diet and route
  • Specific clinical features (seizures, etc)
  • Age
• Verification
  • Targeted analysis of identified metabolites
• Orthologous methods on same samples
• Targeted analysis
  • Isoprostanes
• Validation
  • New sample collection from current NHS
Moving forward

- Current iteration of Natural History Study
  - Systematic sample collection from all enrollees
  - Verification and additional analyses
  - Large enough sample for genotype correlation
  - Longitudinal assessment
- Comparison to other disease groups
  - MECP2 Duplication
  - CDKL5
  - FOXG1
- Assessment during clinical trials
- Utilization of animal models
Acknowledgements

Rett Natural History Study

Blue Bird Circle Rett Center
  Jeffrey Neul
  Daniel Glaze
  Kathleen Motil
  Judy Barrish

Blue Bird Circle Research Team
  Robert McNeil
  Aryn Knight
  Chris Ward

Greenwood Genetics
  Steve Skinner
  Lauren Baggett
  Fran Annese

Children’s Hospital Boston
  Walter Kaufman
  Katherine Barnes

University of Alabama-Birmingham
  Alan Percy
  Jane Lane

Funding
ORD, NICHD U54 HD061222
Genetics of people with RTT without *MECP2* mutations

- DNA from 22 people in Natural History Study without *MECP2* mutations
  - 11 classic RTT/11 atypical RTT
  - 13 complete trios, 9 incomplete

- Exome sequencing
  - *de novo* – not found in dbSNP138, 1000 Genomes, ESP6500, or ExAC
  - Heterozygous variants – MAF<0.005

- CNV assessment
  - Illumina Omni 2.5M SNP Array

*Sajan et al, Genetics in Medicine, in press*
Genetics of people with RTT without *MECP2* mutations

- 3 people actually had mutations in *MECP2* not previously discovered with clinical sequencing!
  - 2 in exon 1, which was not included in clinical sequencing until ~2003-4
- 17/19 remaining had at least one pathogenic mutations in other genes
- 13/17 had mutations in genes linked to other neurodevelopmental disorders
  - TCF4, 22q13 (SHANK3), IQSEC2, WDR45
- *de novo* mutation rate
  - 1.70 per trio
  - $2.57 \times 10^{-8}$ per base per generation (binomial $p=0.0009$)

*Sajan et al, Genetics in Medicine, in press*
Genetics of people with RTT without \textit{MECP2} mutations

Sajan et al, \textit{Genetics in Medicine}, in press

<table>
<thead>
<tr>
<th>Gene</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTL15B</td>
<td>Exome</td>
</tr>
<tr>
<td>ALTS2</td>
<td>CNV</td>
</tr>
<tr>
<td>BRAF</td>
<td>Exome</td>
</tr>
<tr>
<td>BRD1</td>
<td>CNV</td>
</tr>
<tr>
<td>CELSR1</td>
<td>CNV</td>
</tr>
<tr>
<td>CHD4</td>
<td>Exome</td>
</tr>
<tr>
<td>CHL1</td>
<td>CNV</td>
</tr>
<tr>
<td>CNTN6</td>
<td>CNV</td>
</tr>
<tr>
<td>FAM151A</td>
<td>Exome</td>
</tr>
<tr>
<td>HAF1</td>
<td>CNV</td>
</tr>
<tr>
<td>GABRB2</td>
<td>Exome</td>
</tr>
<tr>
<td>GRIN2A</td>
<td>Exome</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>Exome</td>
</tr>
<tr>
<td>HDAC1</td>
<td>Exome</td>
</tr>
<tr>
<td>HERC2</td>
<td>CNV</td>
</tr>
<tr>
<td>IMPDH1</td>
<td>Exome</td>
</tr>
<tr>
<td>IQGAP3</td>
<td>Exome</td>
</tr>
<tr>
<td>IQGAP2</td>
<td>Exome</td>
</tr>
<tr>
<td>KCI110</td>
<td>Exome</td>
</tr>
<tr>
<td>LAMB2</td>
<td>Exome</td>
</tr>
<tr>
<td>LRRC40</td>
<td>Exome</td>
</tr>
<tr>
<td>NAGA</td>
<td>CNV</td>
</tr>
<tr>
<td>NCO2</td>
<td>Exome</td>
</tr>
<tr>
<td>NDNL2</td>
<td>CNV</td>
</tr>
<tr>
<td>OTUD7A</td>
<td>CNV</td>
</tr>
<tr>
<td>PLXNB2</td>
<td>CNV</td>
</tr>
<tr>
<td>PPP6R2</td>
<td>CNV</td>
</tr>
<tr>
<td>SBF1</td>
<td>CNV</td>
</tr>
<tr>
<td>SGCG2</td>
<td>CNV</td>
</tr>
<tr>
<td>SLC19A1</td>
<td>CNV</td>
</tr>
<tr>
<td>SLC22A3</td>
<td>CNV</td>
</tr>
<tr>
<td>SLC39A13</td>
<td>Exome</td>
</tr>
<tr>
<td>SLC39A1</td>
<td>CNV</td>
</tr>
<tr>
<td>SMARCB1</td>
<td>CNV</td>
</tr>
<tr>
<td>SMARCE1</td>
<td>CNV</td>
</tr>
<tr>
<td>SMCC1A</td>
<td>Exome</td>
</tr>
<tr>
<td>SPECT1L</td>
<td>CNV</td>
</tr>
<tr>
<td>STXBP1</td>
<td>Exome</td>
</tr>
<tr>
<td>TCF30</td>
<td>CNV</td>
</tr>
<tr>
<td>TCF4</td>
<td>Exome</td>
</tr>
<tr>
<td>TRPM1</td>
<td>CNV</td>
</tr>
<tr>
<td>TRRAP</td>
<td>CNV</td>
</tr>
<tr>
<td>TUBGCP5</td>
<td>CNV</td>
</tr>
<tr>
<td>WDR45</td>
<td>Exome</td>
</tr>
<tr>
<td>WNT7B</td>
<td>CNV</td>
</tr>
<tr>
<td>ZNF536</td>
<td>Exome</td>
</tr>
</tbody>
</table>

\textbf{KEY}
- \textcolor{red}{\textbullet} Physical interaction
- \textcolor{blue}{\textbullet} Co-localization
- \textcolor{cyan}{\textbullet} Common pathway step
- Input gene
- Related gene
- Related gene with mutations in other NDDs and/or expression change in MECP2 mutant models

Sajan et al, \textit{Genetics in Medicine}, in press
MeCP2 gene structure

100’s disease causing mutations
8 point mutations account for 65% cases
Genotype/phenotype relationship

Neul et al, 2008
What biological factors exist to allow milder presentation? Could we find genetic modifiers?
Proposed sources of variation in RTT

- Non-random X chromosome inactivation
  - Rare familial cases inherited from unaffected mother
  - All cases show extreme (>99%) skewing

- *Brain-derived neurotropic factor (Bdnf)* polymorphism
  - Val66 – major population allele
    - normal function
  - Met66 – minor allele (rs6265, ExAC MAF=0.1937)
    - interferes with BDNF secretion
Sources of variation

The graph shows the severity (CSS) plotted against XCI (%) for VAL and MET BDNF polymorphism. The R² values are 0.04 and 0.12 for VAL and MET respectively.
Finding genetic modifiers in RTT

A suppressor screen in MeCP2 mutant mice implicates cholesterol metabolism in Rett syndrome

Christie M. Buchovecky¹, Stephen D. Turley², Hannah M. Brown¹, Stephanie M. Kyle¹, Jeffrey G. McDonald³, Benny Liu², Andrew A. Pieper⁴,⁵,⁸, Wenhui Huang⁶,⁸, David M. Katz⁷, David W. Russell³, Jay Shendure⁶ & Monica J. Justice¹

Indicated cholesterol important in RTT

Novel concept – creates new approach to therapy
Finding genetic modifiers in RTT

**Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis**

Mary J Emond, Tin Louie, Julia Emerson, Wei Zhao, Rasika A Mathias, Michael R Knowles, Fred A Wright, Mark J Rieder, Holly K Tabor, Deborah A Nickerson, Kathleen C Barnes, National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project, Lung GO, Ronald L Gibson & Michael J Bamshad

- “Extreme phenotypes” strategy
- Exome sequencing on 91 individuals
  - 43 with early age of onset of Pseudomonas infection
  - 48 with no Pseudomonas beyond median age of onset
- Compare two extremes for enrichment of variants
Genotype/phenotype relationship
Genotype/phenotype relationship

![Graph showing genotypes and clinical severities](image-url)
Biological sources of clinical variation
Overall design plan

**Current Proposal**

**Discovery**
- Identify Extreme Cases with Random XCI
- WES + Standard Pipeline
- Select Variants PC to control ancestry
- Variant(s) with large effect FDR<0.05
- Multiple variants FDR>0.05
- Candidate Pathways
  - Cholesterol metabolism
  - Chromatin modification
  - Oxidative stress
- Look for evidence to support putative candidate pathways

**Future Work**
- Restrict to putative pathogenic variants
- Look for candidate gene association FDR<0.05 for candidate loci

**Validation**
- Sequence putative loci in all RTT samples
- Look for association between variants and severity
- Isolate Additional Samples Lymphocytes/Skin Biopsies
- Look for evidence to support putative candidate pathways
- Isolate Additional Samples Lymphocytes/Skin Biopsies
- Look for evidence to support putative candidate genes
Extreme cases for sequencing

- Extreme cases defined by mutation specific distributions
- Top and bottom 20%
- Random to moderately skewed XCI (<90%)

<table>
<thead>
<tr>
<th>ROUND 1</th>
<th>CTT</th>
<th>early trunc</th>
<th>lrgdel</th>
<th>R106W</th>
<th>R133C</th>
<th>R168X</th>
<th>R255X</th>
<th>R270X</th>
<th>R294X</th>
<th>R306C</th>
<th>T158M</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>mild</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>severe</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>ROUND 2</td>
<td>CTT</td>
<td>early trunc</td>
<td>lrgdel</td>
<td>R106W</td>
<td>R133C</td>
<td>R168X</td>
<td>R255X</td>
<td>R270X</td>
<td>R294X</td>
<td>R306C</td>
<td>T158M</td>
<td>Total</td>
</tr>
<tr>
<td>mild</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>severe</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>ROUND 1/2</td>
<td>CTT</td>
<td>early trunc</td>
<td>lrgdel</td>
<td>R106W</td>
<td>R133C</td>
<td>R168X</td>
<td>R255X</td>
<td>R270X</td>
<td>R294X</td>
<td>R306C</td>
<td>T158M</td>
<td>Total</td>
</tr>
<tr>
<td>mild</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>severe</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>27</td>
</tr>
</tbody>
</table>
- Baylor standard exome capture (Nimblegen) – 42Mb
- Custom capture of genomic regions of interest – 1.1 Mb
Pipeline details

- Illumina HiSeq 2000 at Baylor Genome Center
- Alignment using Burrows-Wheeler Aligner to hg19 reference
- Variants identified using GATK
- Annotated using ANNOVAR
- Variants of interest need the following:
  - >10 high quality reads in all samples
  - Localized to coding, IVS, or UTR (except 5 RTT genes), excluding synonymous changes
  - MAF<0.15
Top level results Round 1

- Average of 8.1 Gb uniquely aligned sequence per individual
- Average coverage per base 110X
  - 94% bases covered >20X
  - 83% bases covered >40X
- Total identified variants: 59448
- Variants MAF<0.15: 38011 in 13669 genes
- Variants MAF<0.15 and predicted deleterious: 9531 in 5348 genes
- Gene level analysis – Fisher Exact Test
- Multiple testing correction: Benjamini-Hochberg
- 4 genes
- Voltage gaited chloride channel
- Down regulated in iPSC from RTT
- Homozygous loss of function causes Bartter Syndrome type III
  - Defect in ability of kidney to absorb Na
  - Ca wasting in urine
- rs12746751 causes Pro – Leu change

<table>
<thead>
<tr>
<th>SNP</th>
<th>Mutation_type</th>
<th>Deleterious</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>S12</th>
<th>Ratio</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12140223</td>
<td>nonsynonymous</td>
<td>V/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>V/R</td>
<td>6/13</td>
</tr>
<tr>
<td>rs35747151</td>
<td>nonsynonymous</td>
<td>V/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>6/13</td>
</tr>
<tr>
<td>rs34471231</td>
<td>nonsynonymous</td>
<td>V/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>6/13</td>
</tr>
<tr>
<td>rs12746751</td>
<td>nonsynonymous</td>
<td>V/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>6/13</td>
</tr>
<tr>
<td>rs68127248</td>
<td>UTR3</td>
<td>V/R</td>
<td>V/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>R/R</td>
<td>R/R</td>
<td>V/R</td>
<td>6/13</td>
</tr>
</tbody>
</table>
Candidates/Pathways

**Current Proposal**

**Discovery**
- Identify Extreme Cases with Random XCI
- WES + Standard Pipeline
  - Select Variants PC to control ancestry
  - Variant(s) with large effect FDR<0.05
  - Multiple variants FDR>0.05
- Candidate Pathways
  - Cholesterol metabolism
  - Chromatin modification
  - Oxidative stress

**Future Work**

**Validation**
- Sequence putative loci in all RTT samples
- Look for association between variants and severity
- Isolate Additional Samples Lymphocytes/Skin Biopsies
  - Look for evidence to support putative candidate pathways
  - Look for evidence to support putative candidate genes
Cholesterol synthesis

Cholesterol synthesis gene variants

Graphical representation of the cholesterol synthesis pathway with gene variants highlighted:

- Sgke
  - rs118130263
  - snp138
  - Mutatio
  - nonsyn
  - p.Asn3

- Hmgcr
  - rs5909
  - snp138
  - Mutatio
  - nonsyn
  - UTR3

- Idi2
  - rs41314629
  - rs1044261
  - snp138
  - Mutation
  - stopgain
  - stoploss

- 3-hydroxy-3-methylglutaryl-CoA Reductase
  - HMG-CoA Synthase

- Mevalonate
  - Mevalonate Kinase
  - Phosphomevalonate Kinase

- Dimethylallyl-PP
  - Isopterinyl-PP

- Geranyl-PP
  - Farnesyl-PP Synthase

- Squalene Synthase

- Squalene Monoxygenase

- Protein Prenylation
Modeling cholesterol metabolism by gene expression profiling in the hippocampus†

Christopher M. Valdeza, Clyde F. Phelixa, Mark A. Smithc, George Perryab, and Fidel Santamariaab

Original Article

Isopentenyl diphosphate isomerase, a cholesterol synthesizing enzyme, is localized in Lewy bodies

Segmental copy-number gain within the region of isopentenyl diphosphate isomerase genes in sporadic amyotrophic lateral sclerosis

Takeo Katoa,*,1, Mitsuru Eminiab,1, Hidenori Satoa,1, Shigeki Arawakaa, Manabu Wadaa, Toru Kawanamia, Tadashi Katagiric, Kenji Tsuburayad, Itaru Toyoshimae, Fumiaki Tanakaf, Gen Sobuef, Kenichi Matsubaram
Future genetic modifiers in RTT

• Get 2\textsuperscript{nd} round sequencing results
• Identify additional extreme cases
  • Natural History Study
    • new version has systematic bio-sampling
• Candidate analysis
  • Mouse screens
  • Predicted pathways
• Multimodal analysis – metabolomics/transcriptomics/proteomics
• Validation
  • Variant effects on non-extreme cases of RTT
  • Models-animal and cellular