EXOME SEQUENCING FOR NOVEL DISEASE GENE DISCOVERY IN FAMILIES WITH RARE MENDELIAN PHENOTYPES

F. Sessions Cole, M.D.
Park J. White, M.D., Professor of Pediatrics
Washington University School of Medicine
St. Louis Children’s Hospital

Genetics and Genomics of Disease
26 October 2016
Disclosures

F. Sessions Cole, M.D., has no financial or proprietary conflicts of interest to disclose.

Extramural support

R01 HL065174 (FSC)  R01 HL094520 (MP)
R01 HL082747 (FSC)  R01 HL102968 (MP)
The Saigh Foundation (FSC)  R01 CA08759 (SK)
K08 CA140720 (TD)  R21 DC011332 (SK)
Children’s Discovery Institute (TD, FSC, JAW)
Acknowledgements

Todd Druley, Ph.D., M.D.  Stuart Kornfeld, M.D.  Michael Parmacek, M.D.

Daniel J. Wegner, M.S.  Helena van Meel, Ph.D.  Jian Li, M.D.
Different Strategies to Discover Genetic Causes of Pediatric Diseases

1) Pediatric cancer: sequence tumor and host DNA; computationally subtract; discover potentially treatable target

2) Structural birth defects, neurometabolic disorders, and other known pediatric problems for which candidate genes known: target candidate genes known to be associated with the phenotype

3) Structural birth defects, neurometabolic disorders, and other newborn infant problems without prior disease description: unbiased discovery using exome or whole genome sequencing
   a) Possible role for integration of transcriptome signature with DNA sequencing to improve diagnostic success
Impact of Availability of High-throughput DNA Sequencing Technologies (Nature 2016;536:285)

1) Sequencing of whole genomes or exomes of hundreds of thousands of humans

2) Accessibility reduced by practical, logistical, and ethical reasons (e.g., heterogeneity of variant calling pipelines)

3) Databases of genetic variation important for clinical interpretation of variants observed in patients with rare Mendelian diseases (e.g., variant filtering by minor allele frequency in unselected individuals)

4) Recently, sequence data from the Exome Aggregation Consortium (ExAC) reported 60,706 human exomes
   a) 7,404,909 high quality variants (317,381 insertions or deletions): 1 variant/8 base pairs within exome intervals; 99% of variants have frequency <1%, 54% seen only once in the data set, and 72% absent from the Exome Sequencing Project or 1000 Genomes
Impact of Availability of High-throughput DNA Sequencing Technologies

5) Abundance of rare, functional variants in many disease genes in ExAC suggests that such variants should not be assumed to be causal or highly penetrant without careful segregation or case-control analysis.

6) Discovery of 3,230 high loss of function intolerant genes (protein truncating variants), 72% of which have no established human disease phenotype in OMIM or ClinVar databases of observed human genetic mutations.

7) Note that structural and copy number variants not reported in ExAC.
Exome Sequencing and Variant Discovery Strategies

• Screen by karyotyping and Affymetrix 6.0 genome wide array
• Assume inheritance for computational strategy (dominant, recessive, de novo)
• Sequence using Agilent SureSelect Human All Exon Array (38 Mb) and Illumina HiSeq2000
• Align sequence using Novoalign (human genome build hg19)
• Filter all single nucleotide variants (SNVs) and insertions/deletions (in/dels) based on quality score and coverage (≥5X)
• Computationally identify genetic variants (SNVs and small in/dels) using Samtools
Exome Sequencing and Variant Discovery Strategies

- Computationally annotate genetic variants using Annovar (location in genome and exonic function), Combined Annotation-Dependent Depletion (CADD), public database frequency (Exome Sequencing Project and 1000 Genomes), and functional prediction of non-synonymous variants
- Assess sequencing sensitivity by comparing genome array and exome sequencing at ~2,500 variant positions (98.4%)
- Look for biologic plausibility (model system) and/or unrelated individuals with similar phenotype
Filtering for Rare, Functionally Deleterious, Recessive Variants

All variants

Non-synonymous variants
- <5% frequency in public databases
  - SIFT
  - Polyphen
  - PhyloP
  - LRT
  - Mutation Taster
  - CADD

Frameshift insertion and deletions
- <5% frequency in public databases
  - Genes with multiple variants
    - Parental transmission
      (1 variant from each parent)
  - Final gene list
Rare Mendelian Phenotype

Multi-organ failure
Healthy couple’s first baby developed multi-organ failure in first 9 months of life:

1. Liver failure
2. Transfusion-dependent red cell aplasia
3. Renal tubular dysfunction
4. Interstitial lung disease
5. Hypotonia
6. Central hypothyroidism
Variant Discovery Strategy: Multi-Organ Failure

17,998 total variants

- 7,468 non-synonymous
  - 1,063 variants <2% frequency
    - 305 variants, functional prediction

- 485 insertions/deletions
  - 98 variants <2% frequency
    - 21 genes with multiple variants (45 variants)

- Parental transmission
  - 1 gene, 2 variants

Methionyl-aminocyl tRNA synthetase: compound heterozygous missense mutations (F370L – paternal and I523T – maternal)
MARS Function

- Aminoacyl-tRNA synthetases mediate a critical step in protein biosynthesis
- Couple tRNA to an amino acid
- MARS couples methionine to its cognate tRNA to generate methionyl-tRNA
- Secondary function:
  - Promote rRNA biogenesis
Biologic Validation of MARS Mutations

Expression vectors designed to over express both mutations and wild-type MARS in HEK293 cells

van Meel et al., BMC Med Genet 2013;14:106
Biologic Validation of MARS Mutations: Functional Assay

\[
\begin{align*}
\text{ATP} & \quad \text{AMP + PP}_i \\
\text{[}^3\text{H}]\text{methionine + tRNA} & \quad \rightarrow \quad \text{[}^3\text{H}]\text{methionine-tRNA} \\
\text{MARS} & \\
\text{HEK293 cell lysate overexpressing WT, F370L, or I523T MARS}
\end{align*}
\]

van Meel et al., BMC Med Genet 2013;14:106
Aminoacylation Results

<table>
<thead>
<tr>
<th></th>
<th>Activity (% of wild-type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mock</td>
<td></td>
</tr>
<tr>
<td>WT MARS</td>
<td>18+5.9</td>
</tr>
<tr>
<td>F370L</td>
<td></td>
</tr>
<tr>
<td>I523T</td>
<td>16+6.2</td>
</tr>
</tbody>
</table>

n = 3
Rare Mendelian Phenotype

Extreme microcephaly
Healthy couple has 3 independent conceptions with the same constellation of severe congenital birth defects:

1. Growth retardation (36-37 weeks, birth weights 1,400-1,800 gms)

2. Heart defects (ventricular septal defects, non compaction left and right ventricles, long QT syndrome)

3. Severe microcephaly (unfixed brain weights 14-23.5 gms, normal 318±106 gms); failure of neuronal precursor proliferation and early, severe disruption of cortical development

4. Lung defects (symmetrical, bilobed lungs)

5. Fatal
Rare Mendelian Phenotypes

Extreme microcephaly

Affected infant

Normal infant
Variant Discovery Strategy: Microcephaly

1. Identified 6 candidate genes (non-synonymous, homozygous or compound heterozygous, functional genotype in all 3 daughters)

2. Compound heterozygous variants all originated with father (2 variants on paternal allele – normal brain MRI); mother’s genotypes wild type at all positions and did not transfer any variants to daughters except Mkl2

3. Only the Mkl2 (megakaryoblastic leukemia)/MRTF-B (myocardin related transcription factor-B) variant (rs75963814; CAC(H) -> CAA(Q); His288Gln) was rare (MAF 0.6% in 1000 Genomes, 1.1% in Exome Sequencing Project)
   
   A. 1 homozygote (~40 year old female enrolled in a cardiac study)
   
   B. Affymetrix 6.0 Array: 185 kb deletion 1.2 Mb upstream of Mkl2/MRTF-B mutation on father’s variant allele (16p13.12)
   
   C. No mutations in genes known to cause brain malformations, WDR62 or NDE1 (adjacent to Mkl2 locus on chromosome 16)
Biologic Function:  *Mkl2/MRTF-B*

**Mkl2/MRTF-B**: Regulates actin and cytoskeleton gene transcription; *Mkl2* and *Mkl1* are transcriptional co-activators with serum response factor (SRF)

**SRF**: Binds to CArG box, a 10 bp sequence (CC(A/T)$_6$GG) generally found within 4,000 bp of a targeted gene’s transcription start site that controls disparate programs of gene expression including immediate early genes and actin cytoskeleton genes

**MKL2/MRTF-B-SRF binding**: Mediated through Basic and Glutamine-rich domains: His288Gln falls at the first amino acid of the BASIC domain

No prior reports of humans with *Mkl2/MRTF-B*-associated diseases
Novel Paternal Upstream Deletion

• 185 kb deletion 1.2 Mb upstream of *Mkl2* in cis with the paternal *Mkl2* variant

• Confirmed heterozygous deletion in all three probands

• Deletion contains:
  
  − 4 CArG (CC(AT)$_6$GG) boxes (2 in regions of high regulatory potential)
  
  − 1 entire gene (*CPPED1*) (no reported association with brain development)

  − MIR4718 (156 known targets including 2, microcephalin 1 (*MCPH1*) and strawberry notch homolog 1 (*SBNO1*) with known connections to human microcephaly

• Adult female with same homozygous *Mkl2* variants does not carry this deletion
Proposed Genetic Scheme for Variant Haploinsufficiency

Father
- **MKL2**
  - NORMAL GENE COPY
  - NON-FUNCTIONAL GENE COPY

Mother
- **MKL2**
  - NORMAL GENE COPY
  - WEAK GENE COPY

Daughters
- From Mom
  - **MKL2**
    - WEAK GENE COPY
- From Dad
  - **MKL2**
    - NON-FUNCTIONAL GENE COPY

Heart Attack Woman
- **MKL2**
  - WEAK GENE COPY
  - WEAK GENE COPY

**Legend**
- ⭐ Mutation
- DNA deletion
Similar *Mkl2/MRTF-B* Expression in Control and Affected Infant Brain

Nuclear MKL2 immunostaining observed in a subpopulation of neurons of control and patient brains (not all cells in either sample). Mutant protein is expressed and relatively stable. SRF immunostaining also reveals similar patterns of expression in control and patient brains.
Dysregulation of MKL:SRF Axis Could Impair Human Neurodevelopment

- Lack of Cdk5 phosphorylation: severe cortical dysmorphology
- Rodent Model: Brain-specific Mkl1−/−Mkl2−/− mice: 25% reduction SRF, Pctaire1

\[\text{Cytoplasmic F-actin polymerization (remodeling of dendrites, synapses)}\]

\[\downarrow\]

\[\text{Dissociation of G-actin from MKL1/MKL2}\]

\[\downarrow\]

\[\text{MKL1/MKL2 relocation to nucleus}\]

\[\downarrow\]

\[\text{Heterodimerization of MKL1/MKL2 with SRF}\]

\[\uparrow\text{Pctaire1, protein kinase and brain-specific target of MKL:SRF}\]

\[\downarrow\]

\[\text{Phosphorylation of Cdk5, serine-threonine kinase}\]
Lack of Pctaire1 expression in affected daughters suggests dysregulation of MKL-SRF axis
Sequencing of *Mkl2* and *Srf* in Other Fetuses or Infants with Microcephaly

51 cases of extreme microcephaly from Washington University Fetal Brain Repository and 3 global locations (Bogazici University in Istanbul, University of Washington, Albert Ludwigs University in Germany)

- No individuals with exon 10 *Mkl2* variants
- Two individuals with non-coding, novel *SRF* variants with moderate to high predicted regulatory potential
- Three individuals with multiple *Mkl2* variants of variable frequencies (MAF ≥2.3%) and predicted regulatory impact
- One individual homozygous and 1 heterozygous for rare *Mkl2* variant that abolishes methylated cytosine at highly conserved base
Summary

1. Exome sequencing and computational discovery of novel candidate genes (*Mkl2* and *MARS*) lead to understanding of previously unrecognized human diseases: gene first diagnostic approach

2. Current challenges: imperfect functional prediction algorithms, uneven sequencing coverage throughout human genome, lack of uniform phenotyping strategies, impact of variants in intronic and intergenic regions

3. Future challenges: integration of developmental functions of individual genes, understanding gene x environment impact on phenotype, developing cell-, organ-, and developmental stage-specific catalogue of gene pathways, discovery of oligogenic vs. monogenic causes of disease
## Acknowledgements

<table>
<thead>
<tr>
<th>Washington University</th>
<th>University of Pennsylvania</th>
<th>University of Washington</th>
<th>McDonnell Genome Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. Druley</td>
<td>M. Parmacek</td>
<td>W. Dobyns</td>
<td>R. Wilson</td>
</tr>
<tr>
<td>D.J. Wegner</td>
<td>J. Li</td>
<td>M. Bamshad</td>
<td>E. Mardis</td>
</tr>
<tr>
<td>R. Mitra</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. Cliften</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Corbo</td>
<td>Extramural Support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J.A. Wambach</td>
<td>FSC: R01 HL065174, R01 HL082747, The Saigh Foundation, Children's Discovery Institute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Kornfeld</td>
<td>TD: K08 CA150720, Children's Discovery Institute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Van Meel</td>
<td>JAW: K08 HL105891, Children's Discovery Institute, American Thoracic Society</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MP: R01 HL094520, R01 102968</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SK: R01 CA08759, R21 DC011332</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Affected children and families**
**Neural and Cardiac Phenotypes: Mkl2/MRTF-B**

<table>
<thead>
<tr>
<th>Study Title</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardin-related transcription factors regulate the Cdk5/Pctaire1 kinase cascade to control neurite outgrowth, neuronal migration and brain development</td>
<td>Conditional knockout of Mkl1 and Mkl2 was <em>embryonic lethal due to severe brain malformation</em></td>
</tr>
<tr>
<td>Mayssa H. Mokalled, Aaron Johnson, Yuri Kim, Jiyeon Oh and Eric N. Olson</td>
<td></td>
</tr>
<tr>
<td>Development. 2010 Jul 15; 137(14): 2365-2374</td>
<td></td>
</tr>
<tr>
<td>Mkl Transcription Cofactors Regulate Structural Plasticity in Hippocampal Neurons</td>
<td>Decreased Mkl2 significantly <em>inhibits neuritic growth</em></td>
</tr>
<tr>
<td>Niamh C. O'Sullivan, Mark Pickering, Danika Di Giacomo, Jennifer S. Loscher and Keith J. Murphy</td>
<td></td>
</tr>
<tr>
<td>Cerebral Cortex. August 2010; 20:1915-1925</td>
<td></td>
</tr>
<tr>
<td>Myocardin-related transcription factor B is required in cardiac neural crest for smooth muscle differentiation and cardiovascular development</td>
<td>Mrtf-B null mouse embryos die in the perinatal period with <em>cardiac malformations</em></td>
</tr>
<tr>
<td>Jian Li, Xiaohong Zhu, Mary Chen, Lan Cheng, Deying Zhou, Min Min Lu, Kevin Du, Jonathan A. Epstein and Michael S. Parmacek</td>
<td></td>
</tr>
</tbody>
</table>
Cortical Transcript Expression in Probands, Other Microcephalic Fetuses, and Fetuses without Microcephaly
Relative Cortical Transcript Expression in Probands, Other Microcephalic Fetuses, and Fetuses without Microcephaly

Controls (N=6)
Affected daughters (N=3)
Other infants with diagnosis of microcephaly (N=33)
### Table 1 | Mean number of coding variants in two populations

<table>
<thead>
<tr>
<th>Variant type</th>
<th>Mean number of variants (± sd) in African Americans</th>
<th>Mean number of variants (± sd) in European Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novel variants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missense</td>
<td>303 (± 32)</td>
<td>192 (± 21)</td>
</tr>
<tr>
<td>Nonsense</td>
<td>5 (± 2)</td>
<td>5 (± 2)</td>
</tr>
<tr>
<td>Synonymous</td>
<td>209 (± 26)</td>
<td>109 (± 16)</td>
</tr>
<tr>
<td>Splice</td>
<td>2 (± 1)</td>
<td>2 (± 1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>520 (± 53)</td>
<td>307 (± 33)</td>
</tr>
<tr>
<td><strong>Non-novel variants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missense</td>
<td>10,828 (± 342)</td>
<td>9,319 (± 233)</td>
</tr>
<tr>
<td>Nonsense</td>
<td>98 (± 8)</td>
<td>89 (± 6)</td>
</tr>
<tr>
<td>Synonymous</td>
<td>12,567 (± 416)</td>
<td>10,536 (± 280)</td>
</tr>
<tr>
<td>Splice</td>
<td>36 (± 4)</td>
<td>32 (± 3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>23,529 (± 751)</td>
<td>19,976 (± 505)</td>
</tr>
<tr>
<td><strong>Total variants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missense</td>
<td>11,131 (± 364)</td>
<td>9,511 (± 244)</td>
</tr>
<tr>
<td>Nonsense</td>
<td>103 (± 8)</td>
<td>93 (± 6)</td>
</tr>
<tr>
<td>Synonymous</td>
<td>12,776 (± 434)</td>
<td>10,645 (± 286)</td>
</tr>
<tr>
<td>Splice</td>
<td>38 (± 5)</td>
<td>34 (± 4)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>24,049 (± 791)</td>
<td>20,283 (± 523)</td>
</tr>
</tbody>
</table>

Bamshad et al.  
Nat Rev Genet  
2011
Cortical Transcript Expression in Probands, Unrelated Microcephalic Fetuses, and Fetuses without Microcephaly

Transcript expression of 27 genes known to be involved in the SRF:Mkl2 axis, previously identified in cases of primary microcephaly, and CPPED1 (located in novel, paternal deleted region)

- In probands, decreased expression of CPPED1 and genes known to be involved in microcephaly or lissencephaly: ASPM, MCPH1, MKL1, TUBA1A, WDR62, CENPJ (CPAP), DCX, and CASK

- In unrelated microcephalic fetuses (N=33), increased expression of Mkl2, Srf, NDE1, and CPPED1 expression; decreased TUBA1A, CENPJ, DCX, and CASK expression
3) Only the $\textit{Mkl2}$ (\textit{megakaryoblastic leukemia})/$\textit{MRTF-B}$ (\textit{myocardin related transcription factor-B}) variant (rs75963814; $\text{CAC(H)} \rightarrow \text{CAA(Q); His288Gln}$) was rare (MAF 0.6% in 1000 Genomes, 1.1% in Exome Sequencing Project)

A. 1 homozygote (~40 year old female enrolled in a cardiac study)

B. Affymetrix 6.0 Array: 185 kb deletion 1.2 Mb upstream of $\textit{Mkl2}/\textit{MRTF-B}$ mutation on father’s variant allele (16p13.12)

C. No mutations in genes known to cause brain malformations, $\textit{WDR62}$ or $\textit{NDE1}$ (adjacent to $\textit{Mkl2}$ locus on chromosome 16)
**MARS Mutation and Disease Mechanism**

- **GST_C-like domain**: protein-protein interaction
- **Catalytic domain**: Rossmann fold
- **Conserved functional motifs**: HIGH and KFSKS
- **RNA binding domain**: tRNA binding and protein-protein interactions
- **Nuclear localization signals**: promotes rRNA biogenesis
Human and Murine *Mkl2/MRTF-B* Cardiac Phenotype

- **IHC:**
  - Wild type
  - Affected daughter
  - Green: Mkl2, Blue: DAPI

- **H&E:**
  - Wild type
  - Affected daughter

400x

Mouse Gene trap (hypomorphic) E18.5 embryos

- **Wild type**
- **Hypomorph**

Dilated vessels
Exome Sequencing To Discover Candidate Genes for Rare, Mendelian Phenotypes

Successful strategy for identification of >100 rare, monogenic diseases

Underlying causative gene for >50% of rare, monogenic diseases (>2,500) still unknown

Determinants of successful disease discovery

• Mode of inheritance (recessive, dominant, de novo)
• Specificity of disease phenotype
• Sequencing coverage of causative gene
• Computational variant calling and filtering
• Large structural variant or insertion/deletion
• Replication of gene locus in patients with similar phenotype
• Functional causality established