Providing a More Comprehensive and Personalized Approach to Genetic Disorders through Next Generation Sequencing

Karl V. Voelkerding, MD
Professor of Pathology
University of Utah
Medical Director for Genomics and Bioinformatics
ARUP Laboratories

ARUP Institute for Learning Webinar
June 18, 2013
voelkek@aruplab.com

Notice of Faculty Disclosure

The individual below has no relevant financial relationships with commercial interests to disclose

Karl V. Voelkerding, M.D.

Learning Objectives

• Describe how NGS has provided a new technological approach that has expanded the ability to improve the diagnosis of genetic disorders.

• Relate the essential and complex role of bioinformatics in deriving diagnostic results from NGS data.

• Discuss the impact of exome sequencing in the diagnostic evaluation of patients with undiagnosed disorders.
**Genome sequencing in microfabricated high-density picolitre reactors**

- **NGS Process Steps**
  - Genomic DNA
  - Fragmentation (150–400 bp)
  - Repair Ends and Ligate Oligonucleotide Adapters
  - "Randomly Overlapping Fragment Library"

- **NGS Process Steps**
  - "Fragment Library"
  - Clonal Amplification of Each Fragment

- **NGS Process Steps**
  - Sequencing of Clonal Amplicons in a Flow Cell
Next Generation Sequencing
Flow Cell – High Throughout Process
Sequential Introduction of Nucleotides to Build Sequence

- Luminescence (Roche)
- Fluorescence (Illumina)
- H+ Ion Detection (Ion Torrent)

Signal to Noise Processing
↓
Cyclic Base Calls
C G A T G C

Base Quality Scores
C30 G28 A33 T30 G28 → FASTQ File

Qualitative and Quantitative Information

Qualitative and Quantitative Information

Cystic Fibrosis

NGS Platform Summary

<table>
<thead>
<tr>
<th>Platform</th>
<th>Methodology</th>
<th>Ref Seq</th>
<th>Coverage</th>
<th>Qualitative and Quantitative Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina</td>
<td>cPCR</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → Base Quality Scores</td>
</tr>
<tr>
<td>NGS</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
<tr>
<td>Illumina</td>
<td>PCR</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
<tr>
<td>NGS</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
</tbody>
</table>

FDA Submission
Cystic Fibrosis

Table of NGS Platform Summary

<table>
<thead>
<tr>
<th>Platform</th>
<th>Methodology</th>
<th>Ref Seq</th>
<th>Coverage</th>
<th>Qualitative and Quantitative Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina</td>
<td>cPCR</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → Base Quality Scores</td>
</tr>
<tr>
<td>NGS</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
<tr>
<td>Illumina</td>
<td>PCR</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
<tr>
<td>NGS</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes</td>
<td>GCATGC - - - → BASE QUALITY SCORES</td>
</tr>
</tbody>
</table>
Genetic Testing Paradigm Shift

Sanger Sequencing
Qualitative

Next Generation Sequencing
Qualitative and Quantitative
High Throughput

New Landscape of Genetic Testing

Single-Gene Diagnostics

Multi-Gene Diagnostics

Exome

Whole Genome

Increasing Complexity

Multi-Gene Panel Diagnostics

Clinical Phenotype

Multiple Genes Responsible

Locus Heterogeneity

Allelic Heterogeneity

Multiple Mutations Possible

Technically Difficult to Test for by Sanger Sequencing
Multi-Gene Panel Diagnostics

- Cardiomyopathies
  - Hypertrophic
  - Dilated
  - Arrhythmias
  - 10-35+ Genes Each

- Mitochondrial Disorders
  - Mitochondrial Genome
  - Nuclear Genes > 100 Genes

- Primary Immune Deficiencies
  - 40+ Genes

Next Generation Sequencing Technology Makes Multi-Gene Panel Diagnostics Feasible

Hypertrophic Cardiomyopathy – Model for Multi-Gene Diagnostics

Prevalence = ~ 1 in 500 – 1,000

Teenage to Adult Onset
Autosomal Dominant
Arrhythmias/Angina
Sudden Death

Normal
Hypertrophic

HCM – Genetic Disorder of Cardiac Sarcomere

Myofibril
Sarcomere
### Hypertrophic Cardiomyopathy Genes

<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene</th>
<th>Mutations</th>
<th>Gene Size bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myosin, heavy chain 7</td>
<td>(MYH7)</td>
<td>193</td>
<td>33,828</td>
</tr>
<tr>
<td>Myosin binding protein C</td>
<td>(MYBPC3)</td>
<td>138</td>
<td>28,280</td>
</tr>
<tr>
<td>Troponin T type 2</td>
<td>(TNNT2)</td>
<td>33</td>
<td>25,673</td>
</tr>
<tr>
<td>Troponin I type 3</td>
<td>(TNNT3)</td>
<td>75</td>
<td>12,963</td>
</tr>
<tr>
<td>Cysteine and glycine-rich protein 3</td>
<td>(CSRP3)</td>
<td>12</td>
<td>27,024</td>
</tr>
<tr>
<td>Tropomyosin 1, α</td>
<td>(TPM1)</td>
<td>11</td>
<td>36,274</td>
</tr>
<tr>
<td>Myosin, light chain 2</td>
<td>(MYL2)</td>
<td>10</td>
<td>16,758</td>
</tr>
<tr>
<td>Actin</td>
<td>(ACTC)</td>
<td>7</td>
<td>14,631</td>
</tr>
<tr>
<td>Myosin, light chain 3</td>
<td>(MYL3)</td>
<td>5</td>
<td>12,617</td>
</tr>
<tr>
<td>Protein kinase, AMP-activated, y2</td>
<td>(PRKAG2)</td>
<td>6</td>
<td>328,114</td>
</tr>
<tr>
<td>Phospholamban</td>
<td>(PLN)</td>
<td>2</td>
<td>19,112</td>
</tr>
<tr>
<td>Troponin C type 1</td>
<td>(TNNC1)</td>
<td>1</td>
<td>9,041</td>
</tr>
<tr>
<td>Titin</td>
<td>(TTN)</td>
<td>2</td>
<td>281,434</td>
</tr>
<tr>
<td>Myosin, heavy chain 6</td>
<td>(MYH6)</td>
<td>2</td>
<td>32,628</td>
</tr>
<tr>
<td>Titin-cap</td>
<td>(TCAP)</td>
<td>2</td>
<td>9,361</td>
</tr>
<tr>
<td>Caveolin 3</td>
<td>(CAV3)</td>
<td>1</td>
<td>20,199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>405</td>
</tr>
</tbody>
</table>
Hypertrophic Cardiomyopathy – Model for Multi-Gene Diagnostics

Value of Genetic Testing
- Confirm Genetic Etiology
- Specific Mutation Identification
- Family Risk Counseling/Testing

Medical Management
- Beta and Calcium Channel Blockers
- Antiarrhythmics – Cardioversion – Implantable Defibrillators
- Transplantation

Multi-Gene Panel Diagnostics

More Comprehensive
Compared to Single Gene Sanger Sequencing

Gene Content = Based on Current Knowledge

Facilitated by New Platforms
Lower Capital Costs
Faster Sequencing Process

Multi-Gene Diagnostics Require Gene Enrichment

Genomic DNA

Amplification Based
- PCR or LR-PCR
- RainDance ePCR
- Fluidigm
- Ion Torrent and Illumina
- Agilent Haloplex

Array Capture Based
- In Solution Target Probes
- Agilent
- Nimblegen
- Illumina

Enriched Genes → NGS
Multi-Gene Diagnostics Require Gene Enrichment

- Amplification Based
  - PCR or LR-PCR
  - RainDance ePCR
  - Fluidigm
  - Ion Torrent and Illumina
- Array Capture Based
  - In Solution Target Probes
  - Agilent
  - Nimblegen
  - Illumina

Enrichment Method - Difficult Choice - Substantial Cost Investment

Considerations in Designing Multi-Gene Panels

Suitability of Enrichment Method for Laboratory

- Is the Technical Workflow (Manual) Adoptable in Your Setting?
  - Is it possible to automate the workflow?
- Is the enrichment method compatible with your sequencing platform?
  - How many samples can be barcoded and pooled for sequencing?
  - What data analysis pipeline will be required?
  - Vendor supplied or in-house custom developed

Perform In Silico Designs with Enrichment Methods

- Free designs using vendor software
  - Valuable to compare design results between method options
- What percentage of gene targets will be enriched?
  - Are there In Silico predicted problem areas?
Considerations in Designing Multi-Gene Panels

Expect *In Silico* versus Empiric Results Differences

- Characterize Problem Areas
  - Inadequate Sequence Coverage of Some Target Regions
  - Regions where Data Analysis indicates Homologous Sequence Interference

Case Example Multi-Gene Panel Design

Project Goal

- Multi-Gene Panel for Primary Immune Deficiencies
- Sequencing Platform – Illumina MiSeq
- *In Silico* Designs Performed and Agilent Haloplex Chosen
  - In House Custom Data Analysis

Haloplex Enrichment Theory and Practice

1. Digest and Denature Genomic DNA

2. Hybridize Biotin Target Probe Library to Form "Tri-Molecular" Circular Complexes

3. Capture and Ligate to Form Closed Target Circles
4. **PCR Amplify Targets and Incorporate Sequencing Adapters and Indexes**

![Diagram of PCR Amplification]

**Haloplex Enrichment Theory and Practice**

**Target (s)**

- Adapter
- Indexes
- NGS

**Target Genes**

- IFNGR1
- IFNGR2
- STAT1
- IL12B
- IL12RB1
- IKBKG
- TYK2
- CYBB
- IRF8
- ISG15

**Coverage (%)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exome Halo 150</th>
<th>Halo 250</th>
<th>Predicted Target Coverage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNGR1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IFNGR2</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>STAT1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IL12B</td>
<td>37</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IL12RB1</td>
<td>99.9</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IKBKG</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>TYK2</td>
<td>37</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>CYBB</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRF8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISG15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Addressing “Gaps” in Multi-Gene Panels**

1. Genomic DNA
2. Enrichment
3. Target Genes
4. NGS Library Prep
5. Next Generation Sequencing
6. Bioinformatics
7. Interpretation
8. PCR
10. Big Dye Terminators
11. Sanger Sequencing
12. Bioinformatics
13. Interpretation
Becoming a “New First Tier” Approach

- Application to a Growing Number of Inherited Disorders

Implementation Challenges for Laboratories

- Choosing a Technical Approach
- Assay Optimization and Data Analysis
- Scaling Gene Numbers Increases Interpretive Review Time

New Landscape of Genetic Testing

Single-Gene Diagnostics

Exome

Multi-Gene Diagnostics

Whole Genome

Increasing Complexity

Human Exome

~ 1.5% of the genome

~ 20,500 genes

“Repository” of Mendelian Mutations

“Center of the Genome”

“Journey to the Center of the Earth”

Jules Verne 1864
History of Exome Sequencing

"Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing"
Choi et al PNAS 2009 – Congenital Chloride Diarrhea Gene

>200 Gene Discoveries
Recessive-Dominant-De Novo
June 2013

OMIM Database - June 2013
7430 Disorders with Known or Suspected Mendelian Inheritance

3,805 Disorders with Molecular Basis Known
Potential for Further Molecular Diagnoses is Substantial

Platform Options for Exome Sequencing

Illumina HiSeq 2000 or 2500
Ion Torrent Proton

Exome Sequencing Laboratory Workflow

Genomic DNA

Library Preparation

Next Generation Sequencing Library

Hybridize to Exome Capture Probes

Exome Enriched Library

Next Generation Sequencing

Bioinformatics Analysis
### Exome Sequencing Read Data
- **FASTQ File**
- **Primary Sequence Alignment**
  - BWA/Novoalign
  - SAM/BAM File
- **Refined Sequence Alignment**
  - GATK
- **Variant Calling**
  - SAMTools/GATK
- **Variant Annotation**
  - Annovar

### Workflow for Causal/Candidate Gene Identification
- **Annotated Exome Variants ~ 20,000**
- **Prioritization by Heuristic Filtering**
- **Prioritization by Likelihood Prediction**
  - **VAAST Algorithm**
    - Case vs. Control Allele Comparison
    - Amino Acid Change/Impact
- **Candidate Variants/Genes**
  - Several to Dozens
- **Interpretive Report**
- **Correlation Studies Establishing Causality**
- **Additional Clinical Laboratory Testing**
- **Genetic Screening**
  - Similar Phenotype Patients Compare to Controls
- **Functional Studies**
  - In vitro/In vivo
- **Sanger Confirmation in Patient/Family**
Criteria for Choosing Patients for Exome Sequencing

- Genetic Etiology Strongly Suspected
- Standard Testing Negative or Impractical
- Diagnosis Likely to Impact Treatment and/or Management Decisions
- Diagnostic Yield is Greater in Family Studies
  - Families with Multiple Affected Members

Exome Sequencing – “Diagnostic Yield”

- Difficult to Determine (Yet)
- Currently: Largely Single Case Reports
  - Anecdotal Series ~20-30% Diagnosis
- NIH Undiagnosed Disease Program – 2011 Report
  - 5 Molecular Diagnoses in 30 Patients/Families (17%)
  - Several Compelling Candidate Genes

Exome Sequencing – “Diagnostic Yield”

- Diagnostic Yield Expected to Increase - By How Much?
  - Driving Forces
    - Increasingly Sophisticated Bioinformatics Will Improve Variant Detection
    - Growth in Knowledge Base of Disease Causing Genes and Variants
  - Conversion to Whole Genome Sequencing
    - Filling in the Gaps
Exome Sequencing – Case Vignette

“Diagnostic Odyssey”

8th Century BC

Exomes for “Diagnostic Odyssey”

- First Year of Life: Seizures/Dystonia
- Third Year of Life: MRI with Leukodystrophy

Heuristic Filtering + VAAST + Interpretive Review

Top Three Candidate Genes

1. Recessive
2. X-Linked
**Exomes for “Diagnostic Odyssey”**

- **wt**
- ***/wt***
- **wt/wt**

Dystonia Leukodystrophy

**X-Chromosome PLP1 (Proteolipid Protein 1) Gene Mutation**
c.617T>A, p.M206K – Novel Mutation

**Dystonia Leukodystrophy**

**PMD = Pelizaeus-Merzbacher Disorder**

**Dysmyelination/Leukodystrophy**

**PLP1 Mutations**

| SIFT Score | 0.01 |

**Exome Sequencing – Summary**

**Powerful New Approach to Inherited Disorders**
- Now Available as a Diagnostic in Several Reference Laboratories

**Implementation Challenges for Laboratories**
- Technically Demanding and Capital Equipment Intensive
- Complex and Evolving Data Analysis Requirements
- Diagnostic Yield Needs Management of Expectations
New Landscape of Genetic Testing

Increasing Complexity

Whole Genome
Exome
Multi-Gene Diagnostics
Single-Gene Diagnostics

Thank You

voelkek@aruplab.com