Next Generation Sequencing in the New York State Newborn Screening Molecular Lab

Colleen Stevens, Ph.D.
Research Scientist
NYS Newborn Screening Program
Screen Positive
Referral to Specialty Care Center for Dx
Normal IRT (bottom 95%)
IRT (bottom 99.9%)
Screen Negative

2 Mut Screen Positive:
Most confirmed (30-40 referrals, 19-37 cases)

39 Mutation Panel (Luminex)
IRT Assay
Elevated IRT (top 5%)
Normal IRT (bottom 95%)

1 Mut
VHIRT (top 0.1%)
IRT (bottom 99.9%)

0 Mut

Overall (All Screen Positive)
(900 referrals, 29-65 cases)

1 Mut Screen Positive:
Most healthy single mutation carriers (650 referrals, 9-26 cases)

0 Mut/VHIRT Screen Positive:
Most healthy (250 referrals, 1-4 cases)

D. Kay, Ph.D.
CFTR mutation identification substantially increases test specificity

As # panel mutations increases:

• PPV increases
• FPR decreases

Ideally, refer babies with 2 mutations
Need to refer babies with 0-1 panel mutations (74 babies)
Add 3rd tier CFTR sequencing

Current test
- ↓ PPV, ↑ referral rate, ↑ FP rate
- 2nd tier assays do not include all CF mutations
- Sanger sequencing - labor intensive, expensive ... Need NGS

Ultimate goal → improve CF newborn screening
- Improve test & refine algorithm
- Refer infants most likely to have CF
  - ↓ parental anxiety
  - ↓ healthcare cost
NextGen Newborn Screening for CF?

GEN News Highlights

Nov 20, 2013

FDA Clears Illumina's MiSeqDx as First Clinical NGS System

The FDA granted Illumina premarket clearance for its MiSeqDx system, the first high-throughput DNA sequencing analyzer to receive such approval. Illumina also received FDA premarket clearance for the MiSeqDx Cystic Fibrosis 139-Variant Assay, MiSeqDx Cystic Fibrosis Clinical Sequencing Assay, and MiSeqDx Universal Kit.
Illumina MiSeqDx Cystic Fibrosis Clinical Sequencing Assay (CSA)

- FDA-cleared, next generation sequencing assay
- Amplicon-based panel
- 27 CFTR exons, intron/exon boundaries, 2 deep intronic, 2 large deletions; point mutations, small in/dels
# of Samples per run
- 6 samples
- Positive Control
- Negative Control
- **Total = 8**

Hands on Time
- Library Prep: 2 hours 10 mins
- Run Time: 6 hours 35 mins
- **Total Time: ~34 hours**

Calls
- Variants in protein coding regions, intron/exon boundaries,
  2 deep intronic mutations
  2 large deletions

Output Information
- Position
- Variant Type
- Call/Frequency
- Depth
- dbSNP rsID
- Reference
- cDNA Name
- Protein Name
### MiSeqDx CF Clinical Sequencing Assay

#### MiSeq Reporter 2.2.31

**Samples (8)**

<table>
<thead>
<tr>
<th></th>
<th>Sample ID</th>
<th>Sample Name</th>
<th>Call Rate</th>
<th>Performance</th>
<th>Control</th>
<th>Comment</th>
<th>Coordinates Not Called</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>7</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>8</td>
<td>100</td>
<td>Pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NTC</td>
<td>NTC</td>
<td>0</td>
<td>Pass</td>
<td>negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Variants (6)**

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Chr</th>
<th>Position</th>
<th>Variant Type</th>
<th>Call</th>
<th>Frequency</th>
<th>Depth</th>
<th>Filter</th>
<th>rsSNP</th>
<th>RefGene</th>
<th>cDNA Name (HGVS)</th>
<th>Protein Name (HGVS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>chr7</td>
<td>117199533</td>
<td>SNV</td>
<td>A/A</td>
<td>0.99</td>
<td>2675</td>
<td>PASS</td>
<td>rs213959</td>
<td>G</td>
<td>c.1408G&gt;A</td>
<td>p.Val470Met</td>
</tr>
<tr>
<td></td>
<td>chr7</td>
<td>117199644</td>
<td>Div</td>
<td>ATCT</td>
<td>0.53</td>
<td>7820</td>
<td>PASS</td>
<td>rs113933959</td>
<td>ATCT</td>
<td>c.1521_1523delGTT</td>
<td>p.Phe508del</td>
</tr>
<tr>
<td></td>
<td>chr7</td>
<td>117235055</td>
<td>SNV</td>
<td>T/G</td>
<td>0.49</td>
<td>12779</td>
<td>PASS</td>
<td>rs1042077</td>
<td>T</td>
<td>c.2562T&gt;G</td>
<td>p.Thr854Thr</td>
</tr>
<tr>
<td></td>
<td>chr7</td>
<td>117242019</td>
<td>Div</td>
<td>T/TA</td>
<td>0.51</td>
<td>5600</td>
<td>PASS</td>
<td>T</td>
<td></td>
<td>c.2657+2_2657+3insA</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>chr7</td>
<td>117307105</td>
<td>SNV</td>
<td>G/A</td>
<td>0.44</td>
<td>6277</td>
<td>PASS</td>
<td>rs1800138</td>
<td>G</td>
<td>c.4389G&gt;A</td>
<td>p.Gln1463Gln</td>
</tr>
</tbody>
</table>
Evaluation of Illumina CF Clinical Sequencing Assay

• Validated the assay for Dried Blood Spot (DBS) DNA
  – lower DNA quality and quantity than recommended
• Genotyped 263 infants
  – most with only 1 mutation on 39-mutation panel
• 94.7% samples w/ 100% call rate on first pass
• Median depth = 9,550x (5-121,666x)
• Very high concordance with Sanger sequencing
  – 1 recurrent false positive artifact (low quality indicators)
• 2002-2012: 2.4 million babies screened (IRT)
• 150,000 tested for mutations
• 13,541 referred for sweat test
• 392 with definite CF

39- Mutation Panel

2 MUT
N=256

1 MUT
N=114

VHIRT
N=22

Clinical Seq. Assay

2 MUT
N=377

1 MUT
N=14

VHIRT
N=1

*2 MUT samples were not all run on CSA

Limitation of CSA:
Only detects 2 large deletions
# Supplemental Deletion Mutation Assays

<table>
<thead>
<tr>
<th>Deletion</th>
<th>Method</th>
<th>When Performed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949del84</td>
<td>Gap-PCR/ Gel analysis</td>
<td>All samples</td>
</tr>
<tr>
<td>Exon 2</td>
<td>qPCR</td>
<td>All samples</td>
</tr>
<tr>
<td>Exon 2 - 3</td>
<td>Gap-PCR/ Gel Analysis</td>
<td>Only to confirm NGS</td>
</tr>
<tr>
<td>Exon 17a - 17b</td>
<td>Gap-PCR/ Gel Analysis</td>
<td>All samples</td>
</tr>
<tr>
<td>Exon 17a - 18</td>
<td>Gap-PCR/ Gel Analysis</td>
<td>All samples</td>
</tr>
<tr>
<td>Exon 17b</td>
<td>qPCR</td>
<td>All samples</td>
</tr>
<tr>
<td>Exon 22</td>
<td>qPCR</td>
<td>Only to confirm NGS</td>
</tr>
</tbody>
</table>
• 2002-2012: 2.4 million babies screened (IRT)
• 150,000 tested for mutations
• 13,541 referred for sweat test
• 392 with definite CF

39- Mutation Panel
- 2 MUT N=256
- 1 MUT N=114
- VHIRT N=22

Clinical Seq. Assay
- 2 MUT N=377
- 1 MUT N=14
- VHIRT N=1

+ Supplemental Assays
- 2 MUT N=387
- 1 MUT N=5
- VHIRT N=0
Spring: implementation of Illumina NGS as a 3rd tier

IRT immunoassay

→ top 5%

39-mutation panel

→ 1 mut, VHIRT

Illumina NGS

Refer *only* if 2 mutations

- Dx at screening
- ↓ false positive referrals
- ↓ healthcare costs
- impact on families

# referrals/year 900 → 100
Blinded Pilot Study began January 23, 2017:

- After 39-mutation panel, samples with 1MUT and VHIRT are blinded to person performing CSA
- Work out logistics for batching of samples
- Evaluate workflow for supplemental assays
- Evaluate TAT
- Generate mock reports for each sample tested
- Determine impact on referral rate
Anticipated Impact on Turnaround Time

- Accessioning (1)
- IRT test (1)
- Abnormal (2)
- Repeat IRT test (2)
- Extract DNA (2)
- 39-mutation screen (3)
- Extract fresh punch (3)
- 39-mutation screen (3)
- Enter results (4)
- Mailer (5)
- Accessioning (1)
- IRT test (1)
- Abnormal (2)
- Repeat IRT test (2)
- Extract DNA (2)
- 39-mutation screen (3)
- Extract fresh punch (3)
- Next Gen / Suppl (3-5)
- Sanger (5-6)
- Enter results (6)
- Mailer (7)*

*These times don’t account for any batching

D. Kay, Ph.D.
Proposed Screening Algorithm Changes

**Tier 1**
- IRT 270,000

**Tier 2**
- Luminex 39 Mut 15,000

**Tier 3**

<table>
<thead>
<tr>
<th># Referred</th>
<th>Jan 2018 –</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 2 Mut</td>
<td>&lt;100 2 Mut/Var</td>
</tr>
<tr>
<td>650 1 Mut</td>
<td>0 1 Mut</td>
</tr>
<tr>
<td>250 VHIRT</td>
<td>0 VHIRT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50 CF</td>
<td>50 CF</td>
<td>50 CF</td>
<td>50 CF</td>
</tr>
</tbody>
</table>
**Why switch to TruSeq?**

- Custom design → more comprehensive (rare NYS mutations, deletions, duplications)
- Multiplex greater # samples
  - (CSA only 6 samples per run)
- IRT top 5% run on TSCA
  - "2\textsuperscript{nd} tier" - read only set of common NYS mutations; \textit{CFTR} seq masked
  - "3\textsuperscript{rd} tier" - read entire \textit{CFTR} seq in 1Mut and VHIRT
- Needs to be tested & validated
- Logistics - TAT, cost
NGS for SCID

Current NBS for severe combined immunodeficiency:

- Measure T-cell receptor excision circles (TRECs)
- <125 TRECs constitutes a referral
- Immunologists order CBC, flow, mitogen studies
- Molecular tests order by candidacy, multi-gene panel(s), insurance issues, available labs
- Slow, stressful process
2015 Referral Data

- 80 referrals
  - 3 Classic SCID
  - 2 Syndrome with T cell impairment
  - 7 DiGeorge syndrome
  - 2 Trisomy 21
  - 1 CHARGE syndrome
  - 8 Secondary T cell lymphopenia
  - 3 Variant SCID
  - 3 Idiopathic T cell lymphopenia
  - 37 No disease
  - 3 Lost to follow-up
  - 2 Other
  - 9 Expired w/o diagnosis
February 22, 2017

SCID is a group of disorders, caused by mutations in different genes.

- Various pathways
- Lymphocytes (T/B/NK)
- AR, XL, AD
- Phenotypes related to genes/mutations
- Treatment can differ depending on gene

Design a second-tier molecular assay for SCID NBS
- Sequence all SCID genes in infants who screen positive
- Rapidly rule-in a diagnosis, inform treatment

Cossu, 2010, I J Peds

D.Kay, Ph.D.
Specific Aims

- Validate 2 platforms for 39-gene NGS immunodeficiency panel

- Evaluate Next Gen Sequencing Utility and TAT
  - Shortened time to diagnosis?
  - Fewer visits to Specialist?
  - Earlier, targeted treatment?
  - Long-term follow-up

- Create and disseminate educational materials for parents and providers to state programs
Ion Torrent PGM

Illumina MiSeq

NGS Platforms
Multi-gene SCID panel: Design & Validation Study

- Literature/case reports, commercial SCID panels
- 39 T-cell, B-cell deficiency, syndrome genes

### Multi-gene SCID panel: Design & Validation Study

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORO1A</td>
<td>CD3Z</td>
<td>ATM</td>
<td>LIG4</td>
<td>DOCK2</td>
</tr>
<tr>
<td>PRKDC</td>
<td>DCLRE1C</td>
<td>CHD7</td>
<td>NHEJ1</td>
<td>GATA2</td>
</tr>
<tr>
<td>TBX1</td>
<td>IL2RG</td>
<td>MTHFD1</td>
<td>RAC2</td>
<td>IGHM</td>
</tr>
<tr>
<td>ADA</td>
<td>IL7RA</td>
<td>MTR</td>
<td>CD3G</td>
<td>BTK</td>
</tr>
<tr>
<td>AK2</td>
<td>JAK3</td>
<td>RMRP</td>
<td>STAT5B</td>
<td>CD40LG</td>
</tr>
<tr>
<td>CD3D</td>
<td>RAG1</td>
<td>SLC46A1</td>
<td>ZAP70</td>
<td>WAS</td>
</tr>
<tr>
<td>CD3E</td>
<td>RAG2</td>
<td>DOCK8</td>
<td>PNP</td>
<td>DKC1</td>
</tr>
<tr>
<td>CD45</td>
<td>FOXN1</td>
<td>NBN</td>
<td>BLNK</td>
<td></td>
</tr>
</tbody>
</table>

### Panel Validation

- Design, test, compare panels
  - AmpliSeq (Ion Torrent)
  - TruSeq Custom Amplicon (Illumina)

**Validation: Sanger**

Make available to community
Validation Plan

- Sequence 20 samples from babies confirmed to have an immunodeficiency using both platforms
- Sanger sequence all 39 genes in 8 of the 20 samples to establish sensitivity and specificity of the assays and to establish QC criteria for evaluating NGS data
- Include NA12878 Coriell cell line (NIST; Genome in a Bottle) as a normal control
- Establish reproducibility of the assays
Preliminary Results

NA12878 Concordance with NIST Genome in a Bottle (GIAB) high confidence calls

<table>
<thead>
<tr>
<th>Panel</th>
<th>Quality Filters</th>
<th>True Positives</th>
<th>False Positives</th>
<th>False Negatives</th>
<th>Total # bases evaluated</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmpliSeq (Ion Torrent)</td>
<td>Default</td>
<td>225</td>
<td>37</td>
<td>13</td>
<td>140,999*</td>
<td>0.95</td>
</tr>
<tr>
<td>TruSeq (MiSeq)</td>
<td>Default</td>
<td>198</td>
<td>49</td>
<td>36</td>
<td>188,863*</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*Approximation = # Targeted bases - # bases in poor coverage areas

Data for NA12878 and fully Sanger-sequenced samples will guide us in setting quality filters to decrease FP and FN calls.
Preliminary Results

- Identified Gaps in Coverage
  - 107 regions in TruSeq and 161 regions in Ampliseq
  - <30X (# reads at a position); >MQ30 (map quality score) and BQ>20 (base quality score)
  - Some gaps were only a few bases; some entire amplicons
  - “White glove” redesign of the panels

While awaiting the new panels, can we determine the causative mutations in any of our 20 validation samples with the data in hand?

Staff performing the analysis were blind to final diagnosis and results from diagnostic molecular testing.
## Pathogenic/ Likely Pathogenic Variants Identified in Samples 1-8

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gene</th>
<th>Variant</th>
<th>Zygosity</th>
<th>Sanger</th>
<th>MiSeq</th>
<th>Torrent</th>
<th>ClinVar</th>
<th>CAVA Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>ADA</td>
<td>p.Arg101Trp</td>
<td>Hom.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Pathogenic</td>
<td>missense_variant</td>
</tr>
<tr>
<td>S2</td>
<td>JAK3</td>
<td>p.Glu958*</td>
<td>Hom.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>stop_gained</td>
</tr>
<tr>
<td>S3</td>
<td>IL2RG</td>
<td>p.Arg289*</td>
<td>Hom.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Pathogenic</td>
<td>stop_gained</td>
</tr>
<tr>
<td>S4</td>
<td>IL7R</td>
<td>c.83-2A&gt;T</td>
<td>Het.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>VOUS</td>
<td>splice_acceptor_variant</td>
</tr>
<tr>
<td>S4</td>
<td>IL7R</td>
<td>p.Cys118Tyr</td>
<td>Het.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Pathogenic</td>
<td>missense_variant</td>
</tr>
<tr>
<td>S5</td>
<td>RAG2</td>
<td>p.Gly95Arg</td>
<td>Het.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Pathogenic</td>
<td>missense_variant</td>
</tr>
<tr>
<td>S5</td>
<td>RAG2</td>
<td>p.Arg167Ser</td>
<td>Het.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>VOUS</td>
<td>missense_variant</td>
</tr>
<tr>
<td>S6</td>
<td>CHD7</td>
<td>p.Trp145*</td>
<td>Het.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Pathogenic</td>
<td>stop_gained</td>
</tr>
<tr>
<td>S7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Likely causative variants not identified by either platform – poor sample quality; areas of low coverage</td>
</tr>
<tr>
<td>S8</td>
<td>IL2RG</td>
<td>p.Arg328*</td>
<td>Hom.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>VOUS</td>
<td>stop_gained</td>
</tr>
</tbody>
</table>

Concordant with Diagnostic Testing!
Multi-gene SCID panel: *Pilot Study*

1st tier screen
(real-time PCR for TREC)

↓

Screen positive infants
(low/absent TREC)

↓

Informed consent

↓

2nd tier mutation screen

↓

Report
Multi-gene SCID panel: Benefits

- Single test (couldn’t be done using traditional genotyping)
  - all known genes
  - most mutations
- Cheaper to do all genes at once, by NBS
- Molecular diagnosis at birth
- Differential diagnosis (syndromes, T- vs. B-cell deficiency)
- Phenotype predictions using genotype
- Treatment and prognosis
- Inform genetic counseling (AR, XL, AD; prenatal, carrier testing; family planning)
- Eventually rule out some false positives (screen positive, no mutations)
- Genes /regions can be added or removed/blinded
Future Application of NGS:

APHL Peer Network Grant
Development of Custom NBS NGS Panel

- Pompe
- Krabbe
- X-ALD
- MPS1
- MCAD
- VLCAD

Add in......
- HBB
- SCID
- CFTR
- And more....
Acknowledgments:

Michele Caggana, Sc.D., FACMG

Molecular Lab
Carlos Saavedra, M.D.
Denise Kay, Ph.D.
Erin Hughes
Robert Sicko
Allison Madole
Lea Krein
Sandra Levin
Lisa DiAntonio
Jason Isabelle
April Parker
Matt Nichols
Ritu Jain, Ph.D.

Wadsworth Center Applied Genomic Technologies Core