An evaluation of machine intelligence tools to diagnose genetic diseases in critically ill infants

Michelle Clark PhD
Statistical Scientist, Rady Children’s Institute for Genomic Medicine
Background

- 55% disease of unknown etiology

Search for etiological diagnosis

Interim empirical treatment

Treatment modification

Improvement or worsening

- Discharged home undiagnosed
- Delayed diagnosis: Anxiety, suffering
- Unnecessary morbidity and mortality
- Delayed palliative care
- Unnecessary cost

Page 2
Diagnostic and clinical utility of rapid whole genome sequencing

- 55% disease of unknown etiology
- Search for etiological diagnosis including rWGS
- Interim empirical treatment
- Treatment modification

- 20% Rapid Whole Genome Sequencing
- 42-57% Genetic Disease Diagnosis
- 24-34% Precision Treatment
- 8-15% Avoid Mortality and Long-term Morbidity
- 24% Improved outcomes
- 75% Shorter clinical stay
- 85% Decreased cost

2-7 days
Barriers to broad adoption

- Capital & labor intensity of rapid genomic sequencing
  - Shortage of expert medical geneticists, genetic counselors
  - Not scalable
  - Delays rapid changes in patient care

- Unfamiliarity with rapid genomic medicine
  - 13,000 genetic diseases – most of them too rare to have been seen before by pediatricians

- Insufficient evidence of efficacy
  - Delayed authorization, failure of reimbursement

- Many genetic diseases lack effective treatments
  - Most treatments have not undergone rigorous testing
Solution: automated diagnostic platform using machine intelligence

Order rapid WGS

Blood sample or dried blood spot

EHR CNLP, automated translation to HPO

Library prep

QA, normalize conc.

Sequencing

Sequence alignment

Variant calling and genotyping

Automated, supervised provisional diagnosis

Literature review

Precision medicine

Time from blood draw to provisional diagnosis: 19.5 hours
Automated deep phenotyping

1. Order rapid WGS
2. Blood sample or dried blood spot
3. EHR CNLP, automated translation to HPO
4. Library prep
5. QA, normalize conc.
6. Sequencing
7. Sequence alignment
8. Variant calling and genotyping
9. Automated, supervised provisional diagnosis
10. Literature review
11. Precision medicine

Clinitink
Automated variant interpretation

1. Order rapid WGS
2. Blood sample or dried blood spot
3. EHR CNLP, automated translation to HPO
4. Library prep
5. QA, normalize conc.
6. Sequencing
7. Sequence alignment
8. Variant calling and genotyping
9. Automated, supervised provisional diagnosis
10. Literature review
11. Precision medicine

Automated, supervised provisional diagnosis

Rady Children's Institute
Genomic Medicine

Page 7
Evaluation of the automated diagnostic platform

1. Retrospective study – 84 children
2. Timed study – 10 children
3. Reanalysis study – 48 children
4. Prospective study – 50 children
1. Performance in a retrospective cohort: 99% precision, 97% recall

- 95 children with 97 genetic diseases diagnosed manually by rapid whole genome or whole exome sequencing with manually extracted phenotypes and manual interpretation
- Excluded incidental findings
- 99% precision (93 of 94)
- 97% recall (94 of 97)
2. Timed study: 100% precision/recall
Mean time savings: 22hrs

<table>
<thead>
<tr>
<th>Use Type</th>
<th>Retrospective Patients</th>
<th>Prospective Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject ID</strong></td>
<td>263 6124 3003</td>
<td>6194 290 352 362 374 7052 412</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>8 days 14 years 1 year</td>
<td>5 days 3 days 7 weeks 4 weeks 2 days 17 months 3 days</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>♀ ♂ ♂ ♀ ♂ ♂ ♂ ♂ ♂ ♂</td>
<td></td>
</tr>
<tr>
<td><strong>Abbreviated Presentation</strong></td>
<td>Neonatal seizures</td>
<td>Rhabdomyolysis</td>
</tr>
<tr>
<td><strong>Number of Phenotypic Features</strong></td>
<td>51 115 148</td>
<td>14 2 257</td>
</tr>
<tr>
<td><strong>Molecular Diagnosis</strong></td>
<td>Early Infantile Epileptic Encephalopathy</td>
<td>Glycogen Storage Disease V</td>
</tr>
<tr>
<td><strong>Gene and Causative Variant(s)</strong></td>
<td>KCNQ2 c.727C&gt;G</td>
<td>PYGM c.2262delA c.1726C&gt;T</td>
</tr>
<tr>
<td><strong>NovaSeq Loading (hours)</strong></td>
<td>0:20 0:17 0:16</td>
<td>0:20</td>
</tr>
<tr>
<td><strong>1st &amp; 2nd Analysis (hours)</strong></td>
<td>1:03 1:02 0:59</td>
<td>0:59</td>
</tr>
<tr>
<td><strong>3rd Analysis Processing (hours)</strong></td>
<td>0:06 0:05 0:07</td>
<td>0:05</td>
</tr>
</tbody>
</table>
3. Reanalysis study: 4.2% diagnostic yield

- Automation of these reanalysis steps reduced the number of variants under consideration by an average of 99.9%.
- In two cases, diagnoses were made upon reanalysis, representing a yield of 4.2% (2 of 48).
- Four additional cases were flagged with a possible diagnosis to be considered during periodic reanalysis.
- An untrained analyst identified these six diagnoses with specificity = 0.83 and sensitivity = 0.76.
4. Prospective performance: 100% recall

• Out of 50 patients, the standard diagnostic workflow resulted in 16 (32%) diagnoses

• Automated analysis correctly diagnosed all 16 patients (100% recall)

• In addition to the standard workflow, analysts found automation to be very helpful in 4% of cases

“How helpful was automated analysis in addition to the standard workflow?”

- Not at all helpful: 18%
- Slightly helpful: 44%
- Somewhat helpful: 34%
- Very helpful: 4%
- Extremely helpful: 0%
Rady Children’s Institute for Genomic Medicine – the clinical lab perspective

• Hesitation when machine intelligence tools undergo rapid updates
• Goes against how clinical lab directors were trained to validate tools
• Need sufficient warning prior to updates
• Request increased transparency
Moon’s response to requests for transparency

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BTD</strong></td>
<td><strong>GT</strong></td>
<td><strong>p.Glu48Stop</strong></td>
<td><strong>3 publications</strong></td>
</tr>
<tr>
<td>3:15,677,022</td>
<td>ref: G</td>
<td>Biotinidase deficiency</td>
<td>AR - compound heterozygous</td>
</tr>
<tr>
<td>rs397514336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClinVar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HBB</strong></td>
<td><strong>AA</strong></td>
<td><strong>p.Glu7Val</strong></td>
<td><strong>272 publications</strong></td>
</tr>
<tr>
<td>11:5,248,232</td>
<td>ref: T</td>
<td>Sickle cell anemia</td>
<td>AR</td>
</tr>
<tr>
<td>rs334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClinVar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BTD</strong></td>
<td><strong>GC</strong></td>
<td><strong>p.Asp446His</strong></td>
<td><strong>78 publications</strong></td>
</tr>
<tr>
<td>3:15,686,693</td>
<td>ref: G</td>
<td>Biotinidase deficiency</td>
<td>AR - compound heterozygous</td>
</tr>
<tr>
<td>rs13078881</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClinVar</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reported variants

<table>
<thead>
<tr>
<th>Position</th>
<th>Genotype</th>
<th>Gene</th>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>22:38511635</td>
<td>G/A</td>
<td>PLA2G6</td>
<td>Infantile neuroaxonal dystrophy 1</td>
</tr>
<tr>
<td>22:38512190</td>
<td>G/A</td>
<td>PLA2G6</td>
<td>Infantile neuroaxonal dystrophy 1</td>
</tr>
</tbody>
</table>

Variant discussion

PLA2G6

22:38,511,635 ▼ GA ▼ p.Arg845*

Infantile neuroaxonal dystrophy 1

AR

Note

Two variants, a novel stop gained variant and a novel stop gained variant, were detected in heterozygous state in the PLA2G6 gene (ENST332509: c.1933C>T; p.Arg645* and 332509: c.1933C>T; p.Arg645*). Parental DNA analysis is required to establish a compound heterozygous state of these two variants.

Mutations in PLA2G6 have been shown to cause Infantile neuroaxonal dystrophy 1 (MIM: 256600), an autosomal recessive condition. The reported clinical phenotype of this patient overlaps with the manifestations of this condition regarding neurodegeneration, developmental regression, nystagmus, spastic tetraplegia, and cerebellar atrophy. The typical age of onset of Infantile neuroaxonal dystrophy 1 ranges from 0 to 10, which is in line with the reported age of onset in this patient (1 y). Further clinical evaluation of the patient will give more insight into the phenotypic overlap with Infantile neuroaxonal dystrophy 1.

The detected variant causes stop gained. It is absent from gnomAD and absent from dbSNP, but has not previously been associated with disease. Parental DNA analysis (trio analysis) and DNA analysis of other (un)affected relatives, could establish co-segregation of this variant with the reported clinical phenotype.

Classification ▼ Unassigned ▼

References

Enter PubMed ID or PubMed URL
The clinical lab perspective continued

- High sensitivity with automation, but unsure about sensitivity
  - Trust will come from large studies from other groups of hundreds of thousands of cases

- Development of publically available benchmarks to validate methods after every update
  - Example: Genome in a Bottle for clinical validation of genome sequencing
Conclusion

• Although the automated diagnostic system is “hands-free”, it’s supervised at every step by expert bioinformaticians, clinical medical geneticists and clinical lab directors.

• May enable effective first-tier, provisional diagnoses or automated re-analysis of unsolved cases.

• Wide-spread adoption would allow valuable cognitive resources of molecular laboratory directors and analysts to be reserved for difficult cases, manual curation of variants, and clinical report generation.
Acknowledgments

Michelle M. Clark¹, Kiely N. James¹, Amber Hildreth¹,²,³, Sergey Batalov¹, Yan Ding¹, Shimul Chowdhury¹, Kelly Watkins¹, Katarzyna Ellsworth¹, Brandon Camp¹, Cyrielle I. Kint⁴, Calum Yacoubian⁵, Lauge Farnaes¹,², Matthew N. Bainbridge¹,⁶, Curtis Beebe⁷, Joshua J. A. Braun¹, Margaret Bray⁸, Jeanne Carroll¹,², Julie A. Cakici¹, Sara A. Caylor¹, Christina Clarke¹, Mitchell P. Creed⁹, Jennifer Friedman¹,¹⁰, Alison Frith⁵, Richard Gain⁵, Mary Gaughran¹, Shauna George⁷, Sheldon Gilmer⁷, Joseph Gleeson¹,¹⁰, Jeremy Gore¹¹, Haiying Grunenwald¹², Raymond L. Hovey¹, Marie L. Janes¹, Kejia Lin⁷, Paul D. McDonagh⁸, Kyle McBride⁷, Patrick Mulrooney¹, Shareef Nahas¹, Daheoon Oh¹, Albert Oriol⁷, Laura Puckett¹, Zia Rady¹, Martin G. Reese¹³, Julie Ryu¹,², Lisa Salz¹, Erica Sanford¹,², Lawrence Stewart⁷, Nathaly Sweeney¹,², Mari Tokita¹, Luca Van Der Kraan¹, Sarah White¹, Kristen Wigby¹,², Brett Williams⁵, Terence Wong¹, Meredith S. Wright¹, Catherine Yamada¹, Peter Schols⁴, John Reyners⁸, Kevin Hall¹², David Dimmock¹, Narayanan Veeraraghavan¹, Thomas Defay⁸, Stephen F. Kingsmore¹*.  
¹Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA;  
²Department of Pediatrics, University of California San Diego, San Diego, CA 92093, USA;  
³Department of Pediatrics, University of Washington, Seattle, WA 98195, USA;  
⁴Clinithink Ltd., London N1 6DR, UK;  
⁵Codified Genomics, LLC, Houston, TX 77033, USA;  
⁶Rady Children's Hospital, San Diego, CA 92123, USA;  
⁷Alexion Pharmaceuticals, Inc., New Haven, CT 06510, USA;  
⁸University of Kansas School of Medicine, Kansas City, MO 66160, USA;  
⁹Department of Neurosciences, University of California San Diego, San Diego, CA 92093, USA;  
¹⁰Tessella, Needham, MA 02494, USA;  
¹¹Illumina, Inc., San Diego, CA 92122, USA;  
¹²Fabric Genomics, Inc., Oakland, CA 94612, USA.

Support:  
NICHD  
NHGRI  
Illumina  
Alexion  
Diploid  
Clinithink  
Fabric