Assuring the Quality of New DNA Sequencing Technologies in the Clinical Laboratory

Ira M. Lubin, Ph.D., FACMG
Laboratory Research and Evaluation Branch
Division of Laboratory Science and Standards
LSPPPO, OSELS, CDC

CLIAC
Wednesday, March 6, 2013

The findings and conclusions in this presentation are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry

Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Agency for Toxic Substances and Disease Registry, the Public Health Service, or the U.S. Department of Health and Human Services
Outline

New Sequencing Technologies (for the purpose of this talk)
  = Next-generation sequencing (NGS)
  = Massively-parallel sequencing

- NGS in clinical settings
  - Applications
  - Technology: from Sanger to next-generation sequencing

- Challenges and approaches to assuring NGS quality
  - Next Generation Sequencing-Standardization of Clinical Testing (Nex-StoCT) Workgroup and outcomes – CDC facilitated
  - Complementary efforts of other groups
  - New projects

- Discussion
NGS of the Human Genome

- Diagnose rare diseases
- Directing cancer therapy
- Other clinical applications in development

Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease


Identification of a Novel TP53 Cancer Susceptibility Mutation Through Whole-Genome Sequencing of a Patient With Therapy-Related AML


Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing

Sci Transl Med. 2011: 3 (65) 65ra4

Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units

Sci Transl Med. 2012: 4, 154ra135
NGS in Clinical and Public Health Microbiology

- **Species identification**
- **Drug susceptibility testing and detecting virulence determinants**

Didelot X. et al., *Nature Reviews Genetics*, 2012


NGS is Transforming the Landscape of Public Health

- Genomic Epidemiology
- Pathogen Evolution
- Culture-Independent Microbiology

Gardy J.L. et al.

Rasko D.A. et al.

hmPC. *Nature* 2012; 486:215–221
Sanger Sequencing

- **Mature chemistry**
  - 800 + base length reads

- **Usually sequence is bi-directional**
  - forward and reverse strand

- **Established base calling algorithms**
  - quality scores

- **Accuracy Approaches 100%**
  - Sanger Sequencing is currently considered the **Gold Standard**

Next Generation Sequencing (NGS)

- All NGS technologies are based on massively parallel sequencing of DNA fragments.
- Platforms capable of >1 billion reads per run
- Average read is ~50-400 bp (human genome is 3x10^9 bases)
- Need to “connect” sequences by alignment or assembly
- Informatics needed to determine what is a clinically relevant finding

The coverage necessary to make accurate variant calls across targeted regions should be established empirically during the validation of each NGS application.
NGS in Clinical Practice

1. Indication for testing / counseling / test order

2. Sequence analysis
   a) Sample preparation
   b) Machine sequencing
   c) Alignment / Assembly
   d) Variant call
   e) Variant/gene annotation
   f) Determine findings of Clinical significance
   g) Result reporting

3. Integration into clinical decision making

Physical patient sample

Digital patient Sample
(informatics cost can be equal to or greater than the machine sequence analysis)
Challenges for Assuring the Quality of NGS in Clinical Practice

- Meeting existing regulatory requirements and professional standards
  - Defining performance specifications
  - Test validation
  - Quality control procedures
  - Independent assessment of test performance (proficiency testing/ alternate assessment)
  - Reference materials

- Common standards and guidelines for ensuring the reliability of NGS results are beginning to emerge
  - CDC - Next Generation Sequencing- Standardization of Clinical Testing (Nex-StoCT)
  - College of American Pathologists LAP Checklist for NGS
Next Generation Sequencing- Standardization of Clinical Testing (Nex-StoCT) Workgroup (April 2011)

- Develop a set of principles and guidance useful as a framework for implementing NGS into clinical settings
- Emphasis on human genetic disease applications
- 41 participants with extensive knowledge and experience in NGS that included:
  - Clinical and research laboratory professionals
  - Physicians
  - Test platform and software developers
  - Bioinformatics experts
  - Professionals from government agencies (FDA, CMS, NIST, NIH) and an accrediting body
  - Active members of professional organizations
Nex-StoCT: Human Genetic Applications

Next-Generation Sequencing

Gene Panel
- To 100+ genes

Exome
- 180,000 Exons
- ~25,000 genes; 3x10^6 bases

Whole Genome
- 3x10^9 bases

Gene panels by Next-Generation Sequencing
- Congenital muscular dystrophy
- Congenital disorders of glycosylation
- X-linked intellectual disability
- Autism spectrum disorder
- Cardiomyopathy (Dilated / Hypertrophic)
- Mitochondrial encephalopathy
Assuring the quality of next-generation sequencing in clinical laboratory practice

To the Editor: We direct your readers’ attention to the principles and guidelines (Supplementary Guidelines) developed by the Nex-StoCT Workgroup. These guidelines provide initial steps to ensure that results from tests based on next-generation sequencing (NGS) are reliable and useful for clinical decision-making. The Clinical and Laboratory Standards Institute (CLSI) recognizes the national workgroups within the Clinical Laboratory Standards Institute (CLSI) platform, Independent approaches for establishing technical processes involving quality management systems (QMS) to ensure the analytical validity and compliance of NGS tests with existing regulatory and professional quality standards. The workgroup identified and addressed gaps in quality practices that could compromise the quality of both clinical laboratory services and translational efforts needed to advance the implementation and utilization of NGS in clinical settings.

The workgroup was composed of experts with knowledge of NGS and included clinical laboratory directors, clinicians, pathologists, and informatics and biotechnology experts, and individuals actively engaged in NGS guidelines development from accreditation bodies and professional organizations. Representatives from US government agencies also participated. These guidelines address four topics that are components of quality management in a clinical environment: (1) NGS test validation, (2) quality control (QC) procedures to ensure and maintain accurate results, (3) the independent assessment of test performance through proficiency testing (PT) or alternative approaches (4) reference materials (RMs). Discussions were limited to the analytical and informatics processes required for accurate variant calling. The workgroup did not address how variants are prioritized, interpreted, or reported.

Table 1: Selective workgroup recommendations for establishing NGS test systems for clinical use

<table>
<thead>
<tr>
<th>Objective</th>
<th>NGS-specific recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test validation</td>
<td>Quality control (QC) procedures to ensure and maintain accurate results.</td>
</tr>
<tr>
<td>Performance testing</td>
<td>The independent assessment of test performance through proficiency testing (PT) or alternative approaches.</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Reference materials (RMs) are used in the analysis of test results.</td>
</tr>
</tbody>
</table>

Supplementary Guidelines

Table of Contents

1. Introduction..................................................2-3
2. Background..................................................3-4
   a. Differences between NGS and Sanger sequencing........3-4
   b. NGS applications..........................................4-5
   c. NGS Platform characteristics............................5-6
   d. Data management and downstream informatics analysis....6-9
3. Methods....................................................9-10
4. Validation................................................11-13
   a. Platform, test, and informatics pipeline validation......11-13
   b. Limitations: considerations for homologous sequences...13-14
   c. Establishing performance specifications................14-25
   d. Informatics: assessing the data analysis pipeline......25-28
   e. Indications for repeating a validation....................28-29
5. Quality Control (QC).......................................29-31
   a. Process steps to be addressed............................29-31
   b. Measuring analytical performance during the run: metrics and their applications...31-33
   c. Use of confirmatory testing................................33-34
6. Strategies for Proficiency Testing (PT)..................34-35
   a. Purpose of PT.............................................34-35
   b. Alternate assessment (AA): considerations for laboratories prior to availability of a PT program...35-36
   c. How to provide PT for NGS..............................36-37
   d. Sources of PT Samples.....................................37-39
   e. Characterization of PT samples............................39-40
   f. Possible approaches to provide PT for NGS..............40-43
   g. PT frequency and strategy................................43
7. Development and use of reference materials.............43-45
   a. Importance of RMs for NGS...............................43-45
   b. Approaches to develop characterized RMs................45-47
8. Conclusions...............................................47-48
9. Figure and Tables..........................................49-58
10. References..............................................59-64
Other Efforts in the US
(Professional/Standard Setting Organizations)

- **Food and Drug Administration**
  - Ultra High Throughput Sequencing for Clinical Diagnostic Applications- Approaches to Assess Analytical Validity (June 23, 2011) ([http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm255327.htm](http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm255327.htm))

- **College of American Pathologists**
  - Checklist for laboratory accreditation standards (Released July 2012)
    - Covers analytical and bioinformatics workflow

- **American College of Medical Genetics and Genomics**
  - Points to consider in Clinical Application of Genomic Sequencing (2012)
  - Other technical guidance (anticipated 2013)

- **Clinical and Laboratory Standards Institute**
  - MM09 - Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline (anticipated, latter 2013)

- **Association for Molecular Pathology**
  - Technical guidance in development
Nex-StoCT Workgroup
Principles and Recommendations
(http://www.nature.com/nbt/journal/v30/n11/full/nbt.2403.html)

- Test validation
- Quality control procedures
- Independent assessment of test performance
- Reference materials
# Performance Specifications for NGS: Nex-StoCT Definitions

<table>
<thead>
<tr>
<th>Performance Characteristics</th>
<th>CDC Workgroup established definitions for NGS applications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>The degree of agreement between the nucleic acid sequences derived from the assay and a reference sequence.</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>The degree to which repeated sequence analyses give the same result- repeatability (within-run precision) and reproducibility (between-run precision).</td>
</tr>
<tr>
<td><strong>Analytic Sensitivity</strong></td>
<td>The likelihood that the assay will detect the targeted sequence variations, if present.</td>
</tr>
<tr>
<td><strong>Analytic Specificity</strong></td>
<td>The probability that the assay will not detect a sequence variation when none are present (the false positive rate is a useful measurement for sequencing assays).</td>
</tr>
<tr>
<td><strong>Reportable Range</strong></td>
<td>The region(s) of the genome for which the NGS technology can accurately produce sequence information (e.g. multiple genes, exome, large genomic regions).</td>
</tr>
<tr>
<td><strong>Reference Range</strong></td>
<td>Reportable sequence variations the assay can detect that are expected to occur in an unaffected population.</td>
</tr>
</tbody>
</table>
Validation Framework for the Implementation of Clinical NGS Testing

TEST DEVELOPMENT / OPTIMIZATION

VALIDATION

PLATFORM

TEST

Informatics PIPELINE

PATIENT TESTING

QC

PT/AA

DAILY

PERIODICALLY

IT / BIOINFORMATICS INVOLVEMENT
Quality Control

- Quality control procedures monitor each step of the test to ensure that results are reliable

NGS process steps

Sample Preparation | Library Preparation | Sequence Generation | Sequence Analysis | Result Reporting

Metrics (examples)
- Coverage
- Allelic read percentage
- Quality Scores
- Reads mapped to reference
- Strands bias
- GC bias
- Transition/Transversion Ratio

- Use of alternate methods (e.g., Sanger sequencing)
- Use of confirmatory testing
- Data management
Proficiency Testing or Alternate Assessment

- Clinical laboratories are required to demonstrate the independent assessment of test performance through PT/AA

- A mechanism for comparing inter-laboratory test performance to identify:
  - Analytical and interpretive errors
  - Problems with quality control
  - Instrument calibration problems
  - Assay design issues
Proficiency Testing/Alternate Assessment

NGS Offers a Unique Paradigm for PT/AA

- Ideally, PT materials should represent the diversity and distribution of sequence variations comparable to ones that the assay is designed to detect.

- Methods-based PT challenges (rather than analyte-based) may be optimal (laboratories use different methods, test different regions of the genome).

- The cost and time needed to establish and run a PT program for NGS may be significant.

PT/AA materials useful for NGS:

- DNA from a well characterized cell line or patient sample
- Electronic data
Informed consent

Genomic DNA isolation; quantification and aliquots for distribution

1. Compile
2. Resolution of discordant variants by alternate technology (e.g., Sanger sequencing)

- Highly characterized samples
- Master list of variants in CAP genome
- Variants will cover spectrum of clinical testing (panel, exome, whole genome)

Sequencing by multiple vendors

Slide content courtesy of:
Nazneen Aziz, Ph.D., Director, Molecular Medicine Transformation Program Office, College of American Pathologists

NGS Proficiency Testing
Pilot Projected for 2013
### Availability of Reference Materials for NGS

**Types of materials:**
Genomic/mitochondrial DNA, Synthetic DNA, Electronic files

<table>
<thead>
<tr>
<th>Availability of characterized reference materials</th>
<th>DNA tests (not sequencing)</th>
<th>Sequence test (Sanger + NGS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited</td>
<td>Limited</td>
<td>Limited</td>
</tr>
</tbody>
</table>

| Discreet set of sequence variations                | Yes                       | No*                          |
| Inclusion of variants that account for significant number of clinical findings | Yes | No* |
| Well characterized reference material             | Yes                       | NO (reference build changes) |
| Need for electronic reference materials           | No                        | Yes (primarily for NGS)      |

*Nex-StoCT Recommendation: Make use of both disease-associated and naturally occurring sequence variations*
Reference Materials and Characterization

Genetic Testing Reference Materials Program (CDC)
National Center for Biotechnology Information (NIH)

Partners: Coriell Cell Repositories, Clinical and research laboratories

Lymphoblastoid cell lines (NA19240, NA12878)

Made available to clinical/research laboratories

Sequence analysis

A web-based tool for clinical laboratories
1. Compare all data sets + quality metrics
2. Consensus track w/quality indicator for each position

Purpose: Aid to clinical laboratories to determine if their results agree with others and to identify regions difficult to sequence by NGS
Genome in a Bottle Consortium

- NIST hosted August 2012 workshop to form 4 working groups:
  - Reference Material Selection & Design
    - Andrew Grupe, Celera
  - Measurements for Reference Material Characterization
    - Elliott Margulies, Illumina
  - Bioinformatics, Data Integration, and Data Representation
    - Steve Sherry, NCBI
  - Performance Metrics
    - Justin Johnson, EdgeBio

- Progress
  - Working on informed consent issues for reference materials
  - Selected CEPH family and 3 Personal Genome Project trios as prospective samples

Variation list, Performance metrics

Sample preparation → Sequencing → Bioinformatics

Slide courtesy of: Justin Zook, Ph.D., Biochemical Science Division, National Institute of Standards and Technology
Nex-StoCT II – Informatics Workgroup
October 11-12, 2012 at CDC in Atlanta, GA

- Focus: to define issues and potential solutions to NGS bioinformatics implementation in the clinical lab setting

- Considerations:
  - Optimizing a bioinformatics pipeline
  - Benefits and limitations of existing software tools
  - Defining good practices for accurately calling variants
  - Consider what is needed in terms of “electronic” reference materials for optimization
What is Clinically Relevant?

Sequence variants called (the total set)

- Variant Annotation (automated)
  - Variant Assessment (automated + manual)

Tag with properties
- Examples:
  - Change protein structure
  - Prevalence
  - Reported as disease associated

Assess relevance
- Which variants are important for
  The patient tested?

Benign
- Likely Benign
- Variation of Unknown Significance
- Likely Clinically Relevant
- Clinically Relevant
Clinical NGS testing is analytically and clinically valid for a number of medical scenarios.

The technology and applications of clinical NGS testing is evolving.

Major stakeholders are actively collaborating to advance guidance and standards.

CDC Principles and Guidance Manuscript (Nat Biotech 2012;30:1033)
- Application of definitions for performance specifications
- Test validation, quality control procedures, proficiency testing, reference materials
Current Projects

- Nex-StoCT II — Informatics (manuscript in preparation)
- Genomic reference materials and web-based tools for clinical NGS
- Standardizing NGS file formats (CDC, NIH-NCBI, NIST, FDA, others)
- Exploring expansion to clinical applications for clinical microbiology and public health
Next Generation Sequencing: Standardization of Clinical Testing (Nex-StoCT) Working Groups

Outcomes from Nex-StoCT I published

Background
Next Generation DNA sequencing technologies are currently transitioning from research into clinical and public health settings. While the healthcare and public health benefits that can be achieved using these complex technologies are significant, adoption will require robust quality assurance and control procedures to ensure reliable test results. In the US, the Clinical Laboratory Improvement Amendments (CLIA) regulations require that specifications be established for specific performance characteristics to ensure reliable test results. These performance characteristics include accuracy, precision, analytical sensitivity, analytical specificity, reportable range, reference intervals, and other applicable metrics. At present, guidance to implement quality practices necessary to establish the required performance specifications is lacking and laboratories are challenged with translating next-generation sequencing to fit with existing regulatory and professional standards. In an effort to address these issues and to assure high-quality testing, the Genetics Team within the Division of Laboratory Science and Standards (DLSS) has initiated an effort to identify principles and develop consensus guidance to help laboratories achieve reliable test results.

Nex-StoCT I Working Group
- The Genetics Team within DLSS convened a national workgroup for the Next Generation Sequencing Standardization of Clinical Testing (Nex-StoCT) meeting on April 11-15, 2011 in Atlanta, Georgia, tasked with identifying principles and developing recommendations for assuring quality laboratory practices relevant to the clinical use of NGS for the detection of Sarcoma.
Discussion

- Questions/comments about the presentation

- In addition to what was presented, what additional challenges and possible approaches do you envision for assuring the quality of NGS in the clinical laboratory setting?