Challenges in Developing Healthcare IT Data Standards for Genomics

Mollie Ullman-Cullere
Dana-Farber Cancer Institute and Partners Healthcare
NHGRI Workshop
Genomics and HIT systems: Exploring the issues
April 27-28, 2011
Requirements for Genetic/Genomic Data in Clinical Care

• Accessible in the EHR as Structured Data
• Integrated into Clinical Workflows
• Available to Clinical Decision Support
• Integrated into Data Warehouses
  – patient panel management, outcomes analysis, quality assurance, reporting and discovery research
• Maintained Up-to-Date Interpretations

Function Like Other Laboratory Test Results
External Communications:
• Billing & Pre-approvals
• Public Health Reporting
• Patient Referrals
• Pharmacy Orders

Electronic Health Record (EHR) – Components & Data Flow

- Patient Applications
- Clinical Applications
- Electronic Health Record (EHR)
- Knowledgebases
- Clinical Decision Support (CDS)
- Clinical Data Warehouse
- Research Data Warehouse
- Advanced Molecular Diagnostics
- Radiology, Cardiology, etc.
- Pathology
- Laboratory Tests

Electronic Health Record (EHR) – Components & Data Flow

- Patient Applications
- Clinical Applications
- Electronic Health Record (EHR)
- Knowledgebases
- Clinical Decision Support (CDS)
- Clinical Data Warehouse
- Research Data Warehouse
- Advanced Molecular Diagnostics
- Radiology, Cardiology, etc.
- Pathology
- Laboratory Tests

Dana-Farber / Partners HealthCare – Copyright 2011 – All Rights Reserved
How do we not make this a picture of Genomic Medicine and Healthcare?

# Current HIT Clinical Genomics Data Standards

<table>
<thead>
<tr>
<th>Ensure transfer of data between systems ...</th>
<th>Health Level Seven – HL7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure standard description of tests, results, and interpretations ...</td>
<td>LOINC, Logical Observation Identifiers Names and Codes&lt;br&gt;HGVS Nomenclature, Human Genome Variation Society&lt;br&gt;HGNC, Human Gene Nomenclature Committee&lt;br&gt;RefSeq, Reference Sequences NCBI&lt;br&gt;dbSNP, Single Nucleotide Polymorphism&lt;br&gt;ISCN, International Society for Cytogenetics Nomenclature</td>
</tr>
<tr>
<td>Ensure standard context for interpretations (i.e. associations) ...</td>
<td>SNOMED &amp; RxNORM</td>
</tr>
<tr>
<td>Other References</td>
<td>LRG, OMIM, COSMIC, PubMed...</td>
</tr>
</tbody>
</table>
Portion of LOINC Panel for DNA Variant Details

<table>
<thead>
<tr>
<th>LOINC Code</th>
<th>Name</th>
<th>Example value</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>51958-7</td>
<td>Transcript Reference Sequence Identifier</td>
<td>NM_005228.3</td>
<td>NCBI DB</td>
</tr>
<tr>
<td>48018-6</td>
<td>Gene identifier</td>
<td>EGFR</td>
<td>HGNC Nomenclature</td>
</tr>
<tr>
<td>48004-6</td>
<td>DNA Sequence Variation</td>
<td>c.2573T&gt;G</td>
<td>HGVS Nomenclature</td>
</tr>
<tr>
<td>48003-8</td>
<td>DNA Sequence Variation identifier</td>
<td>rs121434568</td>
<td>NCBI dbSNP</td>
</tr>
<tr>
<td>48002-0</td>
<td>Genomic source class</td>
<td>Somatic, Likely Somatic, Unknown Origin, Likely Germline, Germline</td>
<td>LOINC Answer List</td>
</tr>
<tr>
<td>51961-1</td>
<td>Drug efficacy sequence variation interpretation</td>
<td>Resistant, Responsive, Presumed Resistant, Presumed Responsive, Unknown Significance, Benign, Presumed Benign, Presumed Non-Responsive</td>
<td>LOINC Answer List</td>
</tr>
</tbody>
</table>
Portion of HL7 Genetic Results v2 Message

**Header**

- **Medication assessed**: Erlotinib
- **RxNorm**: 537525

**OBX**

- **OBX|1|CWE|51964-5**
  - **Gene identifier**: EGFR
  - **Transcript reference sequence identifier**: NM_005228.3
  - **DNA sequence variation Identifier**: rs121434568

**OBX|2|CWE|51964-5**

- **Drug efficacy analysis overall interpretation**: Responsive

**OBX|3|CWE|51964-5**

- **Genetic analysis summary report**

**OBX|4|FT|51969-4**

- **Genetic analysis discrete result panel**
  - **Gene identifier**: EGFR
  - **Transcript reference sequence identifier**: NM_005228.3
  - **DNA sequence variation Identifier**: rs121434568

**OBX|10|CWE|51964-5**

- **Genomic source class**: Somatic

**OBX|11|ST|47998-0**

- **DNA sequence variation display name**: c.2573T>G (p.Leu858Arg), Exon 21, EGFR, Responsive

**OBX|12|CWE|51964-5**

- **Drug efficacy sequence variation interpretation**: Responsive

---

Dana-Farber / Partners HealthCare – Copyright 2011 – All Rights Reserved
Co-Creat Human Readable & Computer Readable Elements for the EHR

Launched August 2006

This slide is intended to illustrate general IT functionality. Content displayed is not intended for clinical use. Screen configurations may not reflect the current version of the application.
HL7 Implementation Guides for Structuring Clinical Genetic Test Results within Established Standards

Published
• Gene Variants associated with Disease/Risk, Drug Metabolism and Drug Efficacy

In Ballot
• Cytogenetics and Array CGH
• Genetic Test Report (alt. format for transmission of codified findings)

Under Development
• Gene Variants for Tumor Profiling
• Expression Profiling

Piloting Organizations
• Healthcare Providers: Partners Healthcare, Dana-Farber/Brigham and Women’s Cancer Center, and Intermountain Healthcare
• Laboratories: Laboratory for Molecular Medicine at Partners Healthcare, Dana-Farber/Brigham and Women’s Cancer Center at Harvard Medical School, and ARUP Laboratories at University of Utah
What’s Needed for Standards Development?

Roadmap, Pilot Projects, Tools, Collaborators, and Community Involvement
Healthcare IT (HIT) Standards Development Best Practices

1. Participate in the healthcare standards communities (HIT and Genetics), as well as national initiatives
   - Participate in community review and publication (which is a separate track from journal publication).
   - Use tools for generation/translation/validation into standard representation e.g., HGVS’s Mutalyzer tool: http://www.mutalyzer.nl/2.0/

2. Align standards development and real-world implementation projects
   - Does it enable useful functionality while supporting professional, legal and policy requirements?

3. Collaborate with key stakeholders
   - **Vocabulary/Message Standards**: NCBI, NLM, HGVS, HL7 and LOINC
   - **Practitioners**: Geneticists, clinicians, pathologists, and molecular diagnostic laboratories
   - **IT Professionals**: EHR, Clinical Research, Bioinformatics, and LIMS Developers
Mutalyzer 2.0 β-8
released on 31 Jan 2011

HGVS nomenclature version 2.0

Welcome to the Mutalyzer web site

The aim of this program suite is to support checks of sequence variant nomenclature according to the guidelines of the Human Genome Variation Society.

Different interfaces are provided to collect the information necessary for the checks:

- The Name Checker takes the complete sequence variant description as input and checks whether it is correct.
- The Syntax Checker takes the complete sequence variant description as input and checks whether the syntax is correct.
- The Position Converter can convert chromosomal positions to transcript orientated positions and vice versa.
- The GenBank Uploader allows you to upload and use your own reference sequence.
- The SNP converter allows you to convert a dbSNP rsld to HGVS notation.
- The Webservices page provides instructions for the webservices.
- The Batch Checkers are interfaces that accept a list of inputs. These interfaces can be used for large quantities of checks.

GenBank sequences are retrieved from the NCBI (Copyright and Disclaimers). This project is sponsored by SUN Microsystems with server hardware within the scope of the Academic Excellence Grant (AEG) program (award EDUD-7832-080223-CNE).
NCBI’s dbSNP – Enhanced for the Clinical Perspective

- Single Nucleotide Polymorphism
- OMIM

Reference SNP Cluster Report: rs113488022

Allele
- Variation Class: SNP: single nucleotide polymorphism
- RefSNP Alleles: A/C/T
- Allele Origin: T: Germline, A: Somatic, C: Somatic
- Ancestral Allele: T
- Clinical Source:
- MAF/MinorAlleleCount: NA
- MAF Source:

O.001 MELANOMA, MALIGNANT, SOMATIC [BRAF, VAL600GLU] dbSNP:rs113488022

COLORECTAL CANCER, SOMATIC, INCLUDED, THYROID CARCINOMA, PAPILLARY, SOMATIC, INCLUDED, ASTROCYTOMA, LOW-GRADE, SOMATIC, INCLUDED

The val600-to-glu (V600E) mutation caused by a 1799T-A transversion in the BRAF gene was previously designated VAL599GLU (1796T-A). Kumar et al. (2003) noted that an earlier version of the BRAF sequence showed a discrepancy of 3 nucleotides in exon 1; based on the correct sequence, they proposed a change in nucleotide numbering after nucleotide 94 (the ATG codon by +3 and a corresponding codon change of +1. §

Malignant Melanoma

Davies et al. (2002) identified a 1799T-A transversion in exon 15 of the BRAF gene that leads to a val600-to-glu (V600E) substitution. This mutation accounted for 92% of BRAF mutations in...
Translational Frameworks

- Adheres to HIT standards development best practices
- Focus on structured/codified data and terminologies extending HIT data standards for inclusion of genetic standards used in patient care
Recommendations to Extend Healthcare IT (HIT) Standards for Personalized Medicine

1. Define roadmap for parallel development of Electronic Health Records and Personalized Genomic Medicine

   For example:
   - Standard(s) for coding genetic based disease
   - Standard(s) for representation of genetic data (for human and computer consumption)
   - Minimal core data sets
   - Standard(s) for representation of clinical associations
   - Standard(s) for representation of clinical decision support rules

2. Fund tool development generating/translating/validating standard representation of data for Personalized Genomic Medicine (e.g. HGVS’s Mutalyzer)

3. Make HIT resources easier to find by listing published standards and implementation guides within PubMed

4. Incorporate HIT standards into grants for genomic medicine