Evaluation of LOINC for Representing Constitutional Cytogenetic Test Result Reports

Yan Z. Heras, MS[1, 2], Joyce A. Mitchell, PhD [2], Marc S. Williams, MD, FACMG[1], Arthur R. Brothman, PhD, FACMG[3], Stanley M. Huff, MD[1, 2]
[1]Intermountain Healthcare, Salt Lake City, UT;
[2]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT;
[3]Departments of Pediatrics, Human Genetics and Pathology, University of Utah, and ARUP Laboratories, Salt Lake City, UT

Abstract

Genetic testing is becoming increasingly important to medical practice. Integrating genetics and genomics data into electronic medical records is crucial in translating genetic discoveries into improved patient care. Information technology, especially Clinical Decision Support Systems, holds great potential to help clinical professionals take full advantage of genomic advances in their daily medical practice. However, issues relating to standard terminology and information models for exchanging genetic testing results remain relatively unexplored. This study evaluates whether the current LOINC standard is adequate to represent constitutional cytogenetic test result reports using sample result reports from ARUP Laboratories. The results demonstrate that current standard terminology is insufficient to support the needs of coding cytogenetic test results. The terminology infrastructure must be developed before clinical information systems will be able to handle the high volumes of genetic data expected in the near future.

Introduction

The successful completion of the Human Genome Project on April 14, 2003, marked the beginning of the "genome era", and subsequent gene discoveries are leading to major advances in both diagnosis and treatment. The number of clinically available genetic tests is rapidly growing. When GeneTests, supported by the National Institutes of Health, started tracking laboratories providing genetic tests in 1993, there were 110 disease tests available. Today there are about 1700 disease tests available.[1] Genetics is becoming increasingly important to health care providers and genetic testing is being integrated into medical practice in many areas of medicine. Even though genomic advances promise to improve patient care, the explosion of information and knowledge in the areas of genetics, genomics, and health care can be overwhelming. This information and knowledge explosion, coupled with the lack of integration of genetic testing information with traditional patient data, presents great challenges if we are to take full advantage of genomic advances in medical practice.

Many physicians have reported a lack of basic knowledge and confidence about medical genetics, which limits their ability to appropriately counsel their patients and accurately interpret genetic tests.[2] Missed opportunities for health professionals to educate patients and families regarding genetics have been identified.[3] In addition to the competency of medical staff, the variation and format of test requisitions and result reports have contributed to poor communication between testing laboratories and clinicians.[4] The quality of patient care may be compromised as a consequence.

The importance of standardizing genetic test result reports is well recognized. Efforts have already begun to address this issue within the laboratory testing industry. For example, some model reports for molecular genetic testing have been developed[4] and the College of American Pathologists (CAP) provides a checklist for result reporting.[5] However, little has been done to address how to use information technology to improve the use of genetic test results in medical practice. In particular, the use of standard controlled terminology and information models for exchanging and storing genetic test result reports in Electronic Medical Records (EMRs) remains relatively unexplored.

It is widely agreed that information technology, especially Clinical Decision Support Systems (CDSS), has the potential to reduce medical errors, and to improve quality, safety, and efficiency of health care. Bringing genetic tests results into the patient's EMR is one of the essential first steps in translating genetics and genomic knowledge into daily medical practice. However, it will be very difficult to apply decision support if the genetic test results are simply transmitted and stored as narrative text or as images in the EMR. Establishing standard

logical representations for genetic data using controlled terminologies and information models is a prerequisite to establishing genetic CDSS as part of an EMR system.

The Logical Observation Identifiers Names and Codes (LOINC) system was adopted by the Consolidated Health Informatics (CHI) initiative as the standard vocabulary for observation identifiers for use in electronic exchange of laboratory test results. Health Level Seven (HL7) version 2 is considered to be the most widely implemented standard for healthcare information in the world. LOINC was designed to provide universal identifiers for observations sent in messages in data exchange standards like HL7 and Digital Imaging and Communication in Medicine (DICOM). For example, LOINC provides a code system for the observation identifier field (OBX-3) of the HL7 observation reporting message. Other fields in the HL7 messages provide additional semantic structures that are needed to reflect a model of laboratory testing orders and results observations. Since the first release of LOINC over ten years ago, LOINC content has continued to grow and LOINC has become the most widely adopted standard for laboratory test result names in the United States and internationally.

Clinical cytogenetics is the study of the genetic constitution of individuals by examining the structure and organization of chromosomes. Chromosome tests were introduced into clinical practice in the late 1950s. Constitutional cytogenetic tests can detect pre-existing numerical and structural abnormalities prenatally or after birth. Chromosomal abnormalities have been found to be the etiology for a number of multiple congenital anomaly syndromes as well as isolated mental retardation and developmental delay. Certain chromosomal abnormalities are consistently associated with medical conditions that require screening and management for the affected patient. Given their rarity and the lack of readily available clinical information, these conditions present excellent opportunities for CDSS.

The International System for Human Cytogenetic Nomenclature (ISCN) was created by the International Standing committee on Human Cytogenetic Nomenclature to represent the outcome of cytogenetic tests. The latest version of ISCN was published in 2009. One of the aims of ISCN is to prevent confusion in reporting research cytogenetics results. ISCN is accepted as a standard within the industry. It specifies the nomenclature to describe karyotypes, chromosome abnormalities, in situ hybridization, etc. The CAP checklist for cytogenetics includes an item to assure that current ISCN is used correctly in a final report.

The goal of the current study is to formulate a model for the electronic exchange of coded cytogenetic test results and to determine how LOINC codes fit into the model, and to evaluate whether current LOINC codes are adequate to support this use case.

Materials and Methods

The latest LOINC database release Version 2.27 was selected as the basis for this evaluation. This version contains 55,058 terms. We first searched the LOINC database using RELMA (a mapping and browsing tool provided with the LOINC database) to retrieve genetic related LOINC concepts. We used the key word "MOLPATH" to select the relevant content. "MOLPATH" represents Molecular Pathology, the class under which genetic related LOINC terms are grouped. To confirm the search results, we also searched the LOINC table directly. The LOINC table was filtered using "MOLPATH" and any of its subclasses as the filter values for the "class" column. The subclasses of MOLPATH are "MOLPATH.MUT", "MOLPATH.DEL", "MOLPATH.TRISOMY", "MOLPATH.TRNLOC", "MOLPATH.TRINUC", "MOLPATH.REARRANGE", "MOLPATH.GENERAL", and "MOLPATH.MISC". The "class" filter was also used to select three additional classes: "PANEL.MOLPATH", "HL7.GENETICS", and "PANEL.HL7.GENETICS". The same number of LOINC terms was returned from the filter results as from the original RELMA query. We then manually went through each of the genetic LOINC concepts to select the ones that are specifically for cytogenetic testing.

To evaluate whether the current LOINC terminology is sufficient to represent constitutional cytogenetic test names and their results, we tried to represent a list of key data elements found in cytogenetic result reports by using the existing LOINC concepts. We obtained sample constitutional cytogenetic test result reports from the Cytogenetics Section of ARUP Laboratories. ARUP is a national clinical and anatomic pathology reference laboratory owned by the University of Utah.[6] The sample result reports were chosen so they would cover tests that were done using different cytogenetic techniques including: conventional G-banding, fluorescence in situ hybridization (FISH), and microarray based comparative genomic hybridization (array-CGH). The sample reports also represented a variety of results, including normal, abnormal, and findings of unknown

clinical significance. We examined these sample result reports and extracted a list of key data elements that should be coded. We also obtained the names of constitutional cytogenetic tests offered by ARUP from its online test menu.

Results

Table 1 shows the list of key data elements extracted from the constitutional cytogenetic test result reports that should be coded. We did not include some standard data elements in lab result reports, such as patient date of birth, sex, the specimen type, specimen collection date, reason for referral, etc. These elements should be sent in other fields in the HL7 message, and should not be sent as test results in the observation segment using LOINC codes.

Data Element
Test Performed
Chromosome Result (expressed in ISCN)
FISH Result (expressed in ISCN)
Array-CGH Result (expressed in ISCN)
Number of cells counted
Number of colonies counted
Number of cells analyzed
Number of cells karyotyped
ISCN Band Level
Banding Method
Copy number change
Chromosome bands involved
Base pair coordinates

Table 1. Key data elements in constitutional cytogenetic test result reports.

The constitutional cytogenetic tests offered by ARUP are listed in table 2.

Test #	Test Name
0097779	Prenatal FISH (Chromosomes X, Y, 13, 18 &21)
0097615	Chromosome Analysis, FISH-Metaphase
0092615	Chromosome Analysis, FISH-Interphase
0040201	Genomic Microarray, U-Array Chip
0097640	Chromosome Analysis, Peripheral Blood
0097601	Chromosome Analysis, Amniotic Fluid
0097610	Chromosome Analysis, Chorionic Villus Sampling (CVS)
0097620	Chromosome Analysis, Fetal Blood (PUBS)
0097645	Chromosome Analysis, Products of Conception (POC)
0097655	Chromosome Analysis, Skin Biopsy
0097650	Rule Out Mosaicism

Table 2. Constitutional cytogenetic tests offered by ARUP.

A total of 1047 genetic related LOINC terms were found in the database, 14 of them are inactivated terms with status of "DEL". Among the 1033 active terms, the majority was related to mutation analysis; only 38 terms were cytogenetic test related concepts. The first part of the LOINC name is the component or analyte measured. Table 3 lists the 20 distinct LOINC components from the 38 LOINC names. Some of the components were used in several LOINC names in combination with different systems, properties, scales, or methods.

LOINC Component
18q chromosome deletion
19q chromosome deletion
1p chromosome deletion
20q chromosome deletion
Chromosome 12 trisomy
Chromosome 12p tetrasomy
Chromosome 21 trisomy
Chromosome 7 trisomy
Chromosome 8 trisomy
Chromosome 9 trisomy
Chromosome analysis.interphase
Chromosome number
Chromosome region
Clinical cytogeneticist
Karyotype
Maternal cell contamination
Microdeletion syndromes
Subtelomere analysis
Telomere analysis
Y chromosome deletion

Table 3. Distinct LOINC components from the 38 existing cytogenetic test related concepts.

We found that the current LOINC terms for cytogenetic tests are not consistent with how the ARUP cytogenetic tests are named or with how the results are represented in actual reports. The existing LOINC terms are not consistent with the vocabulary needed to represent ARUP cytogenetic test names and results.

To report a chromosome analysis result for a male with Trisomy 21 (Down syndrome), the ARUP result report includes "Chromosome Analysis, Peripheral Blood" as the test name. This test name could be

mapped to the LOINC code "Karyotype:Prid:Pt:Bld/Tiss:Nar". For the test result, ARUP reports it as "47,XY,+21", which is the ISCN representation for male, Trisomy 21. The existing LOINC codes do not support this reporting style. Instead, they attempted to pre-coordinate the findings into the result names, e.g. *Chromosome 21 trisomy:Arb:Pt:Bld/Tiss:Ord:Cytogenetics*. This style of pre-coordination implies that the value of the result for this test as named by LOINC would be "Present" or "Absent."

For FISH studies, LOINC codes exist for Chromosome analysis, FISH-Interphase, but no codes exist for Chromosome Analysis, FISH-Metaphase. No codes are currently available to properly represent the results for any of the common microdeletion syndromes using either the LOINC variable approach or the panel approach. For example, consider DiGeorge/Velco-Cardio-Facial syndrome with the ISCN representation "ish del(22)(q11.2q11.22)(HIRA-)". To represent this finding using a panel approach, we would need a LOINC code that pre-coordinates the 22q11.2 deletion into the LOINC name. To represent it using the variable approach, a LOINC term like "chromosome analysis FISH result" would need to be created.

No LOINC codes currently exist to represent the array-CGH tests and their results.

Discussion

The LOINC six-part structure itself is capable of making codes for cytogenetic test names and observation names. However, the number of terms in the latest LOINC release for genetic test observations, especially cytogenetic tests, is minimal. We suspect that the existing LOINC terms are not being used in production systems because the existing LOINC terms and what is being reported from ARUP imply very different models of representation. These terms do not match well with how the tests are named and how the test results are reported.

Recognizing the importance of genetic test result reporting, the LOINC committee recently began developing terms for representing genetic variations. However, there is no specific section in the LOINC Reference Manual that discusses names and codes for cytogenetic tests. We plan to propose developing the needed cytogenetic codes in partnership with the LOINC committee.

Pre-coordination vs. post-coordination

The majority of existing LOINC terms for cytogenetic tests are taking the pre-coordination approach. The current style of LOINC terms seems to have been created to ask questions like whether a given abnormality is found, e.g. 18q chromosome deletion:Prid:Pt:Bld/Tiss:Nom:Molgen, with the expected answers being "Present" or "Absent". Continuing this style of LOINC name creation will be problematic, not only for the representation of cytogenetic test results but also for the representation of genetic test results in general. Due to the ever growing and changing nature of this field, this pre-coordinated style of name creation will likely lead to a large and potentially limitless number of test names being created. For example, the U-Array Chip that ARUP currently uses for its array-CGH test contains close to 150 targeted regions and this number will continue to grow as higher density chips come into practice. In order to avoid combinatorial explosion, a post-coordinated style would be more appropriate for creating LOINC concepts for genetic testing as it will be more sustainable and flexible.

ISCN and Coded Expression Data Type

Compared to molecular genetic tests results, the advantage that cytogenetic test result reporting has is that ISCN has been the gold standard for describing chromosome aberrations for almost 40 years. ISCN provides a list of symbols and abbreviated terms in adjunction with a set of rules, which can be used in the description of chromosomes and chromosome abnormalities, such as *p* for short arm of chromosome, *q* for long arm of chromosome, *cen* for centromere, *del* for deletion, *ish* for in situ hybridization, and plus sign (+) for gain, etc.

Data that are expressed in ISCN nomenclature need to be distinguished from either string values or concepts from a code system. Typical behaviors that are expected for coded concepts do not apply to ISCN expressions. This situation is the use case that would justify a new "coded expression" data type for use in HL7 messages. It might also suggest the need for a new type of scale in the LOINC terminology. When receiving systems encounter coded expressions, tools will need to parse the data rather than to do terminology look ups. This would also imply the need for a new query engine that could query against the ISCN expressions. For example, as new knowledge becomes available it would be desirable to run a query to identify all patients who have chromosome abnormalities that were believed to be clinically insignificant or that have unknown clinical significance at the time of testing where a revised report should be issued. The results review

applications will also need to be able to present this new type of data rather than treating them the same as simple name-value pairs.

Array-CGH

Array-CGH merges molecular diagnostics with traditional chromosome analysis and is transforming the field of cytogenetics. Array-CGH holds the promise of being the initial diagnostic tool in the identification of visible and submicroscopic chromosome abnormalities in mental retardation and other developmental disabilities.[7-8] Therefore, clinical information systems should anticipate receiving more array-CGH results in the very near future. The LOINC standard should examine this rapidly growing area and develop codes for microarray based laboratory tests.

Terminology and Information Models

The LOINC terminology without the context of an information model is not sufficient to unambiguously exchange cytogenetic test results. The LOINC codes need to be developed in the context of an information model, which is similar to putting vocabulary terms into meaningful sentence structures. In addition to the LOINC standard, other bioinformatics standard terminologies such as ISCN are necessary to represent the detailed results of cytogenetic tests.

Limitations

Our evaluation may be limited due to the fact that there is lack of industry wide cytogenetic result report standards available. As a consequence our analysis is based on sample result reports from ARUP only. However, because ARUP result reports contain all the data elements listed on the CAP checklist (which represents the industry standard), this limitation is likely minimal.

Another limitation is that the list of key data elements that we included for analysis is not complete. We did not extract data elements from the free text sections of the report such as the "diagnostic impression" and "recommendation" sections of the reports. This means that our evaluation of current reporting limitations is likely conservative.

Conclusion and Future Work

Current LOINC terminology is insufficient to support the needs of coding cytogenetic test results. With genetic testing becoming an increasingly important part of the daily medical practice, we need to develop this essential infrastructure before clinical information systems will be able to handle high volumes of genetics data.

This study was an initial step in integrating cytogenetic test result reports into EMRs. It demonstrated that a gap exists in LOINC in supporting such integration. Work needs to be done to extend LOINC to cover cytogenetic tests and to continue to expand the codes needed for the broader field of genetic variation testing. Since it is the CHI designated standard for laboratory tests, we suggest enhancing and extending LOINC to represent cytogenetics test result reports rather than creating them in some other existing terminology.

Further analysis needs to be done to develop new LOINC codes and information models to represent the constitutional cytogenetic test result reports. The analysis needs to be expanded to include result reports from other laboratories besides ARUP. Structuring the diagnostic impression and recommendation section of the result report needs to be addressed as well. Our hope is that this will lead to consistency in reporting results, in addition to simplifying access to and understanding of interpretation of those results.

Acknowledgements: We would like to thank Drs. Scott Narus and Peter Haug for reviewing the article, and Cori Nigh for her technical assistance in obtaining ARUP sample result reports.

References

1. http://www.genetests.org
2. Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases a systematic review. JAMA. 2008;299(11)1320-1334.
3. Harvey EK et al. Providers' knowledge of genetics: A survey of 5915 individuals and families with genetic conditions. Genet Med. 2007 May;9(5):259-67.
4. Lubin IM et al. Clinician perspectives about molecular genetic testing for heritable conditions and development of a clinician-friendly laboratory report. J Mol Diagn. 2009 Mar;11(2):162-71
5. http://www.cap.org
6. http://www.aruplab.com
7. Aston E et al. Comparison of targeted and whole genome analysis of postnatal specimens using a commercially available array based comparative genomic hybridization (aCGH) microarray platform. J Med Genet. 2008 May;45(5):268-74.
8. Baldwin et al. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. Genet Med. 2008 Jun;10(6):415-429.