

HL7 VERSION 3
DOMAIN ANALYSIS MODEL:
CLINICAL SEQUENCING, RELEASE 1
DRAFT
(1ST INFORMATIVE BALLOT)

Initial version: January 2013

Next Balloted Version: September 2014

Chapter Chair and Principal Author:	Mollie Ullman-Cullere Dana-Farber Cancer Institute and Partners HealthCare
Chapter Chair and Contributing Author:	Amnon Shabo (Shvo) Standards of Health
Project Chair and Contributing Author:	Grant Wood Intermountain Healthcare
Contributing Author:	Kevin Hughes, MD Massachusetts General Hospital
Contributing Author:	Daryl Thomas Life Technologies
Contributing Author	Larry Babb Partners Healthcare Center for Personalized Genetic Medicine
Contributing Author	Lynn Bry, MD Brigham and Women's Hospital
Seeking Additional Co-Authors/Participants	

The HL7 Clinical Genomics Work Group is actively seeking comments/feedback from the genetics/genomics community. Please contact Mollie Ullman-Cullere at: mullmancullere@partners.org, if you are interested in participating.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
1.1 Purpose.....	1
1.2 Audience.....	1
1.3 Scope	1
1.4 Assumptions	2
2. USE CASE STAKEHOLDERS	3
3. ISSUES AND OBSTACLES	4
4. PERSPECTIVE.....	4
5. USE CASE SCENARIOS.....	5
5.1 Scenario 1: Specimen Identification.....	5
5.1.1 Germline testing for biomarkers/mutations (usually inherited)	5
5.1.2 Tumor testing for somatic (tumor specific biomarkers/mutations)	5
5.1.3 Pediatric testing for biomarkers/mutations causal to rare early childhood conditions.....	6
5.1.4 Prenatal testing which may be reported on the maternal medical record (and should be identified as separate from germline testing)	6
5.1.5 Infectious disease testing, where the biomarker/mutation identified within the disease causing organism is reported into the patient medical record following similar data standards as used for other testing scenarios above.	6
5.2 Scenario 2: Clinical Sequencing - Germline Testing	7
5.2.1 Description of Scenario (following numbers in the diagram above)	8
5.2.2 Alternative Flow 1: Chart Review.....	9
5.2.3 Alternative Flow 2: New Genetic Knowledge	9
5.2.4 Alternative Flow 3: New Clinical Indication.....	9
5.3 Scenario 3: Cancer Profiling - Somatic Testing	10
5.3.1 Description of Scenario Differences from Germline Workflow	10
5.4 Scenario 4: Decision Making Tools - Family History and Drug Dosage Calculators.....	11
5.4.1 Description of Scenario	11
5.5 Scenario 5: Public Health Reporting.....	12
5.5.1 Description of Scenario	12
5.6 Scenario 6: Clinical and Research Data Warehouses.....	13
5.6.1 Description of Scenario	13
6. ADDITIONAL USE CASES <RELEASE 1 OR 2?>	14
6.1 State & Regional HIE.....	14

Table of Contents

6.2 National Marrow Donor Program.....	14
6.3 Cancer Registry workflow.....	14
6.4 Public Health Testing - microbial.....	14
6.5 Newborn Screening.....	14
6.6 commercial testing laboratories.....	15
6.6.1 <i>Defined Genetic Testing vs. Expanding Genetic Tests</i>	15
6.7 patient panel management- analytics for care quality.....	15
6.8 Patient genetic profile - data across all testing paltforms.....	15
6.9 FDA Scenarios in Public Health Reporting.....	15
6.10 Additional variant types.....	15
6.10.1 <i>Structural variants</i>	15
6.10.1.0 Currently using ISCN standards and stored at NCBI in dbVAR.....	15
6.10.2 <i>Copy number change</i>	15
6.10.2.0 Emerging standards with the following suggestions:.....	15
6.10.3 <i>Biomarkers --> Is this far enough along to add</i>	15
6.10.3.0 <Add MedGen/LOINC>.....	15
6.11 Laboratory genomic data standards.....	16
6.12 Extension of sequence VARIATION AND cytogenetic HL7 models.....	16
7. DATA SET CONSIDERATIONS & STANDARDS FROM THE FIELD.....	16
7.1.1 <i>Genes</i>	16
7.1.1.0 HGNC gene symbols (required).....	16
7.1.2 <i>Sequence Variations</i>	16
7.1.2.0 HGVs (required).....	16
7.1.2.1 dbSNP (optional, but highly recommended).....	17
7.1.2.2 COSMIC (optional).....	17
7.1.3 <i>Reference Sequences (required)</i>	18
7.1.3.0 RefSeq.....	18
7.1.3.1 LRG	18
7.1.4 <i>Publicly Available References (valuable for clinical and translational genomics)</i>	19
7.1.4.0 OMIM (optional).....	19
7.1.4.1 PubMed (optional).....	19
7.1.4.2 PharmGKB (optional).....	19
7.1.4.3 ClinicalTrials.gov (optional).....	20
8. VOCABULARY CONSTRAINTS.....	21
9. REVIEW OF EXISTING HL7 CLINICAL GENOMICS SPECIFICATIONS.....	22

Table of Contents

9.1 HL7 V2 Genetic Test result message.....	22
9.2 HL7 CDA Implementaion Guide for Genetic testing reports.....	22
10. HIT DATA STANDARDS.....	23
10.1 FAMILY HISTORY.....	23
10.2 Sequence Variations / Chromosomal change.....	23
10.2.1 Small Genetic Variations within a Gene	23
10.2.2 Structural Variations	23
11. HL7 ENCAPSULATION OF GENOMIC DATA FILES.....	23
12. CLINICAL GRADE-GENOMIC DATA FILE STANDARDS	23
13. GAPS & EXTENSIONS.....	24
13.1 Laboratory order entry.....	24
14. OUTSTANDING QUESTIONS	24
15. GLOSSARY	24
15.1 Extension to Specimen scenarios.....	24
15.1.1 Microbiome analysis of the patient.....	24

1. Introduction

In March, 2008, the United States Department of Health and Human Services, Office of the National Coordinator for Health IT published the *Personalized Healthcare Detailed Use Case* <add reference to publication <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2442266/>> in response to a request and specifications from the American Health Information Community. The use case focuses on supporting secure access to electronic genetic laboratory results and interpretations for clinical care, as well as family history and associated risk assessments by authorized parties and is driven by the need for timely electronic access to ordered, referred and historical genetic laboratory results and family history. Ordering clinicians receive genetic laboratory test results as a response to an order by having the genetic test results sent either directly to the clinician's EHR system (local or remote) or to another clinical data system in support of the provisioning of historical results.

Members of the HL7 Clinical Genomics work group participated in the ONC use case development and in parallel extended HL7 messaging standards and wrote implementation guides to support the described scenarios.

Family History

- Pedigree – Family History
- IG for Family History

Clinical Genetic Testing

- IG for Genetic Variants 2.5.1
- CDA – GTR v3
- IG for Cytogenetics

Much has changed since 2008 and much remains the same. The *HL7 Version 3 Domain Analysis Model: Clinical Sequencing, Release 1* catalogs the breadth of genetic/genomic testing use cases and clinical scenarios, discusses current challenges and lessons learned, and raises questions to consider for future implementations. While this document discusses the use of new technology (Next Generation Sequencing (NGS)), it must be remembered that the vast majority of clinical genetic testing is still performed on testing platforms in use ten years ago, and it is the goal of the Clinical Genomics work group to facilitate platform- independent, interoperability of genetic/genomic data.

1.1 PURPOSE

The *HL7 Version 3 Domain Analysis Model: Clinical Sequencing, Release 1* should be used to inform standards developers and implementers, for the design scalable, interoperable solutions covering the breadth of clinical scenarios.

1.2 AUDIENCE

This guide is designed to be used by analysts and developers who require guidance on incorporation of genomic data in the clinical and clinical research healthcare IT environment. In addition, developers of genomic and healthcare IT data standards may use this guide to extend these standards for support of clinical sequencing. Users of this guide must be familiar with the details of HL7 message construction and processing. This guide is not intended to be a tutorial on that subject.

1.3 SCOPE

This domain analysis model details a variety of use case scenarios key to personalized genomic medicine and translational research, including more typical scenario for testing of a person's inherited or germline genome, cancer genomics/tumor profiling, early childhood developmental delay, neonatal testing, and

Chapter 1: Introduction

newborn screening. In addition, the use case includes two scenarios where test results are manually translated from reports into either a tool for clinical decision making (e.g. family history or drug dosage calculator) or for public health reporting for cancer registries.

1.4 ASSUMPTIONS

Assumptions are summarized as follows:

- Infrastructure is in place to allow accurate information exchange between information systems.
- Providers access laboratory test results through either an EHR or a clinical information system.
- Privacy and security has been implemented at an acceptable level.
- All participants agree to all standards, methodologies, consent, privacy and security.
- Legal and governance issues regarding data access authorizations, data ownership and data use are outside the scope of this document.
- The order, paper or electronic, associated with the laboratory result contains sufficient information for the laboratory to construct the laboratory result message properly.

2. Use Case Stakeholders

Stakeholder	Contextual Description
Anatomic & Surgical Pathology	For cancer profiling (i.e. genetic testing of cancer specimens), the pathologic diagnosis will play a key role in testing and interpretation of the findings.
Geneticist / Medical Geneticist / Molecular Pathologist	Professionals interpreting the clinical implications of a patient's genetic data. These professionals may work within the laboratory setting or outside the laboratory.
Healthcare Entities	Organizations delivering healthcare.
Healthcare Payors	Healthcare Insurers and Centers for Medicare & Medicaid Services
Information Technology Vendors	Vendors supplying information technology solutions and support.
Laboratories - Reference	Testing laboratories outside the hospital environment either as a separate corporate entity or separate unit of the same organization.
Laboratories - Hospital	Testing laboratory which is part of the hospital entity and hospital laboratories.
Manufacturers/Distributors	Entities involved in the development, production, and distribution of products used in healthcare (e.g. <i>in vitro</i> diagnostic tests)
Patients	Members of the public that use healthcare services.
Public Health Agencies	Agencies which help to protect and improve health and healthcare of the public.
Registries	Systems for the collection, analysis, and distribution of data for the improvement of public health.

Chapter 5: Use Case Scenarios

Internal and External Systems for Stakeholder Groups			
Patient Systems	Healthcare Provider Systems	Laboratory Systems	Geneticist...Systems
<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">PHR (Patient Controlled)</div> <div style="border: 1px solid black; padding: 5px;">Patient Portal (Part of EHR)</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">EHR: medical record, lab order entry, lab/pathology results, clinical guidelines, e-prescription, image repository, clinical decision support systems, specialized clinical modules, family history analysis tools</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Genetic Data Repository (containing patient variome or sequence)</div> <div style="border: 1px solid black; padding: 5px;">Genetic Knowledgebase</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">NCBI Databases: RefSeq, dbSNP, dbVAR, OMIM, PubMed, Genetics Home Reference, GeneTests.org,</div> <div style="border: 1px solid black; padding: 5px;">Lab Information Management System, BioInformatic Pipeline, results database</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">NCBI Databases: RefSeq, dbSNP, dbVAR, OMIM, PubMed, Genetics Home Reference, GeneTests.org,</div> <div style="border: 1px solid black; padding: 5px;">Genetic Databases/ Knowledgebase Gene, Disease, Ethnicity (germline & somatic), family history analysis tools</div>

3. Issues and Obstacles

Numerous challenges exist in the area of policy, patient and clinician education, and reimbursement, which are beyond the scope of this document, unless requiring consideration within the information technology solutions (e.g. clinical decision support). Key challenges for information technology include: data security, adoption of electronic health records and laboratory information management systems, and interoperability, and structuring of useful data. This document informs information technology vendors of key functionality for clinical sequencing, and outlines considerations for healthcare providers and laboratories investing in information technology.

4. Perspective

This document includes perspectives of stakeholder groups outlined in section 2. Integration of molecular diagnostics into the clinical workflow is key for safe, efficient and effective adoption. For instance, the potential for medical error during drug order entry is reduced with clinical decision support which alerts the clinician, if ordering a drug which is contraindicated. Developing systems which are capable of consideration of genetic markers associated with drug metabolism, efficacy, and toxicity during the order entry process will reduce medial error, as our knowledge increases.

5. Use Case Scenarios

5.1 SCENARIO 1: SPECIMEN IDENTIFICATION

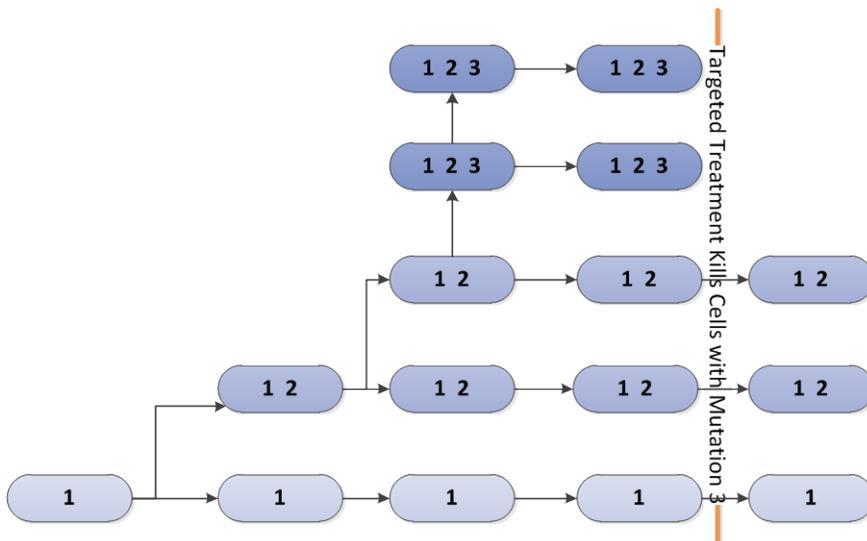
Use Cases for sequencing require explicate identification of 1 or more specimens to be used in laboratory analysis. This likely requires the identification of specimen groups (i.e. separate specimens and associated derivatives) originating from the same patient/subject or related patients/subjects.

5.1.1 Germline testing for biomarkers/mutations (usually inherited)

In terms of specimen identification, this is the most straightforward scenario. Typically a blood sample or cheek swab will be taken from the patient and DNA extracted. Except for low level heterogeneity, the genome/variome/mutations identified in this specimen are ubiquitously present throughout every cell in the patient and are inherited from their mother and father (except in the case of spontaneous mutations). This specimen is not limited in quantity, like a tumor specimen, because the laboratory may request an additional sample.

5.1.2 Tumor testing for somatic (tumor specific biomarkers/mutations)

To identify somatic (i.e. acquired) mutations within a cancer specimen, in general a laboratory will analyze both a germline specimen and somatic specimen. The somatic/cancer specimen contains both germline sequence and mutations as well as the somatic mutations present in cancer. To definitively classify a mutation as somatic the laboratory compares the two sequences and to identify mutations unique to the cancer. Note this can be a complicated process, because cancer cells acquire mutations throughout their lifespan and pass them on to daughter cells.



Simplified representation of cancer cells acquiring mutations or sequence variants, represented as numbers 1 2 and 3, in dividing cancer cells. Note targeted therapy can kill a specific population of cancer cells.

Chapter 5: Use Case Scenarios

Changes in the population of cells with particular mutations will change overtime as well as in conjunction with events such as therapy. For instance, targeted chemotherapy may kill a specific population of cancer cells with specific mutations and other cancer cell populations may survive and continue to divide. Therefore, clearly annotating these specimens as somatic and capturing annotations related to a time relevant to a treatment timeline may be critical for analysis. [In order to explicitly represent these annotations, it is important to be able to associate all data elements into a coherent clinical genomics statement, as described in the Domain Information Model document.](#)

In some scenarios, a laboratory may focus sequence analysis on well studied genes/mutations identified only in cancer. Commonly these mutations are only found in cancer, because they cause extreme behavioral changes at the cellular level (e.g. uncontrolled cell division), which would result in embryonic death if present in the embryo. Specimens, sequence, and identified variants/mutations from these studies should be clearly annotated as somatic.

Summary

- a. Matched specimens for germline and somatic analysis, where comparison will result in the identification of tumor specific mutations/biomarkers
- b. Tumor specimen without a matched germline specimen, where mutations/biomarkers are believed to be specific to tumors.

5.1.3 Pediatric testing for biomarkers/mutations causal to rare early childhood conditions

- a. Matched specimens of patient and maternal and paternal specimens, where comparison aids in identification of original biomarkers/mutations within the patient

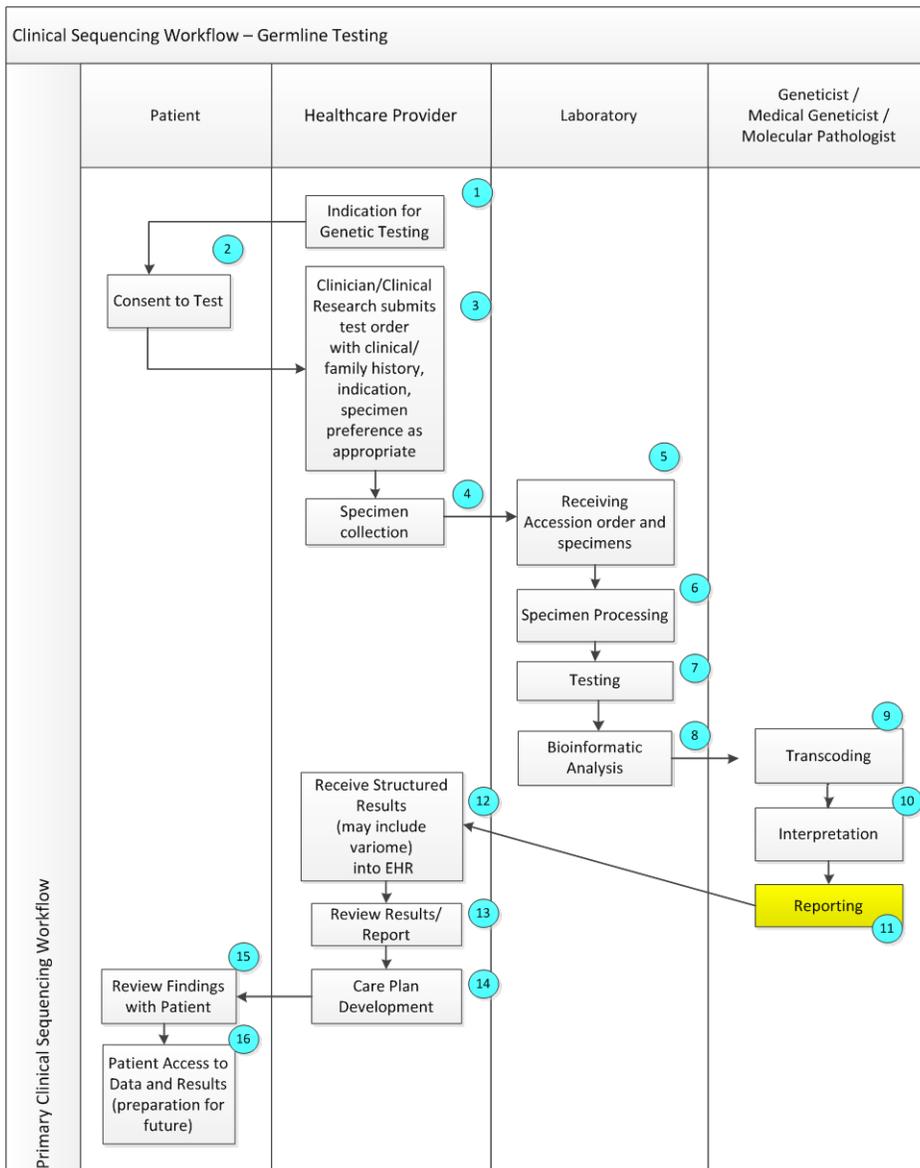
5.1.4 Prenatal testing which may be reported on the maternal medical record (and should be identified as separate from germline testing)

- a. Often have matched prenatal/fetal and maternal specimens for analysis

5.1.5 Infectious disease testing, where the biomarker/mutation identified within the disease causing organism is reported into the patient medical record following similar data standards as used for other testing scenarios above.

Derivatives which may be analyzed from the above testing scenarios include: DNA, RNA, and Protein

5.2 SCENARIO 2: CLINICAL SEQUENCING – GERMLINE TESTING

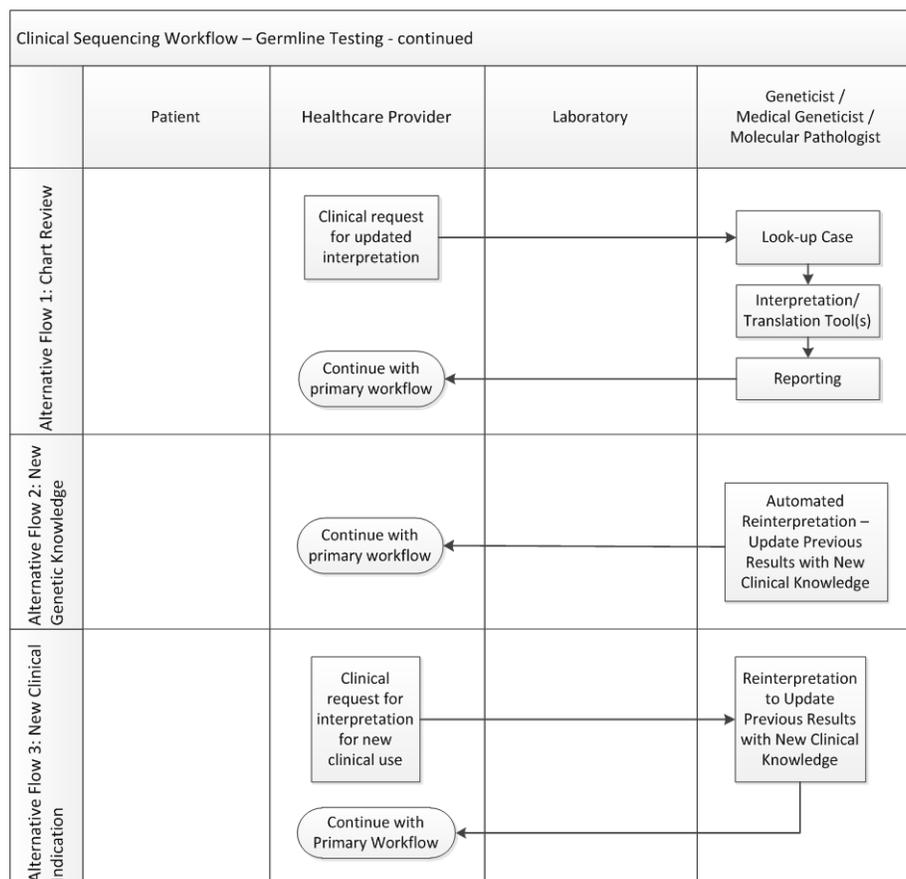


5.2.1 Description of Scenario (following numbers in the diagram above)

1. Clinician determines that a genetic test is needed to inform patient care decisions. Often this includes family history based risk assessment.
2. Clinician obtains patient consent for testing.
3. Order entry for genetic testing, including relevant data to aid in evaluation and interpretation of findings: indication for testing, family history, and relevant clinical data for the patient.
4. Blood is drawn or cheek swabbed for cells containing DNA
5. Laboratory receives the order and specimen(s) for testing
6. Specimens are processed (e.g. DNA extracted) and prepared to be loaded on the sequencing instrument.
7. Specimens are sequenced.
8. Data from the instrument passes through a bioinformatics pipeline for data processing: alignment and identification of sequence variants, as well as quality assurance
9. During the 'Transcoding' process, raw genetic-genomic data is transformed from bioinformatics format into healthcare IT data standards.
 - 9-a. Alternatively, key chunks of the raw genomic data are encapsulated in healthcare standards in their native bioinformatics formats, and only some of these key data sets are transcoded into healthcare standards in order to be better processed by clinical decision support application, as well as be associated with phenotypic data.
10. Genetic results are interpreted for clinical implications
11. Genetic report is created, including narrative findings and interpretation as well as the equivalent information structured in machine readable formats using interoperable healthcare IT data standards.
12. Genetic report and structured results are received in the Electronic Health Record system (EHR_S)
13. Clinician reviews the results/report
14. Clinician develops (or modifies) a care plan taking into consideration the genetic findings
15. Clinician reviews the genetic findings and care plan with the patient
16. Genetic results are made available to the patient in the web-based patient portal

Formatted

Chapter 5: Use Case Scenarios



5.2.2 Alternative Flow 1: Chart Review

If a sequence variant (i.e. mutation) of 'Unknown Significance' were identified in a patient or the clinical implications of an identified variant are suspected of change, then the clinician may contact the testing laboratory prior to a follow-up patient appointment (e.g. annual exam).

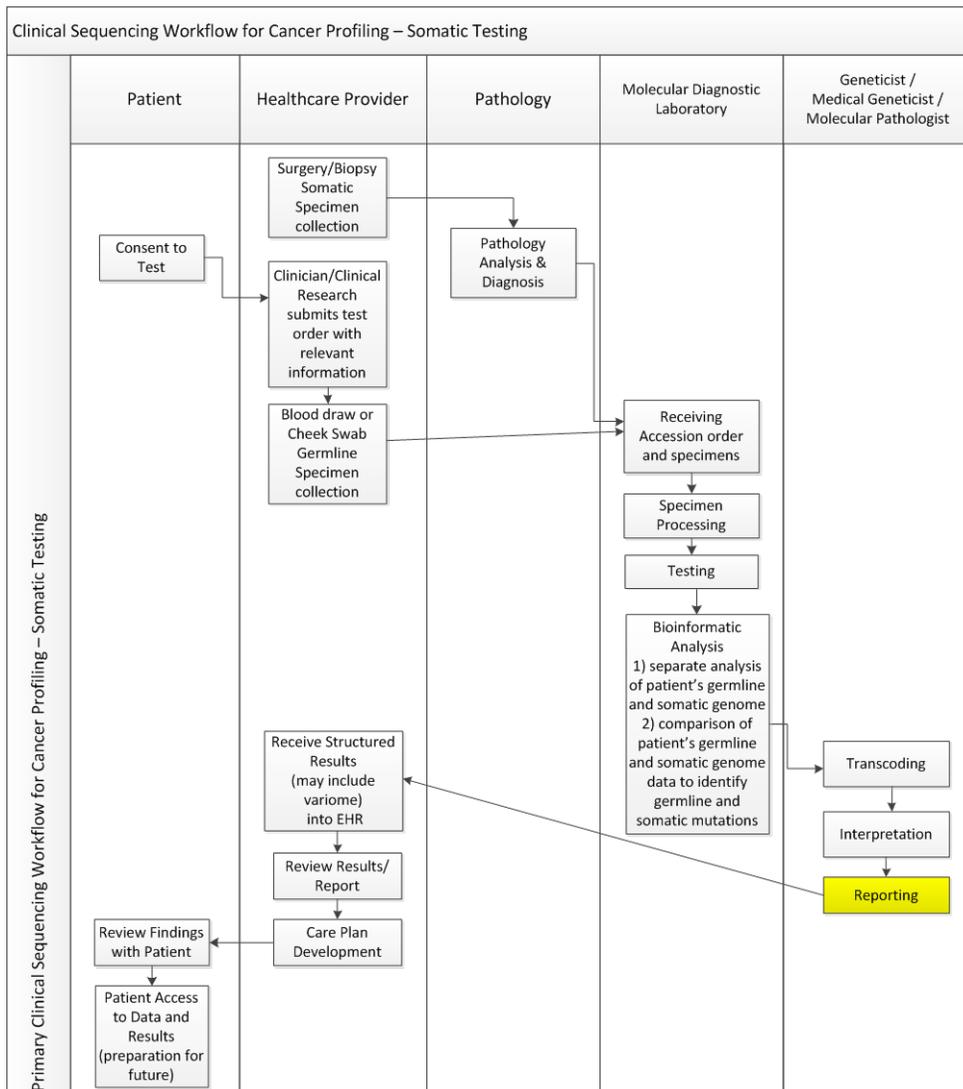
5.2.3 Alternative Flow 2: New Genetic Knowledge

A testing laboratory may contact the ordering clinician, if the clinical implications of a sequence variant (i.e. mutation), previously identified in the patient, have changed.

5.2.4 Alternative Flow 3: New Clinical Indication

If genetic data from previous testing may inform a new clinical decision, the clinician may contact the laboratory for a new interpretation of existing data. As confidence in data quality increases and size of data sets increases, alternative flow may become more common.

5.3 SCENARIO 3: CANCER PROFILING – SOMATIC TESTING

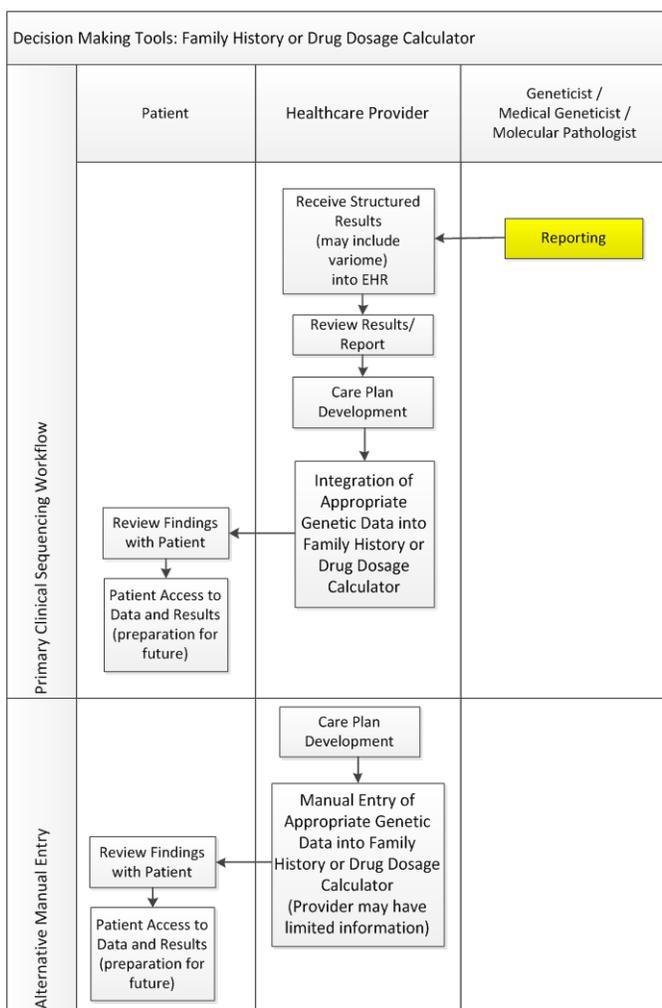


5.3.1 Description of Scenario Differences from Germline Workflow

In cancer profiling, pathology plays a key role. For instance, the same mutation identified in different cancers has different clinical implications. In addition, ideally clinical sequencing will include analysis of

both a germline specimen and a cancer specimen, so that cancer specific mutations can be identified with more certainty. For more information on specimens within this workflow, see section 5.1.2.

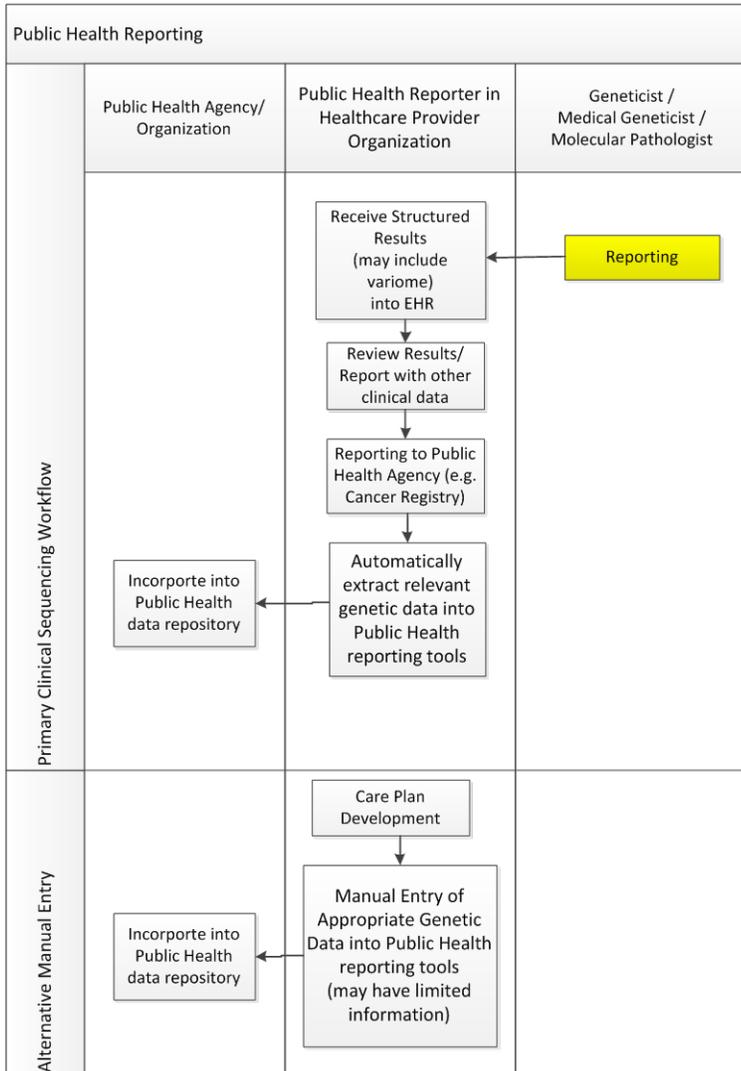
5.4 SCENARIO 4: DECISION MAKING TOOLS – FAMILY HISTORY AND DRUG DOSAGE CALCULATORS



5.4.1 Description of Scenario

Today clinicians translate (i.e. manually reenter) genetic data into tools for decision making. This includes family history tools and drug dosage calculators. In the future, this data will automatically be incorporated into clinical decision making tools.

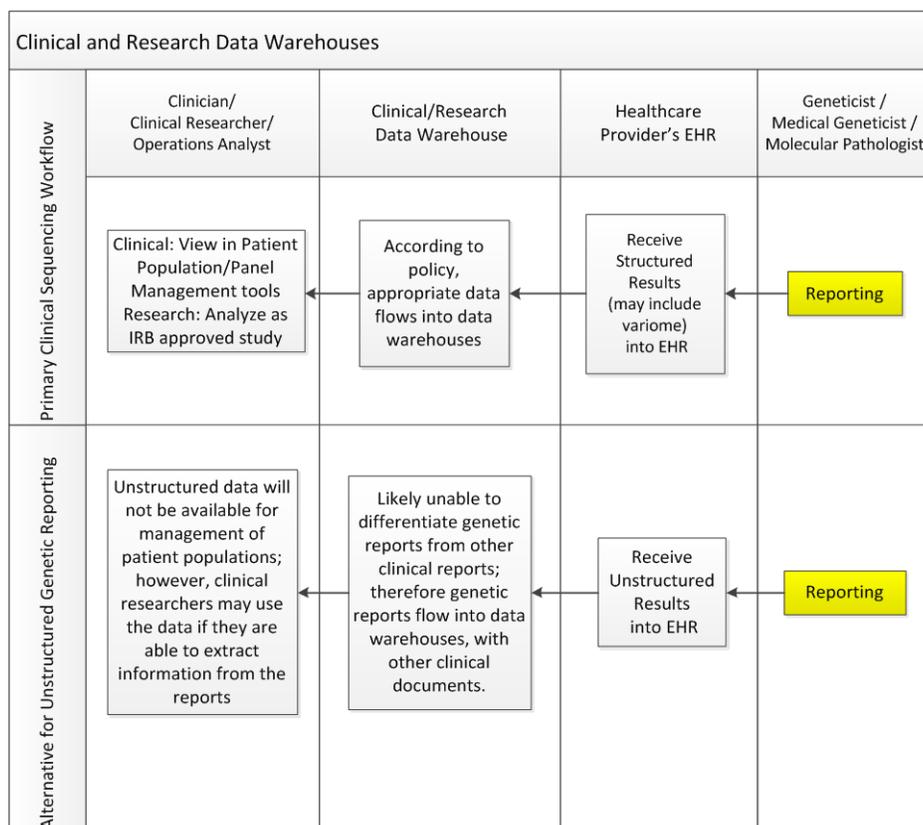
5.5 SCENARIO 5: PUBLIC HEALTH REPORTING



5.5.1 Description of Scenario

Today Registrars manually translate clinical data into public health reporting systems. This data is used to monitor and improve public health (e.g. surveillance and clinical research). In the future, this data will be extracted from the EHR in an automated (or semi-automated) manner.

5.6 SCENARIO 6: CLINICAL AND RESEARCH DATA WAREHOUSES



5.6.1 Description of Scenario

Electronic health records [systems](#) (EHR-[Ss](#)) are optimized for transactional data and working with one patient record at a time. To enable clinicians to view populations of similar patients (e.g. a primary care provider may want to see last mammography dates for all their patients with increased risk of breast cancer), clinical data is incorporated into clinical data warehouses. Similar data warehouses support use of clinical data, for clinical research, according to Institutional Review Board policies. If genetic data is not structured, it doesn't support these activities.

[Health data warehousing should persist data in its standardized formats, while allowing users to export subsets of the data in the warehouse into multiple 'data marts', optimized for specific use cases, analysis type or reporting needs. Warehouse data should be represented in the richest form possible using generic standards, while each data mart is optimized for specific use case, e.g., clinical research, public health registrars, or even EHR systems. In this way, all different 'views' of the data are based on the same standardized semantics, thus achieving consistency and interoperability while avoiding lossy/ousy transformations and duplication of mass data.](#)

6. Additional use cases <release 1 or 2?>

6.1 STATE & REGIONAL HIE

Work with VA HIE - HITSAC

6.2 NATIONAL MARROW DONOR PROGRAM

Work with Bob Millus

Challenges:

- Genomic regions of interest are not included within a genome build; therefore, using a genome build and chromosome in conjunction with genomic location does not support HLA typing
- Clinical genetic standards for communicating a variant (e.g. HGVS) do not support the complexity of HLA typing; therefore the marrow donor program has developed their own standard.
- Marrow donor nomenclature is based on allele naming and continues to evolve as more is understood and technology platforms are capable of more detailed detection.
- Systems must support different versions of the marrow donor nomenclature and various degrees of ambiguity.

Formatted: List Paragraph, Bulleted + Level: 1 + Aligned at: 0.25" + Indent at: 0.5"

6.3 CANCER REGISTRY WORKFLOW

Cancer registrars perform patient chart review translating and summarizing clinical information into public health reporting systems.

Challenges:

- Genetic test results are inconsistently reported due to a number of factors
 - Lack of adherence to guidelines of medical professional organizations (e.g. CAP and ACMG)
 - Granularity of results are tied to testing platform and no known mapping exists to align levels of granularity. For example,
 - Kit based tests often do not output specific identified variants but roll these up into a biomarker (e.g. xxxxx)
 - ABI Sequencing is often reported in HGVS nomenclature at the c. and p. level. Current software makes it difficult to determine the genomic coordinates
 - Next Generation Sequencing (NGS) pipelines first identify variants in genomic coordinates. Translation of genomic coordinates into c., p. and biomarker representation is dependant on tools which are still immature. In addition, many of these tools are developed by groups with strong research backgrounds, and their understanding of clinical standards and practices is still evolving.
 - College of American Pathologists reporting templates currently report variants at the biomarker level without mapping between these other representations.

Formatted: Normal, No bullets or numbering

Formatted: Normal

Formatted: List Paragraph, Bulleted + Level: 1 + Aligned at: 0.25" + Indent at: 0.5"

Formatted: List Paragraph, Bulleted + Level: 2 + Aligned at: 0.75" + Indent at: 1"

Formatted: List Paragraph, Bulleted + Level: 3 + Aligned at: 1.25" + Indent at: 1.5"

6.4 PUBLIC HEALTH TESTING – MICROBIAL

Work with VA public health

6.5 NEWBORN SCREENING

Work with VA public health

6.6 COMMERCIAL TESTING LABORATORIES

6.6.1 Defined Genetic Testing vs. Expanding Genetic Tests

For example, a clinician may order the specific version of a cardiomyopathy test from lab A, which tests specific regions of specific genes for the presence of clinically relevant mutations. If new regions are found to be associated with cardiomyopathy, the patient's DNA will not be retested without a new clinical requisition. However, if the test is ordered from lab B, lab B will retest the patient's DNA as new genes/genetic regions are found to be associated with cardiomyopathy, thereby expanding the genetic test for cardiomyopathy in perpetuity.

6.7 PATIENT PANEL MANAGEMENT– ANALYTICS FOR CARE QUALITY

Work with HL7 QCI

6.8 PATIENT GENETIC PROFILE – DATA ACROSS ALL TESTING PLATFORMS

Challenges:

[See Cancer Registry workflow 6.3](#)

6.9 FDA SCENARIOS IN PUBLIC HEALTH REPORTING

Add workflow for next version.

Add workflow for next version.

6.10 ADDITIONAL VARIANT TYPES

6.10.1 Structural variants

6.10.1.0 Currently using ISCN standards and stored at NCBI in dbVAR.

6.10.2 Copy number change

6.10.2.0 Emerging standards with the following suggestions:

6.10.3 Biomarkers --> Is this far enough along to add

6.10.3.0 <Add MedGen/LOINC>

6.11 LABORATORY GENOMIC DATA STANDARDS

Identify and collaborate with stakeholders for laboratory genomic data standards, to ensure support for required annotations key to clinical processing and reporting (e.g. germline vs. somatic variants).

6.12 EXTENSION OF SEQUENCE VARIATION AND CYTOGENETIC HL7 MODELS

Current HL7 standards for sequence variation and cytogenetic findings use established clinical standards. These will be extended to support inclusion of established bioinformatic representation, to support linking to research and clinical information systems.

7.Data Set Considerations & Standards from the Field

Formatted: Not Highlight

~~Amnon — where field~~The following sub-sections list recommendations for specific nomenclatures (e.g. HGVS), field standards (e.g. reference sequences), and public repositories and knowledge bases ~~are recommended, they should be introduced in the DAM along~~ with a discussion on how to use them (e.g. dbSNP contains somatic and pathogenic variants not just polymorphisms). In addition, ~~OIDs registered at HL7 for these nomenclatures should be listed are listed~~ here.

7.1.1 Genes

7.1.1.0 **HGNC** gene symbols (required)

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	HGNC
OID	2.16.840.1.113883.6.281
Minimum attributes of the component	Gene symbol
Other Comments	Human Gene Nomenclature Committee (HGNC maintains a database of gene names and symbols. They are a non-profit body which is jointly funded by the US National Human Genome Research Institute (NHGRI) and the Wellcome Trust (UK). They operate under the auspices of Human Genome Organization . The database can be found at http://www.genenames.org/ .

7.1.2 Sequence Variations

7.1.2.0 **HGVS** (required)

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	HGVS
OID	2.16.840.1.113883.6.282
Minimum attributes of the component	Sequence variation

Chapter 5: Use Case Scenarios

Other Comments	Human Genome Variation Society (HGVS) Nomenclature standards for the description of sequence variations are maintained at http://www.hgvs.org/mutnomen/ . This standard is well accepted by the clinical genetic community and is extended on an ongoing basis to support genetic findings. HGVS maintains a tool for the generation/translation/verification of HGVS nomenclature. This tool can be found at: http://www.mutalyzer.nl/2.0/index
----------------	---

7.1.2.1 dbSNP (optional, but highly recommended)

TABLE 6-6 - DBSNP

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	dbSNP
OID	2.16.840.1.113883.6.284
Minimum attributes of the component	RS number and nucleotide change
Other Comments	The Single Nucleotide Polymorphism database (dbSNP). Is maintained by National Center for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov/projects/SNP/ Databases and knowledgebases defining sequence variants will be increasingly important. Although sequencing based tests which can result in the identification of novel variants require HGVS nomenclature standards for complete results reporting, genotyping tests which probe for the existence of known variants can additionally report results using an 'RS number' (i.e. identifier in dbSNP) and the associated nucleotide change. (Within the clinical environment results reporting using HGVS nomenclature is required with an option to additionally specify the RS number.)

7.1.2.2 COSMIC (optional)

Variants/Mutations can also be reported with a COSMIC mutation identifier associating the findings with internationally compiled cancer mutation data.

TABLE 6-10 - COSMIC

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	COSMIC (Catalogue Of Somatic Mutations In Cancer)
Responsible Body	Sanger Institute
OID	2.16.840.1.113883.3.912
Minimum attributes of the component	COSMIC ID
Other Comments	Catalogue Of Somatic Mutations In Cancer (COSMIC) serves as a repository for somatic mutations identified in specific cancer specimens. These mutations are recorded associated with structured description of the specimen.

Chapter 5: Use Case Scenarios

Available at: <http://www.sanger.ac.uk/genetics/CGP/cosmic/>

7.1.3 Reference Sequences (required)

Reference sequences are the baseline from which variation is reported. For example, sequence variants are identified in a patient by comparing the patient's DNA sequence to a reference sequence standard, used in the laboratory. Typically, differences between the patient and reference sequence are called sequence variation and are cataloged, interpreted and reported. Documentation of the reference sequence used is becoming increasingly important for normalization of results between laboratories. To meet this need NCBI is cataloging reference sequences used in clinical testing in the Core Nucleotide Database and can be referred to through the RefSeq identifiers. In collaboration with NCBI, the European Bioinformatics Institute (EBI) is also developing a database of reference sequences called Locus Reference Genomic Sequences (LRG). The standard is still in draft status. Importantly, NCBI's RefSeq and EBI's LRG will contain the same reference sequences, annotations and cross references to each other.

7.1.3.0 RefSeq

TABLE 6-7 – REFSEQ

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	RefSeq
OID	2.16.840.1.113883.6.280
Minimum attributes of the component	RefSeq ID
Other Comments	National Center for Biotechnology Information (NCBI) Reference Sequences contained in Core Nucleotide database. (Note version numbers are required to uniquely identify the reference.) Available at: http://www.ncbi.nlm.nih.gov/nuccore?db=nuccore

7.1.3.1 LRG

TABLE 6-8 – LRG

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	LRG
OID	2.16.840.1.113883.6.283
Minimum attributes of the component	LRG ID
Other Comments	Locus Reference Genomic Sequences an emerging standard led by the European Bioinformatics Institute. Available at: http://www.ebi.ac.uk/ebisearch/search.ebi?db=lrg&t=gene And http://www.lrg-sequence.org/page.php

7.1.4 Publicly Available References (valuable for clinical and translational genomics)

7.1.4.0 OMIM (optional)

Clinical genetic/genomic results can be reported with an OMIM id for association to relevant information in the OMIM knowledgebase, which contains a compendium of information on genetic based disease, genes and mutations.

TABLE 6-9 – OMIM

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	OMIM (Online Mendelian Inheritance in Man)
Responsible Body	Johns Hopkins
OID	2.16.840.1.113883.6.174
Minimum attributes of the component	OMIM ID
Other Comments	<p>Knowledgebase for genes, variants/mutations and genetic based phenotypes. Note this information includes somatic or acquired variants/mutations and phenotypes and is not limited to inherited variants/mutations and phenotypes.</p> <p>Available at: http://www.omim.org/ and through NCBI at http://www.ncbi.nlm.nih.gov/omim</p> <p>Additionally, dbSNP contains links to variants in OMIM.</p>

7.1.4.1 PubMed (optional)

Coding of references may include PubMed ids to peer reviewed literature (e.g. publications within a medical journal).

TABLE 6-10 – PUBMED

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	PubMed
Responsible Body	United States National Library of Medicine
OID	2.16.840.1.113883.13.191
Minimum attributes of the component	PubMed ID
Other Comments	<p>"PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites."</p> <p>Available at: http://www.ncbi.nlm.nih.gov/pubmed/</p>

7.1.4.2 PharmGKB (optional)

PharmGKB ids to community curated information on emerging pharmacogenomic associations.

Chapter 5: Use Case Scenarios

TABLE 6-10 – PHARMGKB

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	PharmGKB (Pharmacogenomic Knowledge Base)
Responsible Body	Stanford University, Department of Genetics
OID	2.16.840.1.113883.3.913
Minimum attributes of the component	PharmGKB ID
Other Comments	The mission of PharmGKB is "to collect, encode, and disseminate knowledge about the impact of human genetic variations on drug response. We curate primary genotype and phenotype data, annotate gene variants and gene-drug-disease relationships via literature review, and summarize important PGx genes and drug pathways." Available at: http://www.pharmgkb.org/

7.1.4.3 ClinicalTrials.gov (optional)

ClinicalTrials.gov id maybe transmitted as part of the interpretation indicating which clinical trials the patient may qualify.

TABLE 6-11 – CLINICALTRIALS.GOV

Code sets, vocabularies, terminologies and nomenclatures that need to be constrained:	ClinicalTrials.gov
Responsible Body:	U.S. National Institutes of Health and Lister Hill National Center for Biomedical Communications
OID	2.16.840.1.113883.3.1077
Minimum attributes of the component:	ClinicalTrials.gov Identifier
Other Comments:	"ClinicalTrials.gov is a registry of federally and privately supported clinical trials conducted in the United States and around the world. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. This information should be used in conjunction with advice from health care professionals." http://clinicaltrials.gov/ct2/home

TODOS

- Get OIDs for and discuss use
 - o ENSEMBL
 - o Genome build
 - o ClinVar
 - o dbVAR
 - o Amnon are there EU resources to add other than ENSEMBL?

- ?? EU clinical trials <https://www.clinicaltrialsregister.eu>
-
- Genome builds – need versions
- Gene Symbols – can change but are unique and mapping is maintained. Clinicians in the field a long time maybe more familiar with older symbols, so listing aliases may be helpful
- right alignment left alignment issues
- variant IDs and dbSNP variants
 - o RS#'s do not uniquely id a variant RS#'s identify a location
 - o dbSNP contains somatic and germline variants as well as pathogenic
- Reference Sequences – challenges with interoperability
 - o Do not use mapping tables for automated mapping from UCSC, NCBI, and ENSEMBL works for some genes at some time points, but not always. Need to map from transcripts to genomic coordinates and out.
 - o When comparing variants, make sure comparing apples to apples and oranges to oranges or compare at the genomic coordinate level
 - o EBI and NCBI sync on a frequent basis
- Challenges merging genetic data across testing platforms and implications for clinical use (human and computers)

8.Vocabulary Constraints

Ideally, binding to vocabularies should be part of constraining HL7 Clinical Genomics specifications consistent with the CG DAM and DIM. Constraining is typically done as part of an implementation guide over a universal specification. For example, the HL7 v2.5.1 Lab message was constrained in a US-Realm specific implementation guide for genetic testing results (see in 7). As part of this constraining process, message fields were bound to LOINC codes (see for example in 6.1). Also, the Clinical Document Architecture (CDA) was constrained in a universal implementation guide for genetic testing reports (GTR). In the GTR, the same LOINC codes were given as example vocabularies to bind to from the class attributes of the CDA.

Given the rapidly-changing nature of the clinical genomics field, it is preferably-preferable to have HL7 specifications bound to instances dynamically, so that a code is drawn from the most up-to-date vocabulary /value-set. It is important to note that dynamic binding requires strict compliance with indication of the code system id, name and precise version when binding is done at instantiation time.

Nevertheless, it is important to highlight here the type of concepts already coded in LOINC:

- Designating other coding systems and nomenclatures crucial for genomics, e.g., HGNC, dbSNP, HGVS, RefSeq, LRG, etc.
- Publicly available knowledge bases, e.g., OMIM, PubMed, PharmGKB, ClinicalTrials.gov, etc.
- Codes designating basic concepts, e.g., DNA region name, Amino acid change, Allele name, Medication assessed, Genetic disease analysis overall interpretation, Drug efficacy sequence variation interpretation, etc.
- Value sets designating possible types of a concept, the concept Amino acid change type can be Wild type, Deletion, Duplication, Frameshift, Initiating Methionine, Insertion, Insertion and Deletion, Missense, Nonsense, Silent or Stop codon mutation.

For more information, see <http://loinc.org>.

9. Review of Existing HL7 Clinical Genomics Specifications

9.1 HL7 V2 GENETIC TEST RESULT MESSAGE

The Genetic Test Result Reporting message is defined by a set of four nested LOINC panels, which serve as templates for the messages. In general, LOINC panel definitions include one LOINC code to identify the whole panel and a set of LOINC codes for each child element of that panel. A child element can also be a LOINC panel, and such panels can repeat, to provide a structure that can accommodate many reporting patterns. For each such child element, the panel definition also includes its data type, units of measure, optionality and answer list, as applicable. The definitional information for the four panels used to report Genetics Test Result Reports is included in the HL7 2.5.1 implementation guide (link to HL7 document on website).

In a message, the first panel is the master panel for the reporting of genetic analysis. The first child panel delivers an overall summary of the study results and includes options for reporting the traditional narrative report, the overall study impression, and a few other items. Depending on the study being reported, the summary panel may contain variables required to summarize a pharmacogenomics study, or those required to summarize the genetic findings associated with a disease or the risk of a disease. Next comes the discrete results panel, which contains the detailed results payload in a series of one or more "DNA sequence analysis discrete sequence variation panels". This last panel repeats as many times as needed to report all of the variations of interest.

For more information please refer to:

[Link to v 2.5.1](#)

Comment [as1]: Consider revise this text as an overview.

9.2 HL7 CDA IMPLEMENTATION GUIDE FOR GENETIC TESTING REPORTS

The Clinical Genomics Work Group developed a CDA Implementation Guide (IG) for genetic testing reports, with the support of the Structured Documents Work Group. The main purpose of this IG is to specify a Universal document standard for a Genetic Testing Report (GTR) typically sent out from a genetic laboratory to recipients who ordered the report. The GTR IG targets both human viewing and machine processing by representing the data in a renderable format along with structured entries; these entries are associated by 'clinical genomic statement' templates defined by this guide, which could empower clinical decision support by conveying clinical genomics semantics in an explicit way. This guide is defined as 'Universal' as it is flexible enough to accommodate various use cases, e.g., in translational medicine and clinical environments or of different genetic testing types.

For more information see http://www.hl7.org/implement/standards/product_brief.cfm?product_id=292..

10.HIT Data Standards

10.1 FAMILY HISTORY

A minimal core data set for family history can be found at in the ONC/HHS family history data requirements as developed by the multi-stakeholder workgroup (available at:

http://healthit.hhs.gov/portal/server.pt/community/use_cases_and_requirements_documents/1202/personalized_healthcare/15671)

10.2 SEQUENCE VARIATIONS / CHROMOSOMAL CHANGE

10.2.1 Small Genetic Variations within a Gene

[Point to 2.5.1 and CDA - GTR](#)

10.2.2 Structural Variations

[Point to 2.5.1 for cytogenetics](#)

11.HL7 Encapsulation of Raw Genomic Data Files

Amnon what should go here???? As we face nowadays a constantly growing stream of raw data in both research and clinical environments, it is important to develop approaches to coping with these streams that involve extraction of subsets of the data that might have clinical relevance to the patient. Examples of such data include medical imaging information via new techniques (along with extracted regions of interest), health sensor data (such as data from implanted electrocardiography devices, along with alerts generated through ongoing analysis of that data), or DNA sequences (along with clinically significant variants found in these sequences).

These raw data sets typically have common formats developed by the respective developer communities of medical imaging modalities or associated with personal health devices or genetic testing kits. Such raw data should be encapsulated in medical records, using common formats, so that 1) it can be referenced as evidence supporting analysis results and 2) it can be reassessed when needed.

Furthermore, clinically significant data sets are typically extracted from each type of the raw data. These extracts then become available to the clinical environment and thus, their representation should adhere to common and agreed-upon health information standards.

It is recommended that Clinical Genomics standard specifications support the encapsulation of raw genomic data through specialized constructs capable of holding bioinformatics formats, along with placeholders of key data items extracted from the raw data and optionally associated with phenotypic data. For example, if a patient's DNA sequences are the raw data, then extracted data sets may be a few of the variants found in these DNA sequences that are associated with responsiveness to drugs relevant to the treatment options being considered for that patient.

Formatted: Space After: 0 pt, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

Formatted: Font: (Default) AdvTT182ff89e, 9.5 pt, (Int'l) Times New Roman

12.Clinical Grade-Genomic Data File Standards

<Point to federally-mediated workgroup for development of clinical grade-GVF/VCF files>

13.Gaps & Extensions

13.1 LABORATORY ORDER ENTRY

One significant gap is the need to develop a laboratory order implementation guide for clinical sequencing/molecular diagnostics, which is capable of including relevant clinical history and a fully structured family history with familial mutations and risk assessment. Currently, laboratory orders are paper or pdf based, which has fulfilled the need while volumes remain low. However, as genetic analysis becomes a standards part of clinical care, paper-based order entry will not scale.

14.Outstanding Questions

1. Will electronic health record [systems](#) (EHR-[Ss](#)) incorporate a genomic repository housing a patient's genome/variome for access on demand, in much the same way images are stored in PACS (picture archiving and communication system)? Or will EHRs-[S](#) contain a pointer to a centralized repository? Or will the laboratory continue to sequence a patient's DNA each time a test is ordered?
- 4.2. [A possible solution to these questions is encapsulation of key genomic data into healthcare standards, while keeping pointers to the raw data on the one hand and associations with clinical data \(phenotypes\) on the other hand.](#)

15.Glossary

Genome: Entirety of a patient's inherited genetic information, unless specified as the cancer genome.

Sequence Variation: Variation from a common DNA reference sequence and synonymous with mutation.

Transcoding: Process of converting genetic data from a bioinformatic representation into a clinical representation, following healthcare IT data standards.

Variome: Variation from a reference sequence. That is a patient's DNA sequence can either be stored as a true sequence of nucleotide, or can be stored as a series of variations from a common reference sequence.

15.1 EXTENSION TO SPECIMEN SCENARIOS

15.1.1 Microbiome analysis of the patient

- a. Includes analysis of microorganisms living in the patients gastrointestinal tract or Genitourinary system and may aid in diagnosis

This page intentionally left blank.