This page: http://tinyurl.com/damcgdoc V3DAM_CG_CLINGEN_R1_I1_2018May

HL7 Domain Analysis Model: Clinical Genomics

Clinical Genomics Testing and Applications

May 2018

HL7 Informative Ballot

Sponsored by Clinical Genomics Work Group (CGWG)

Co-Chairs: Gil Alterovitz, Robert Freimuth, Bob Milius, Kevin Power, Amnon Shabo (Shvo)

Questions or comments regarding this document should be directed to Gil Alterovitz at ga@alum.mit.edu

Sponsored by Clinical Genomics Work Group

Principal Author and Co-Chair	Gil Alterovitz Harvard Medical School and Boston Children's Hospital
Co-Chair and Contributing Author	Robert R Freimuth Mayo Clinic
Co-Chair and Contributing Author	Bob Milius National Marrow Donor Program
Co-chair and Contributing Author	Kevin Power Cerner
Co-Chair and Contributing Author	Amnon Shabo (Shvo) Philips
Contributing Author and Former Co-Chair	Mollie Ullman-Cullere OM1
Contributing Author and Former Co-Chair	Daryl Thomas Personlis
Contributing Author	Larry Babb Geneinsight, a Sunquest Information Systems Company
Contributing Author	Scott Bolte Niss Consulting
Contributing Author	Lynn Bry, MD Brigham and Women's Hospital
Contributing Author	Bob Dolin, MD Elimu Informatics
Contributing Author	Kevin Hughes, MD Massachusetts General Hospital
Contributing Author	David Kreda Harvard Medical School
Contributing Author	Alexander Marzilli Cornell University
Contributing Author	Hanna May Wellesley College
Contributing Author	Rachel Patel Dartmouth College
Contributing Author	Matthew Rioth University of Colorado
Contributing Author	Jeremy Warner Vanderbilt University

Contributing Author	Grant Wood Intermountain Healthcare

With contribution of members of the HL7 Clinical Genomics Work Group

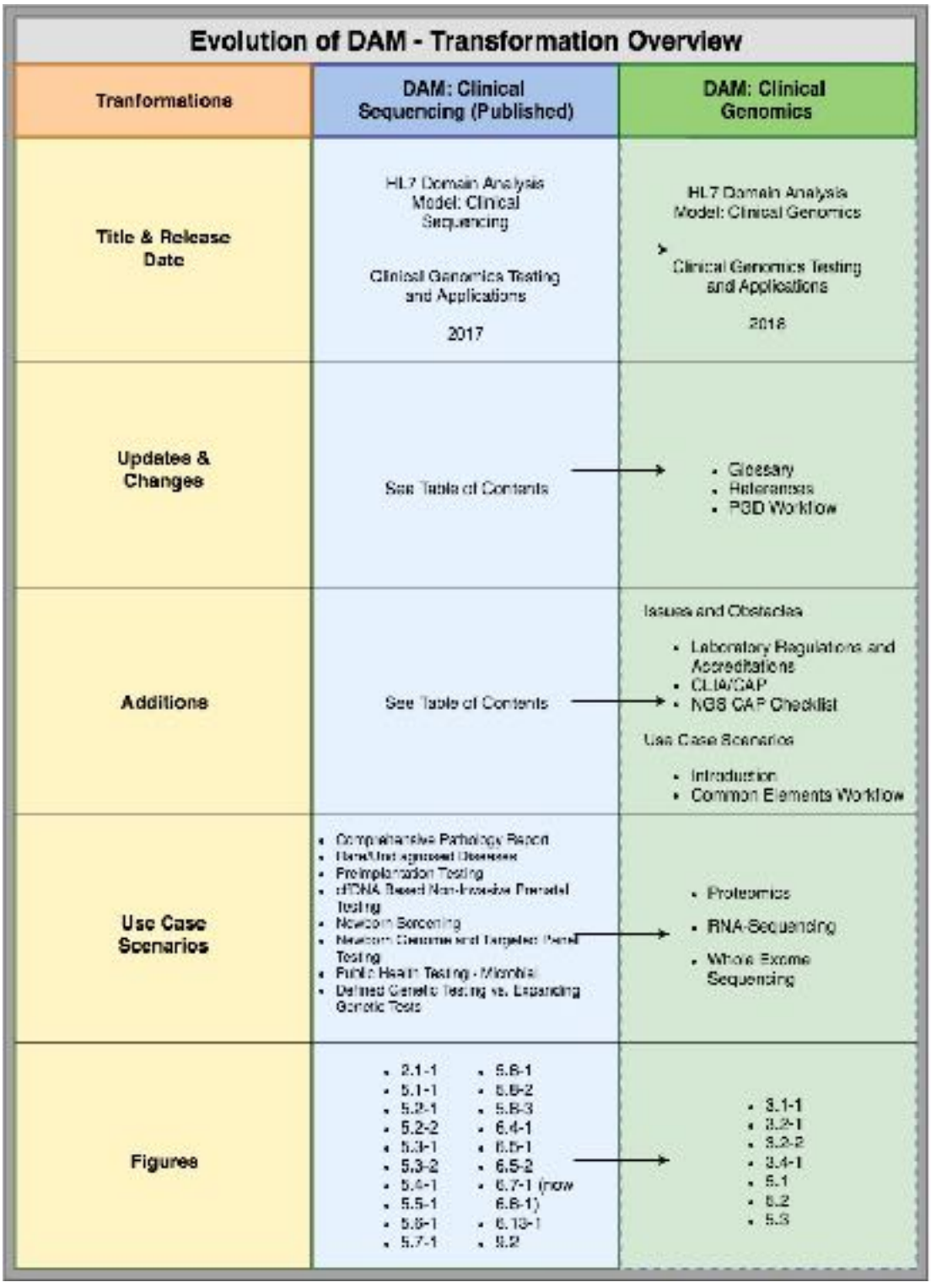
Summary Table of Contents

Detailed Table of Contents

1. Introduction

In 2017 the *HL7 Domain Analysis Model: Clinical Sequencing was published.* It cataloged the breadth of genetic/genomic testing use cases and clinical scenarios with focus on clinical sequencing, drawing use cases from a number of stakeholders. It discussed current challenges and lessons learned, and raises questions to consider for future implementations. While this document discusses the use of new technology (e.g. Next Generation Sequencing or NGS), it must be remembered that the vast majority of clinical genetic testing is still performed on testing platforms that were in use years ago, and it is the goal of the Clinical Genomics Work Group to facilitate interoperability of genetic/genomic data, independent of specific genetic testing platforms or methodologies. Following the Clinical Sequencing Domain Analysis Model: Clinical Sequencing, this document is part of an effort to develop a Clinical Genomics Domain Analysis Model (CG DAM) by expanding the use cases to the overall clinical genomics field.

Figure 1: The figure below summarizes the document's transformations since 2017, with a focus on updated as well as added materials.



1.1. Purpose

The *HL7 Domain Analysis Model: Clinical Genomics* should be used to inform standards developers and implementers for the design of scalable, interoperable solutions covering the breadth of clinical genetics/genomics scenarios.

1.2. Audience

This document is designed to be used by analysts and developers who require guidance on incorporation of genomic data in clinical care and translational research IT environments. 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 document will not serve as a tutorial on that subject.

1.3. Scope

This initial version toward a domain analysis model begins by detailing a variety of use cases key to personalized genomic medicine and translational research, including more typical scenarios for testing of a person's inherited or germline genome, cancer genomics/tumor profiling, early childhood developmental delay, neonatal testing, and newborn screening. In addition, each use case may include several 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. While not in scope for this publication, for future publications of this DAM, it would be useful to add more specific information models for various use cases. Ideally, the DAM use cases and semantics represented in these information models could then be used in the design of specifications across clinical genomics, so that instances compliant with these specs can be mapped with no loss of the intended semantics.

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.
- Trading partners agree to all standards, methodologies, consent, privacy and security.
- The order, paper or electronic, associated with the laboratory result contains sufficient information for the laboratory to construct the laboratory result properly.
- Privacy and security has been implemented at an acceptable level based on specifications that are handled by other standardization work groups.
- Legal and governance issues regarding data access authorizations, data ownership and data use are outside the scope of this document.

2. Use Case Stakeholders

Stakeholder	Contextual Description
Anatomic & Surgical Pathology; Hematopathology	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
Treating Clinicians	Healthcare professionals making a diagnostic, treatment, or preventative decision or recommendation, based on the genetic/genomic information
Healthcare Entities	Organizations delivering healthcare
Prior authorization personnel	Agents or employees of a Healthcare Entity whose responsibility is to identify and enable the prescription of high-cost genetic and genomic treatments
Informaticists	Individuals responsible for the integration of genomic data into local EHR and other clinical systems
Clinical Data and Knowledge Management / Delivery	Entities that provide local governance of clinical data, knowledge management, delivery of knowledge to the point of care (e.g. implementation of genomic-based CDS rules)
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. *in vitro* diagnostic tests)
Patients / Individuals	Members of the public that use healthcare and wellbeing services
Public Health Agencies	Agencies which help to protect and improve health and healthcare of the public (e.g. CDC)
Registries	Systems for the collection, analysis, and distribution of data for the improvement of public health
Genetics Standard Organizations	Organizations that create standards (HGVS, GA4GH, HGNC, LOINC, CDISC SDTM-PGx etc.)
Public reference databases	NCBI (e.g. Gene, ClinVar), EBI, COSMIC, LSDB, CDISC SHARE etc.

Professional organizations	Academies of Medicine, College of American Pathologists, American College of Medical Genetics
Grant-funded consortiums	Grant-funded organizations disseminating data (PCORI, Undiagnosed Diseases Network, etc.)
Producers of open source tools	Broad Institute, NCBI, EBI, SMART, ClinGen

3.

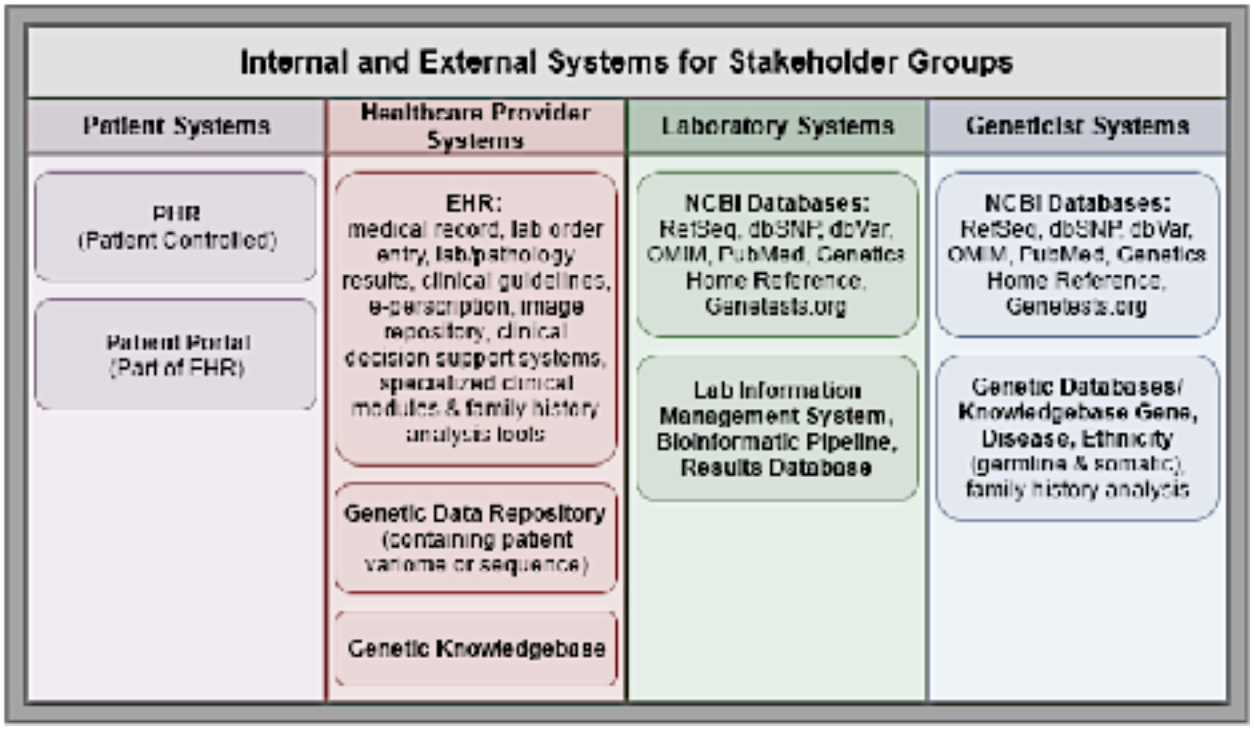


Figure 2.1-1: Systems for stakeholders. The above figure separates various internal and external systems into relevant the healthcare categories of patient, healthcare provider, laboratory, and geneticist.

3. Issues and Obstacles

Numerous challenges exist in the area of policy, patient and clinician education, and reimbursement, which include adoption of electronic health records and laboratory information management systems and data security. These challenges 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 addressed in this document include interoperability among various systems and useful structuring of genomics 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.

3.1. Laboratory Regulations and Accreditations

The laboratory is one of the most stringently regulated entities in the field of healthcare. As such, proper compliance in the laboratory environment is necessary to ensure interoperability and safety. A clear distinction between laboratory regulations and accreditations is important to standardization and efficiency.

Laboratory Regulations and Accreditations				
	Full Name	Type	Goal	Applicability
CAP	The College of American Pathologists Accreditation Program	Accreditation	Helps laboratories meet CLIA requirements and ensure compliance through the guidance of the most comprehensive scientifically endorsed laboratory standards	All laboratories and other healthcare organizations
CLIA	Clinical Laboratory Improvement Amendments	Regulation	To establish standards for all laboratory testing and ensure the accuracy and reliability of patient test results	All laboratories performing testing of human specimens for health assessment; in all settings including commercial, hospital, and physician office laboratories
FDA	Food and Drug Administration	Regulation	Assuring the application of current good manufacturing practice regulations (cGMP) to blood	Laboratory with a blood collection center or transfusion services

HIPAA	Health Insurance Portability and Accountability Act	Regulation	To improve and secure the electronic transmission of health information	All laboratories and other healthcare organizations
TJC	The Joint Commission	Accreditation	Assures that the accredited organization has demonstrated compliance with the most stringent standards of performance (CLIA included), focuses on operational systems critical to the safety and quality of patient care	Hospital laboratories, Reference laboratories, Physician office laboratories, Assisted reproductive technology laboratories, Clinics etc.

3.2. CLIA

The Clinical Laboratory Improvements Amendments of 1988 (CLIA) sets forth the conditions that all laboratories must meet to be certified to perform testing on human specimens in the United States.

CLIA's purpose is to guarantee high-quality laboratory procedures, and its jurisdiction extends to approximately 254,000 laboratory entities. Electronic laboratory data and its formatting under United States law has and will continue to be impacted most by the regulations involving the implementation of CLIA. More information on the specific guidelines and mandates of CLIA and their role in clinical genomics will be outlined further at a later time.

Data indicates that CLIA has helped to improve the quality of laboratory testing in the United States through a survey process that is education-oriented and quality-focused. Within CLIA inspected laboratories, the total number of quality deficiencies has been shown to decrease approximately 40% from first laboratory surveys to the second and further on subsequent surveys (https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Program_Descriptions_Projects.html) .

CLIA Overview:

All entities that meet the definition of a “laboratory” under CLIA regulations are required to obtain an applicable CLIA certificate prior to conducting any testing on human specimens.

According to CLIA, a *laboratory* is defined as: “a facility for the biological, microbiological, serological, chemical, immunohematological, hematological, biophysical, cytological, pathological, or other examination of materials derived from the human body for the purposes of providing information for the diagnosis, prevention, or treatment of any disease or impairment of, or the assessment of the health of, human beings. These examinations also include procedures to determine, measure, or otherwise describe the presence or absence of various substances or organisms in the body.”

If an entity meets the requirements for being defined as a laboratory, CLIA and its subsidiary regulations are applicable. Specifically, a laboratory is applicable for CLIA regulation when:

- The entity reports patient-specific results to another entity AND
- The reported results are utilized or are intended to be utilized for any analysis pertaining to the health of human beings

CLIA & Sequencing

One approach institutions have taken to integrate clinical sequencing into an existing CAP/CLIA-accredited laboratory is to incorporate sequencing with step-wise procedures, as with cytogenetics. The alternative approach is to establish a new CLIA-compliant laboratory or make an existing genomics core facility CLIA-compliant.

Potential issues (Kirkup, Benjamin C., Steven Mahlen, and George Kallstrom. "Future-generation sequencing and clinical microbiology." Clinics in laboratory medicine 33.3 (2013): 685-704) to consider include:

- implementing qualified key laboratory personnel as well as a CLIA-qualified director
- appropriate space to allow for a unidirectional workflow separating pre- and post-amplification processes
- developing a validation study and implementation plan for each assay offered and participating in a CLIA-approved proficiency test or sample exchange program
- having an experienced quality program manager to oversee the quality program and document management system
- having the financial resources to invest in developing and operating this unique regulatory environment

Sample Requirements for Cytogenetic Testing:

A laboratory that provides services in the field of clinical cytogenetics must meet requirements from several sections of the CLIA document. A sampling of CLIA-specific requirements are outlined below:

Procedure	Reporting	Personnel

• Patient specimen identification
• Accessioning
• Cell preparation
• Image reproduction & photography
	• Reporting and storage of results
	  ○ Karyotypes
	  ○ Photographs
	• Records of
	  ○ Procedure media
	  ○ Reactions observed
	  ○ Cell count
	  ○ Karyotype count
	  ○ Chromosome count
	  ○ Banding quality
	• Appropriate tissue or specimen resolution
	• Adequate karyotype number
	• Determination of sex by full chromosome analysis
	• Summary and interpretation of observations
		• Technical supervisors are required to
		  ○ Be a state-licensed MD, DO, DPM OR
		  ○ Hold a doctoral degree in biological sciences from an accredited institution AND
		  ○ Have 4 years experience in genetics, 2 of which must be in clinical cytogenetics

[redacted]

The College of American Pathologists (CAP) Laboratory Accreditation Program offers the most scientifically rigorous, laboratory-customized checklist of requirements to certify and accredit the entire spectrum of laboratory procedures and tests.

Its utility is twofold-- CAP accreditation assists laboratories in meeting CLIA regulatory requirements and helps to assure and maintain accurate test results and patient diagnosis. The Centers for Medicare and Medicaid Services (CMS) the responsible entity for CLIA certification has granted the CAP Laboratory Accreditation Program deeming authority, which allows for a CAP inspection in the place of a CLIA inspection. It is also recognized by the Joint Commission, considered the gold standard of certification programming, and can also be used to meet several state certification requirements.

The College of American Pathologists (CAP) accreditation checklists contain requirements that are updated yearly to reflect advancing technologies. Additionally, these checklists contain are discipline-specific questions. These questions in each checklist cover proficiency testing, quality control, and quality improvement activities including supervision, the procedure manual, specimen collection and handling, reporting of results, reagents, instruments and equipment, individual tests performed in that laboratory section, personnel requirements, physical facilities, and safety. Discipline-specific checklists are available by purchase.

3.4. Next Generation Sequencing: CAP Checklist

The large volumes of data produced by NGS procedures, as well as the associated composite computational analyses, have necessitated a creation of new requirements specific to bioinformatics for assessment and implementation of new technology and software releases, for record keeping and validation, quality control and monitoring, as well as data storage.

Accordingly, CAP has classified NGS testing and procedures into two central analytical processes:

- A wet bench process, including
  - Sequence generation
  - Specimen handling
  - NGS library preparation
- A bioinformatics process, including
  - Software-aided variant prioritization/interpretation
  - Sequence alignment and assembly
  - Variant calling and annotation

Each of these processes have their own specific requirements, and these NGS-specific CAP requirements have been updated as of July, 2015. For ease of understanding, the requirements can be summarized and sorted into five major areas:

Laboratory Policy

• The laboratory creates policies for:
  ○ Selection/evaluation of reference facilities for NGS procedures
  ○ Indications for confirmatory testing
  ○ Performing the wet-bench analytical process to generate NGS data

Records/Reporting

• The laboratory maintains records/reports for: ○ Tracking specimens referred to other facilities involved in NGS testing ○ Patient specimens for which steps differ from written procedures ○ Methods, instruments, and reagents used for processing and analyzing samples ○ Interpretation and reporting of sequence variants ○ Reporting genetic findings unrelated to the clinical purpose for testing
Bioinformatics
• The laboratory ensures secure internal and external transfer and storage of NGS data • The laboratory retains necessary NGS data for a period of two years • The laboratory maintains traceable files for all versions of the bioinformatics pipeline used to generate NGS data
Quality Assurance
• The laboratory follows a written quality management program for: ○ The analytical wet-bench process and its modifications/updates ○ Monitoring and implementing upgrades to NGS instruments and software ○ Recording the bioinformatics pipeline and its modifications/updates
Fetal Aneuploidy NGS
• The laboratory follows specific guidelines pertaining to: ○ Sample requisition ○ Test result reporting

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 applications that alerts a clinician if he/she orders a drug which is contraindicated or unlikely to be effective. Developing systems which are capable of considering genetic markers associated with drug metabolism, efficacy, and toxicity during the order entry process will reduce medical error and will become increasingly relevant as we learn more about specific interactions between human health and genomics.

4.1. Current and Emerging Testing Paradigms

Clinical genomics is now at a paradigmatic crossroads due to improvements in the technological performance, availability, and medical utility of sequencing. In the past, clinical genomic testing typically focused on a **single clinical question**, for example, a clinician may ask whether or not a patient has a variant associated with drug efficacy/resistance, i.e. EGFR, KRAS (NSCLC), BRAF (melanoma). Gene chips can now be used to answer these questions by looking for *a priori*-specified variants. A second question may be whether or not the patient has (a) variant(s) associated with drug metabolism for a particular drug. Examples of this include testing differences in metabolism efficiency via testing for CYP variants involved in cytochrome P450 metabolism. Finally, testing may want to answer whether or not a patient has a variant associated with a particular disease. Examples include cystic fibrosis and cardiomyopathy. Again, a gene chip may help here. Deep sequencing may also be done in order to find de novo or rare variants not present on a particular gene chip. This may result in a variant of unknown significance (VUS). But, in all cases, the patient has consented for a particular test and the physician has a specific hypothesis when ordering the test.

Going forward, the current testing paradigm may end up being continued for a significant duration and should thus be supported throughout its lifespan. Yet, the community must prepare for an emerging paradigm that is compatible with current standards. This emerging paradigm should work for hypothesis-based confirmation and diagnostics through unprecedented interoperability and communication between healthcare professionals and electronic documents, and should also work for the payor/provider relationship and reimbursements. Instead of the clinician asking a single clinical question, the paradigm is shifting to: “What can the genome tell me?”

In addition, there currently is an increase in scenarios where there is a need for investigation of unknown variants and where some may wish to not be reimbursed, and a system should be variable for these scenarios. For example, the Dana Farber Cancer Institute (Boston, MA) may want an interpretation in the report but may not want that interpretation to be structured. Particularly for large tests there may not be a desire to structure an interpretation. In somatic testing, inclusive of an interpretation, one is merely reporting out the variants and using tools such as COSMIC to filter out non-somatic variants. Pathologists and pharmacists can consume variant data visually and/or through a CDS tool that can keep certain data up to date, e.g. if variants or their corresponding interpretations change over time. They can also consume other clinical data such as body weight, diet, etc. in order to affect the manner with which CYP deals with interpretations. Finally, today the ability to encode structured interpretations for EMR is becoming more valuable and advantageous.

In the future, the use of APIs developed and enabled by SMART/FHIR Genomics (Alterovitz, et al. JAMIA 2015) can enable EMR-based apps (Warner, et al., JAMIA 2016) that enable contextualized, dynamic visualization and interpretation of NGS data. In the case of germline testing, NGS coupled with such apps would allow unprecedented insight into the genetic factors and variants that contribute to disease and wellbeing and start a new paradigm of personalized healthcare. In the case of somatic testing, these standardizations and technology will assist in diagnosis, prognosis, and treatment of illnesses and ailments such as cancer on a case-by-case basis, through the use of pharmacogenomics, drug dosage calculators, etc. Developments in fetal testing (either amniotic or through cffDNA) could increase screening efficiency for rare genetic disorders and de novo variants. Apps could enable questions that link to knowledge, e.g. what is the variant associated with tyrosine kinase inhibitor efficacy/resistance (EGFR) or cardiomyopathy risk? Further, in large tests, apps may not only encode

specific interpretations, but can also consume that data and keep interpretations up to date.

4.2. Somatic

The analysis of somatic variants has a variety of use cases and instances where it would be useful to the patient, physician, oncologist, and pathologist. First, an analysis of somatic variants can be done protectively, to determine actual variants. Second, genetic analysis of a variant would allow greater accuracy in prognostics. Third, a genetic analysis will be able to provide a diagnosis of specific cancer subtypes and treatment for those subtypes. Finally, a somatic cell analysis can reveal drug efficacy or resistance, as in the case of EGFR-associated drug resistance and efficacy.

4.3. Germline

Germline analysis provides a key resource in disease diagnosis, risk assessment, and subtype characterization. For a laboratory all would be the same, but an examining physician would want more discrete interpretations. It is important that the interpretations are not separate from the findings themselves. Further there would be a desire to assist the laboratory, not necessarily the medical geneticists, in looking at the data and clinical findings. In addition to diagnosis, germline sequencing could support drug metabolism (in CYP) and in drug toxicity (e.g. hearing loss if prescribed an antibiotic).

4.4. NGS Approach

There are many examples of applications for precision medicine, NGS, and FHIR to work together. Firstly, in an NGS approach, the clinician could use the NGS platform to determine any information or data that the clinician would need in properly treating a specific patient, or in other words, for the current treatment of a current patient. Just one example would be to look at the general category of drug resistance.

From that point, an NGS platform can further be used to determine any information or data that would be useful for any current or future treatment decisions. As an example, a clinician might be inclined not just to order somatic testing for an immediate treatment of cancer, but also to order a germline sequencing to determine if there are any genetic risks that may affect the patient later (i.e. additional cancers). A clinician may order a somatic and germline test together and receive one overall diagnostic report. Another example of NGS utility in future diagnostics would be running multiple samples to assess risks for ovarian cancer in a patient when looking at breast cancer.

To achieve accurate and meaningful diagnoses among the other varied applications of precision medicine, there needs to be a way to determine the variations in the patient's genome compared with standard references by current knowledge. In this case, relevant variants would be stored to the reference. As opposed to a targeted gene panel, a genome sequence could give insight to drug toxicity and metabolisms along all drugs, which could be stored as a variation to the reference sequence.

Finally, NGS would deal with the identification and classification of variants which could yield future use for a patient. In this scenario, the current sequence, including both sequences of clinical relevance and unknown relevance, is stored to be reviewed further in the future.

5. Use Case Scenarios

The use cases for sequencing put forth in this document share several common elements within their respective workflows. Here we offer a standard workflow that is (1) applicable to several sequencing scenarios and (2) adaptable to reflect novel technologies and procedures in new areas of clinical sequencing.

While almost all use cases share certain elements in common (i.e indication for testing, submission of test order), many use cases diverge at notable elements, which are more specific depending on the use case involved:

- Testing
- Interpretation
- Receive structured results into EHR

Additionally, some of the described use cases require auxiliary elements outside of the environments detailed in the common element workflow. For instance, Somatic Testing includes a column for incorporation of an outside Pathologist, and other use cases like Public Health Reporting and Newborn screening account for the additional governmental involvement/regulation in their respective procedures.

Figure 5-1: Common elements of the clinical sequence workflow involving the patient, providers, laboratory, and pathology.

5.1. Scenario 1: Specimen Identification

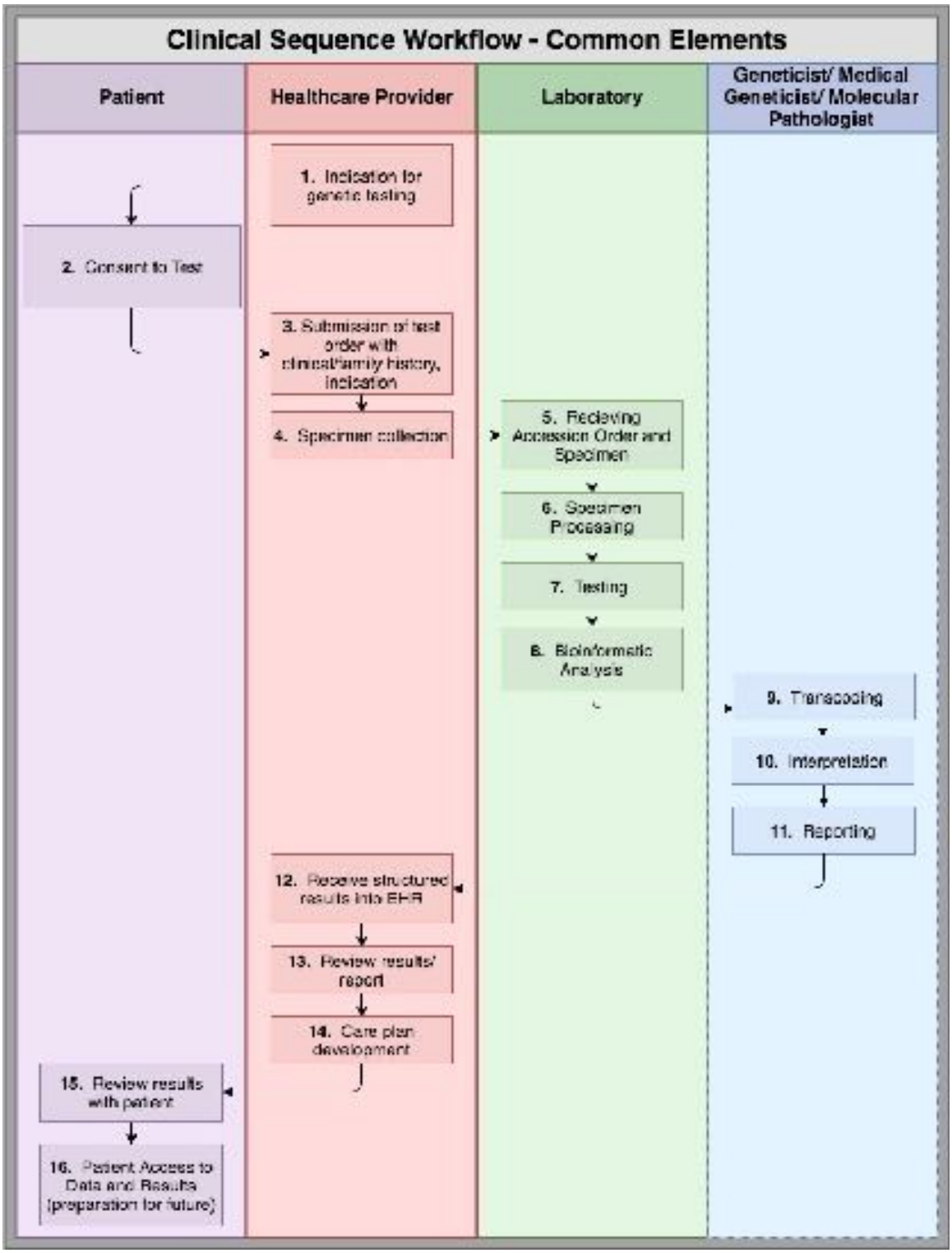
Use cases for sequencing require identification of one or more specimens to be used in laboratory analysis. This likely requires the identification of specimen groups (i.e. separate specimens and associated derivatives) originating from the same or related patients. Derivatives analyzed from these testing scenarios include: DNA, RNA, and Protein.

5.1.1. Germline testing for biomarkers/variants (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 will be extracted from the sample. Except for low level heterogeneity or acquired somatic variants, the variants identified in this specimen are ubiquitous throughout every cell in the patient and are inherited from their mother and father (except in the case of spontaneous variants). The typical genome contains about 4.1 to 5 million variants (The 1000 Genome Consortium, 2015). 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 biomarkers/variants (somatic/tumor specific)

To identify somatic (i.e. acquired) variants within a cancer specimen, a laboratory can choose one of three methods: 1) analyze both a germline “normal” specimen and somatic



"tumor" specimen and curate the differences; 2) "subtract" the somatic specimen from the germline specimen; or 3) analyze a somatic specimen and remove germline findings through bioinformatics post-processing algorithms. Due to the fact that tumor/normal testing is roughly twice the cost of tumor-only testing, many labs do not carry out this procedure rou-

tinely. When they do, the somatic/cancer specimen contains the germline sequence as well as the somatic variants present in cancer. The laboratory compares the two sequences and identifies variants unique to the cancer to definitively classify a variant as somatic. Note that this can be a complicated process because cancer cells acquire mutations throughout their lifespan and pass them on to daughter cells (see Figure 5.1-1). Also, directly tumorigenic germline variants will be present in tumor and normal, but are usually preserved for reporting. This is not the case when approach #2 is used. For the third scenario, testing labs rely on internal knowledge and/or publicly available database (e.g. 1000 Genomes, ExAC) to identify and remove germline variants.

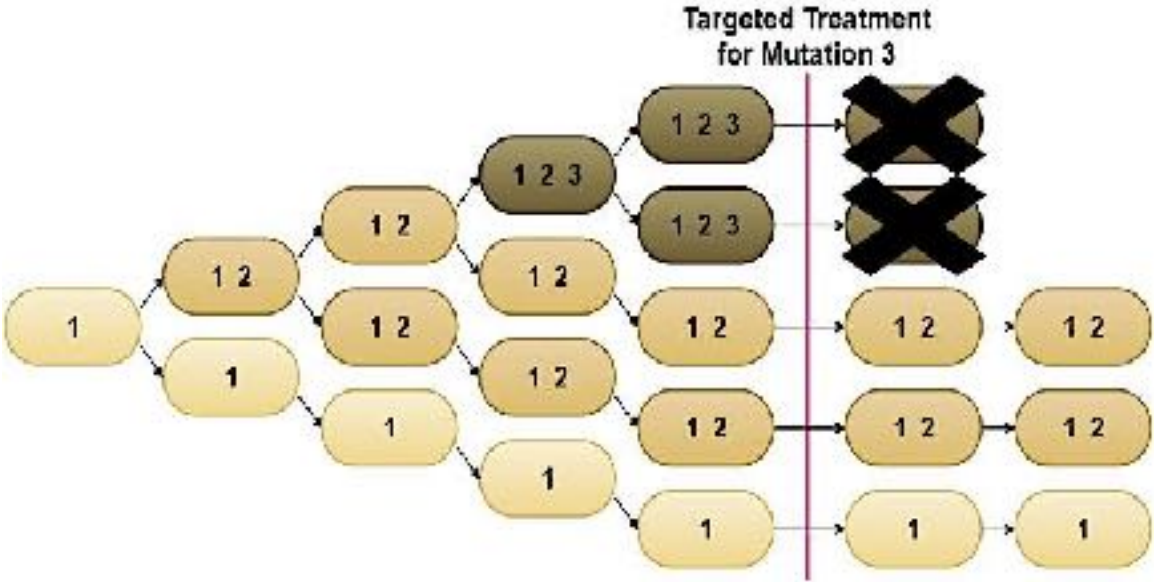


Figure 5.1-2: A simplified representation of cancer cells acquiring mutations or sequence variants, represented as numbers 1, 2, and 3, in dividing cancer cells. Note targeted therapy can kill a specific population of cancer cells while other populations survive.

Changes in the population of cells with particular variants will change over time and space, as well as in conjunction with events such as antineoplastic therapy. In the case of tumor metastasis, each lesion can be considered a separate “founder” population and may not share variants with other lesions; this is called intertumoral heterogeneity (see Meador et al. Clin Cancer Res 2014). Within a single mass of cancer cells, there may also be substantial differences, call intratumoral heterogeneity. Commonly, targeted chemotherapy may kill a specific population of cancer cells with specific variants and other cancer cell populations may survive and continue to divide (see Figure 5.1-1). 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 some scenarios, a laboratory may focus sequence analysis on well-studied genes/variants identified only in cancer. Commonly, these variants 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 utero. Specimens, sequences, and identified variants/variants from these studies should be clearly annotated as somatic.

It should be noted here that somatic specimens are often limited in availability and require a biopsy/surgery to obtain a specimen from the tumor site.

In summary, systems need to support both testing paradigms:

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

5.1.3. Pediatric Testing

Pediatric testing is most commonly used for the identification of biomarkers, variants, and variants causal to rare early childhood conditions. In addition to inheriting maternal and paternal variants, a child's genome typically contains novel variants not found in either parent. Matched patient, maternal, and paternal specimens facilitate a comparison which aids in the identification of original biomarkers, variants, and variants in the patient.

5.1.4. Prenatal Testing

Prenatal testing specimens can come from the amniotic fluid, a maternal serum, or cffDNA circulating in the maternal blood stream. Originating from the trophoblasts making up the placenta, cffDNA (cell-free fetal DNA) is estimated to comprise of an average of 11-13.4% of the DNA in the maternal blood (Wang et al. 2013). Using cffDNA for fetal testing provides a non-invasive (for the fetus) method for fetal genetic testing, thereby significantly reducing risk to the pregnancy.

Prenatal testing is commonly reported on the maternal medical record. Therefore, to avoid mistaking fetal results for maternal results, fetal variants should be clearly labeled as 'prenatal'. Most often prenatal/fetal and maternal specimens are matched and compared for analysis.

5.1.5. Infectious Disease Testing

Infectious disease testing involves the analysis of patient specimens for the presence of infectious organisms through the identification of organism specific genomic biomarkers/variants. These findings may subsequently be used to identify the specific organism, inform prognosis, and/or guide treatment. Where genetic findings are reported into the patient medical record, these genetic findings must clearly differentiate microorganism from human genomic findings, following similar data standards as used for other testing scenarios above.

5.1.6. Emerging Specimen scenarios

5.1.6.1. Microbiome analysis of the patient

This includes analysis of microorganisms living in the patient's gastrointestinal tract or genitourinary system and may aid in diagnosis. A fecal or urine sample is collected from the patient, DNA is extracted from the sample, and a combination of NGS and 16S rRNA gene sequence amplification are analyzed to determine the populations in the microbiome. Whole genome sequencing can be utilized for examining specific populations of interest (e.g. those that display drug resistance).

5.1.6.2. Cell-free circulating tumor DNA (ctDNA)

An emerging non-invasive approach to acquire solid tumor DNA is to extract it from circulating plasma. Some (probably most) cancer cells release oligonucleotides – around 170 bp pieces of DNA – into the patient's bloodstream (ctDNA). Traditional mechanisms for testing of

solid tumors requires biopsy of the tumor, which is invasive and may be impossible for some anatomic locations. Aside from the advantage of a less invasive method of testing, this technique may also overcome some of the concerns surrounding tumor heterogeneity (e.g. direct sampling of a single anatomic site may not reflect all genomic aberrations observed across all anatomic sites). Commercial tests are expected to quickly emerge over the next few years.

5.2. Scenario 2: Clinical Sequencing – Germline Testing

5.2.1. Description of Primary Clinical Sequence Workflow – Germline Testing

Germline testing involves the interaction between patient, health care provider, molecular diagnostic laboratory, and geneticist / medical geneticist / molecular pathologist. The testing is initiated at the discretion of the clinician, being necessary to inform accurate and effective patient care. After the sample is collected by the healthcare provider (for more information see Section 5.1.1), the specimen is received, processed, and sequenced by the laboratory. The data is analyzed and prepared for processing before it is sent to the geneticist/medical geneticist/molecular pathologist that often times will work in the laboratory. There, the data is transcoded for IT standards, interpreted, and compiled into a report. The report is then entered into the patient's EHR where it can be seen by the healthcare provider and a patient care plan can be developed. A figure of the process and more detailed descriptions of the steps can be found on the following pages. It should be noted here that most often the geneticist is working in the laboratory, which is represented in Figure 5.2-1 by a dotted line separating the laboratory and geneticist.

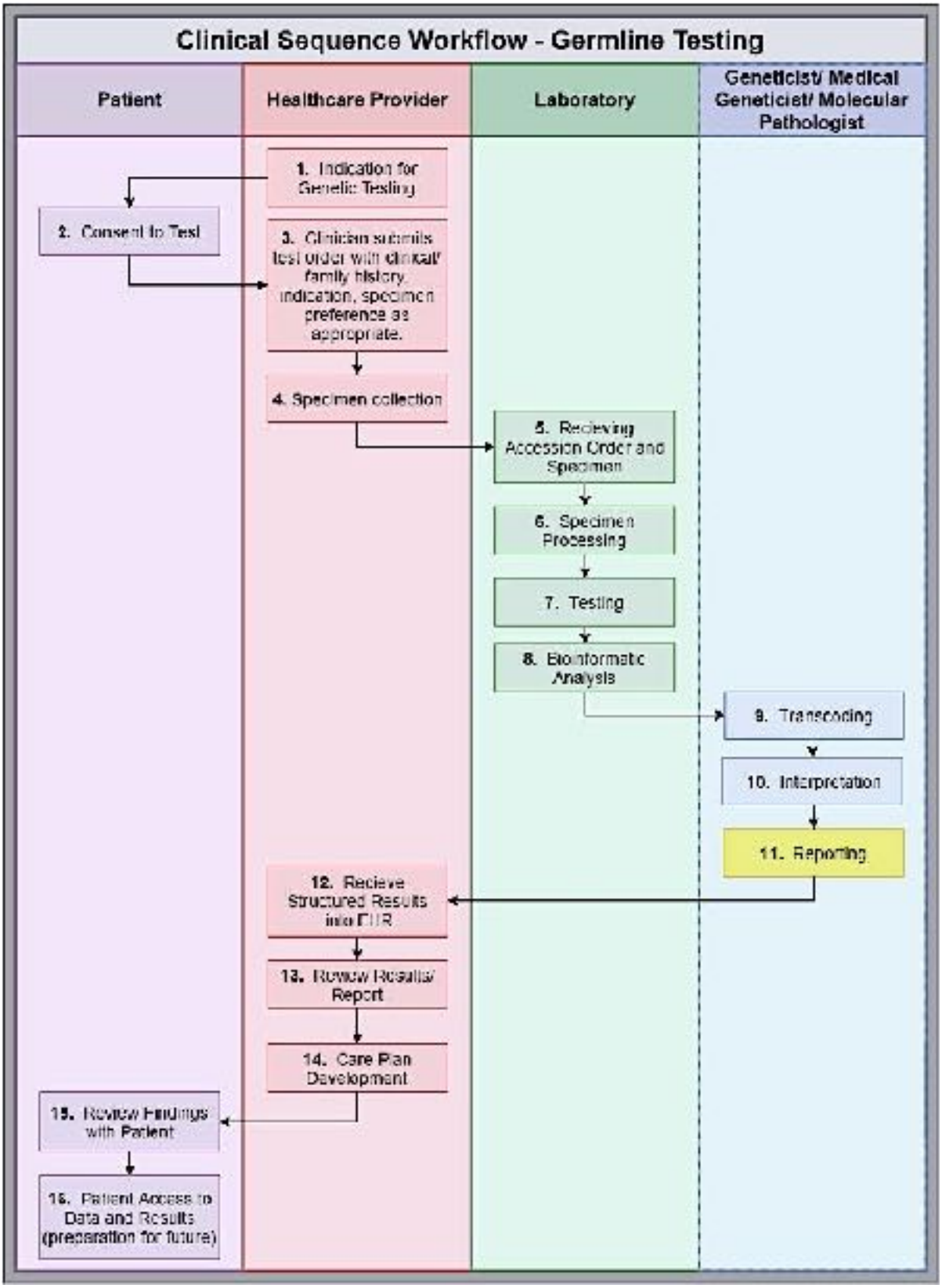


Figure 5.2-1: General workflow of germline testing: 1. Clinician determines that a genetic test is needed to inform patient care decisions. Often this includes family history based risk assessment. 2. Clinician obtains patient consent

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 genomic data is transformed from bioinformatics format into healthcare IT data standards. Not all raw data from the bioinformatics pipeline needs to be shared outside of the lab discretely, and some of this data may be shared in its native format. 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 (EHR). These may include variome data eventually. 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

5.2.2. Alternative Germline Workflows

In addition to the primary germline workflow, alternative workflows exist where genetic information from older germline testing is reviewed and reevaluated as knowledge of the health implications of genetic sequences expands. Alternative workflows may become more common as confidence in data quality increases and size of datasets increases.

5.2.2.1. Alternative Flow 1: Chart Review

If a sequence variant (i.e. variant) of 'Unknown Significance' is 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). . A clinical request is put into the laboratory and geneticist to reinterpret sequence data already a part of the patient's health record/EHR based on a chart review. The geneticist receives the request and looks up the patient's corresponding germline sequence. Using the most updated interpretation or translational tools, the sequence is reinterpreted. The reinterpretation is compiled into a report and entered into the EHR.

5.2.2.2. Alternative Flow 2: New Genetic Knowledge

A testing laboratory may contact the ordering clinician if the clinical implications of a sequence variant (i.e. variant) previously identified in the patient have changed. As interpretation and translational tools are updated with clinical knowledge, patient's germline sequence is automatically reinterpreted.

5.2.2.3. Alternative Flow 3: New Clinical Indication

A clinician may contact the laboratory when documenting a new clinical Indication if they feel the new indication might result in a different interpretation of existing data. A clinical request is put into the laboratory and geneticist to reinterpret sequence data already a part of the patient's health record/EHR based on a new clinical decision. 2c. The results are then reinterpreted with updated clinical knowledge and entered back into the EHR

5.2.2.4. Alternative Flow 4: Clinical Decision Support Initiated Reanalysis

Existing genetic data may be reanalyzed as part of decision support. For example, studies now show that particular variants of the SLCO1B1 gene are associated with increased risk of simvastatin toxicity. Ordering simvastatin triggers a reanalysis of sequencing data for genotyping of the SLCO1B1 gene. [Note that as confidence in data quality increases, this scenario may not need to involve the lab, but instead be executed by direct bioinformatic query of genetic information from older germline testing].

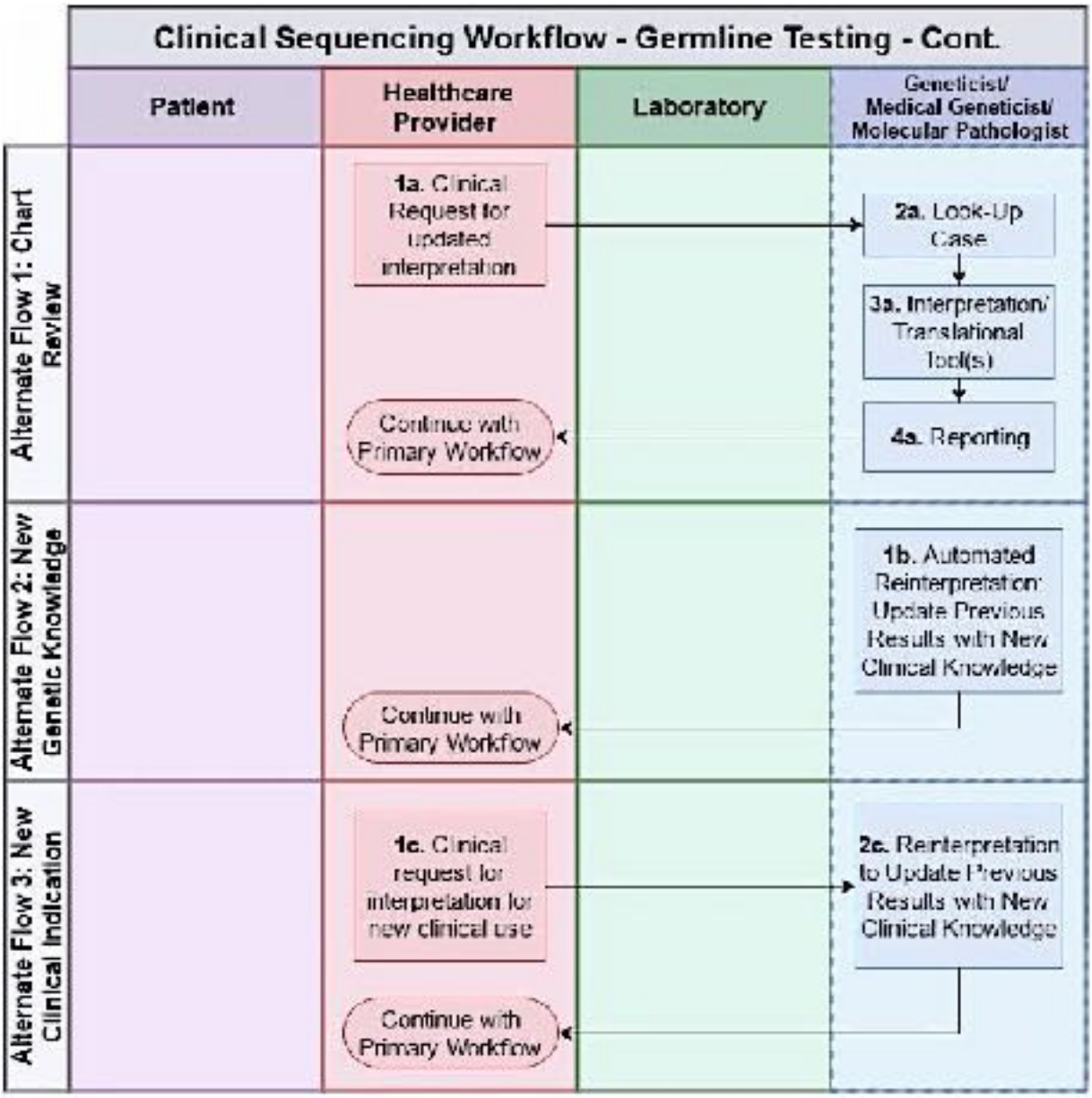


Figure 5.2-2: Alternative flows associated with chart review, new genetic knowledge, and new clinical indication, where the healthcare provider interacts geneticist for new interpretations of previous results. .

5.3. Scenario 3: Cancer Profiling – Somatic Testing

5.3.1. Description of Primary Clinical Sequence Workflow – Somatic Testing

Somatic testing is often times very different and can be more complex than germline testing, with more input from a range of doctors and healthcare professionals. The first difference being that patient consent to somatic testing is not always required, and so a test will be ordered either by a treating physician or as part of a reflex testing pathway (see Section 5.3.2). Additionally, pathology plays a vital role in cancer profiling. The same variant identi-

fied in different cancers has different clinical implications. Thus, a somatic specimen obtained by a biopsy or surgery will be analyzed by a pathologist to provide a diagnosis before moving onto a molecular laboratory for genetic sequencing. A clinician may also recommend germline testing or panels to test for specific germline variants (e.g. MLH1 in MSI-high colorectal cancers) or to compare a somatic sequence to the germline. It is important to note here that germline and somatic testing may not occur at the same time and are not processed at the same time through the bioinformatics pipeline. In most cases, somatic testing will be done prior to germline testing. Throughout the analysis and interpretation of the somatic specimen, the EHR can be updated accordingly. After the report is compiled, the laboratory will ideally complete a College of American Pathologists (CAP) biomarker reporting template for the physician for applicable tumor types. A workflow on the next page shows the primary workflow for somatic testing of a new patient. Optional modifications to the workflow are shown in dotted lines. For more information on specimens within the following workflow, see Section 5.1.2.

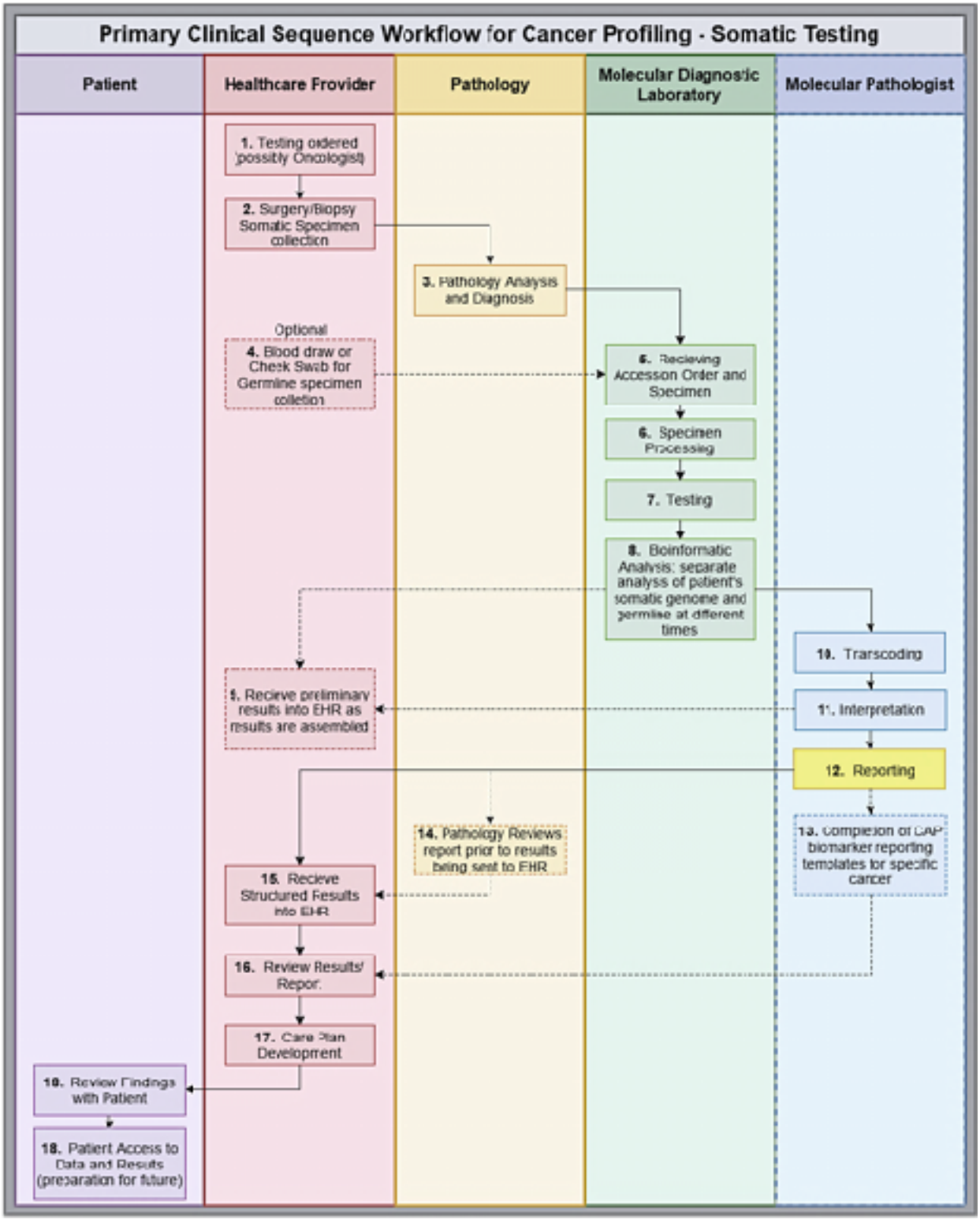


Figure 5.3-1: General workflow of somatic testing: 1. Sequence testing is ordered either by a physician or by an oncologist. 2. Suspected tumorigenic cells are identified and a specimen is collected by a clinician/surgeon. 3. The somatic specimen is sent to pathology where it is tested and analyzed for proper diagnosis. 4. (Optional) Blood is drawn or cheek swabbed for cells containing DNA for germline testing. 5. Laboratory receives specimen(s) and an order for genetic testing, including relevant data to aid in evaluation and interpretation of findings: indication for testing,

cancer type, and relevant clinical/pathological data for the patient. 6. Specimens are processed (e.g. DNA extracted) and prepared to be loaded on the sequencing instrument (both germline and somatic). 7. Specimens are sequenced. 8. The data for the somatic testing from the instrument passes through a bioinformatics pipeline for data processing: alignment and identification of sequence variants, as well as quality assurance. If germline sequencing is required as well, it is done at a separate time. 9. (Optional) Preliminary data is entered into the EHR for care during the analysis and interpretation process. 10. During the 'Transcoding' process, raw genomic data is transformed from bioinformatics format into healthcare IT data standards. 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 applications, as well as be associated with phenotypic data. 11. Genetic results are interpreted for clinical implications. 12. 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. 13. (Optional) The data is also put into a College of American Pathologists (CAP) biomarker report template depending on the cancer type. 14. (Optional) Pathologist (molecular and anatomic) review the report before results are sent into the EHR. 15. Genetic report and structured results are received in the Electronic Health Record (EHR) which may include variome. 16. Clinician reviews the results/report/CAP report. 17. Clinician develops (or modifies) a care plan taking into consideration the genetic findings. 18. Clinician reviews the genetic findings and care plan with the patient. 19. Genetic results are made available to the patient in the web-based patient portal

5.3.2. Alternative Workflows – Somatic Testing

Due to the complexity of care surrounding cancer, often times a different workflow is used or is necessary to provide better or continued care. For example, a patient may have had a biopsy done at a different site before going to a new hospital.

5.3.2.1. Alternate Workflow 1: Referral

In the case of a referral, a patient has already been seen, had their tumor biopsied, and possibly started treatment. It is not uncommon for a patient to have samples at several different hospitals, and in this case the referral hospital would want to compile this information and perform testing prior to the patient coming in. Upon scheduling an appointment, the hospital will reach out to previous care providers and request specimens to be reviewed at the current hospital. This review may include the determination that the patient would benefit from molecular testing, at which point there are two choices: 1) extract DNA from archival tissue, most commonly formalin-fixed paraffin-embedded (FFPE); or 2) obtain a new tumor specimen for testing. It is not uncommon for years to elapse between an initial diagnosis and a recurrence, and if the patient has not yet received targeted therapy or had the original tumor sequenced, this would be a scenario where FFPE would be tested, in many cases.

5.3.2.2. Alternate Workflow 2: Pathologist Ordered Testing

It may be the case that an oncologist or physician may not initially see the need of somatic testing. However, upon examination by a pathologist, somatic testing could be seen as necessary and be ordered by the pathologist.

5.3.2.3.Alternate Workflow 3: Patient Ordered Testing

In some rare situations, a patient may want to obtain their tumor specimen and submit it for somatic testing.

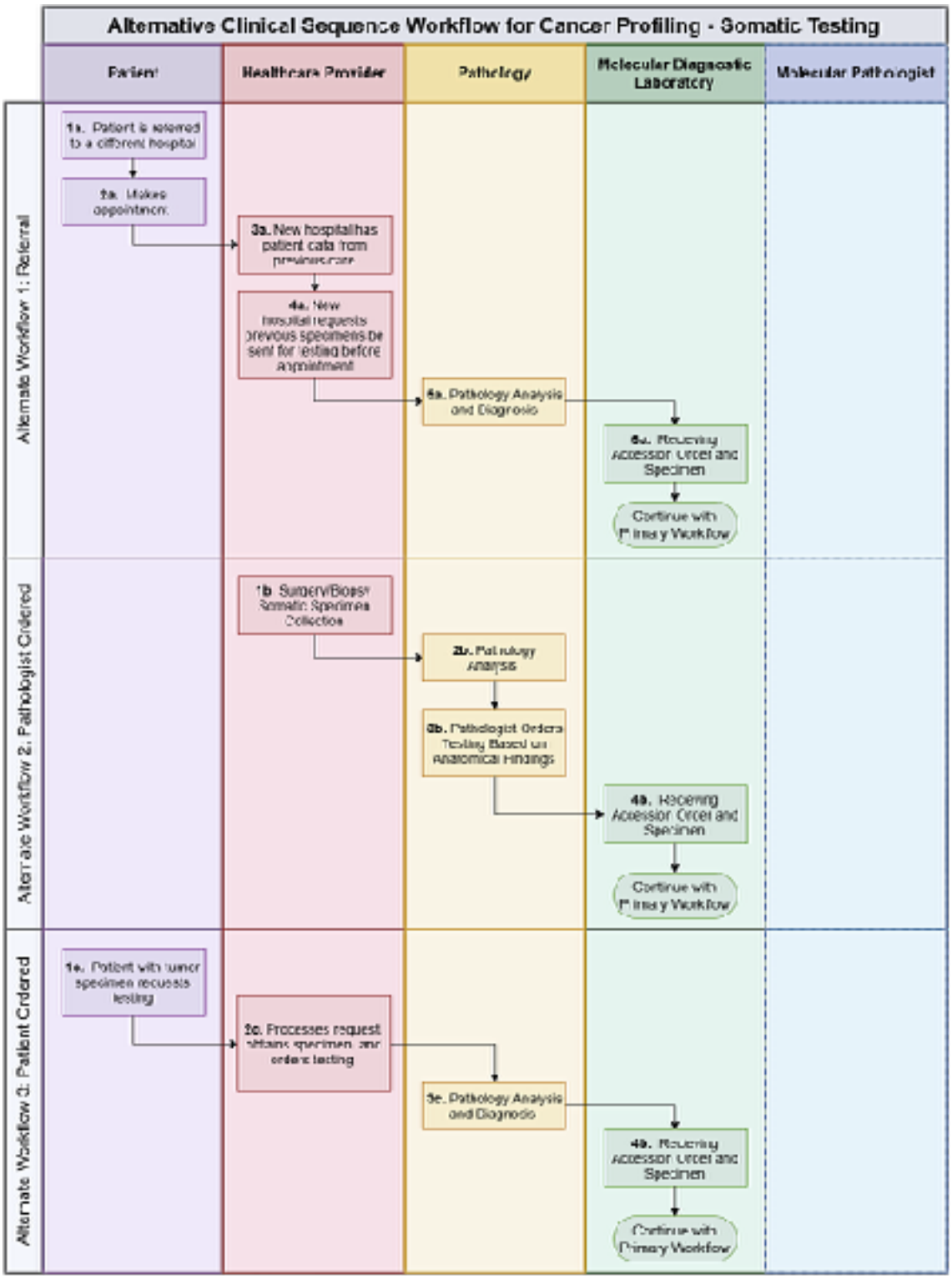


Figure 5.3-2: Alternate workflows for somatic testing. The **first alternate** shows a referral: 1a. The patient is referred to another hospital. 2a. The patient schedules an appointment with the new oncologist. 3a. The new hospital has the data and records of the previous care. 4a. A request is made for somatic specimens from previous hospitals to be rerun prior to the appointment. 5a. The specimen is received and sent to pathology before being sequenced. 6a. The specimen is sent to the lab along with an order containing pertinent information, and then continues the primary workflow. The **second alternate** shows the process for a pathologist ordered sequencing test. 1b. A suspected tu-

morigenic specimen is collected via biopsy/surgery. 2b. The pathologist analyzes the suspected tumor. 3b. Based on anatomical findings and in order to aide in diagnosis or treatment, the pathologist orders somatic testing of the specimen. 4b. The laboratory receives the pathologist's order and continues with the primary workflow. The **third alternate** shows the process for a patient ordered sequencing test. 1c. The patient with tumor specimen requests testing. 2c. Hospital and physician processes the request, receives the specimen, and orders the testing. 3c. The pathologist analyzes the specimen and makes a diagnosis. 4c. The laboratory receives the order and continues with the primary workflow

5.4. Scenario 4: Decision Making Tools – Family History and Drug Dosage Calculators

Genetic sequences coupled with more traditional clinical methods can lead to better decision making through the utilization of family history tools, risk assessment tools, and drug dosage calculators. In some cases, clinicians translate (i.e. manually reenter) genetic data into tools for decision making, but in other cases, patient genetic data from the EHR is automatically incorporated into clinical decision making tools.

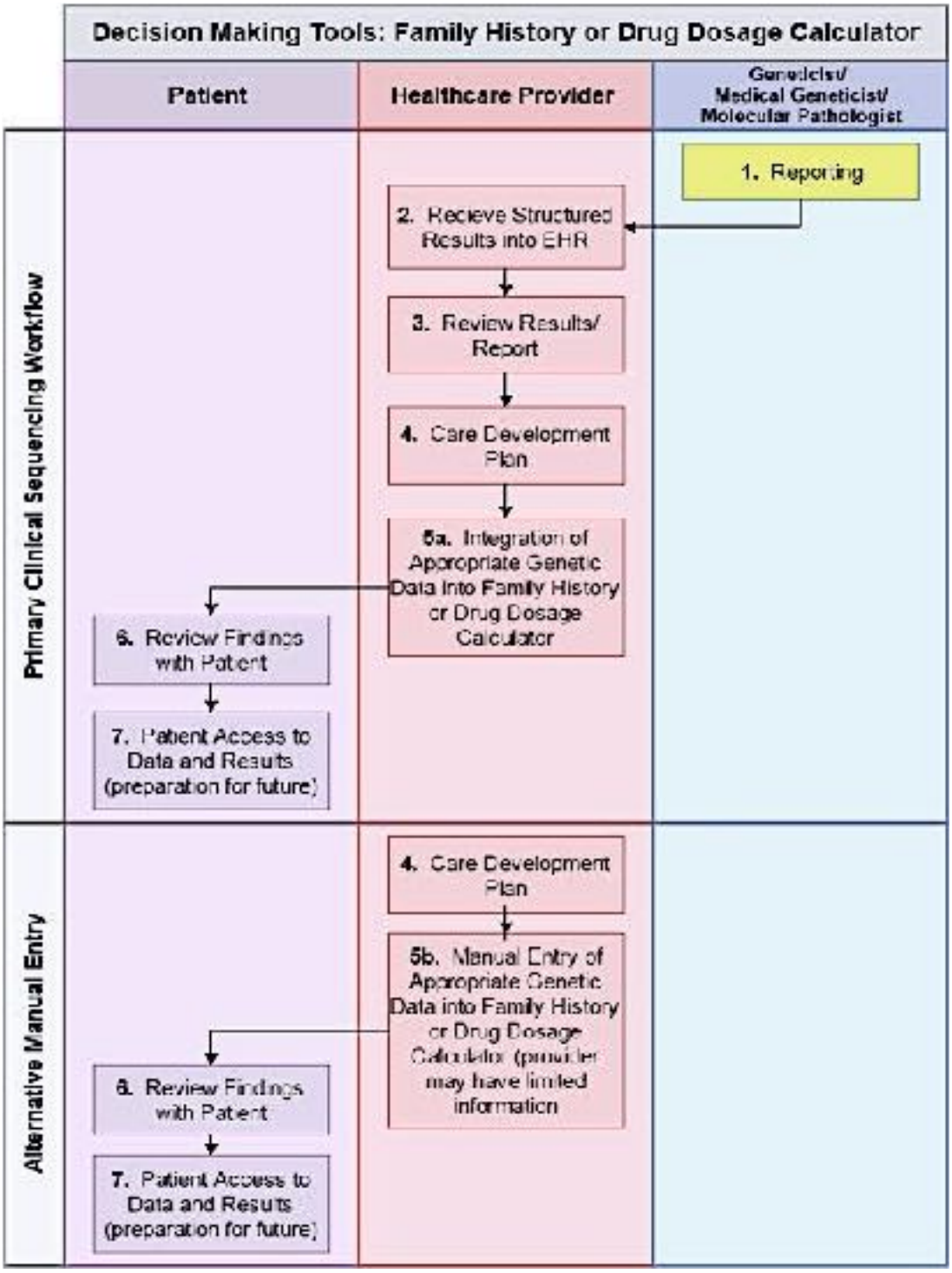


Figure 5.4-1: Primary workflow for automatic integration of genetic data into clinician decision making tools. 1. The genetic report is compiled. 2. The results are received into the EHR; in the future these results may contain variome. 3. The results are reviewed by clinician. 4. A care development plan is developed. 5a. The results are then integrated into the family history, drug dosage calculator, or other decision making tools automatically. 5b. These tools are used by the manual entering of data, and provider may have limited information. 6. The genetic and decision making tool results are reviewed by the patient. 7. The results are made available through web based portal.

5.5. Scenario 5: Public Health Reporting

5.5.1. Description of Public Health Reporting Scenario

Today, registrars manually translate clinical data into public health reporting systems (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5346201/), which can be time consuming and complicated due to different standards across testing platforms. This data is used to monitor and improve public health (e.g. surveillance and clinical research). After the genomic data is reported to the public health reporter, the relevant data will be extracted primarily through the manual process of chart review in order to be incorporated into the Public Health data repository. In the future, this data will be extracted from the EHR in an automated (or semi-automated) manner.

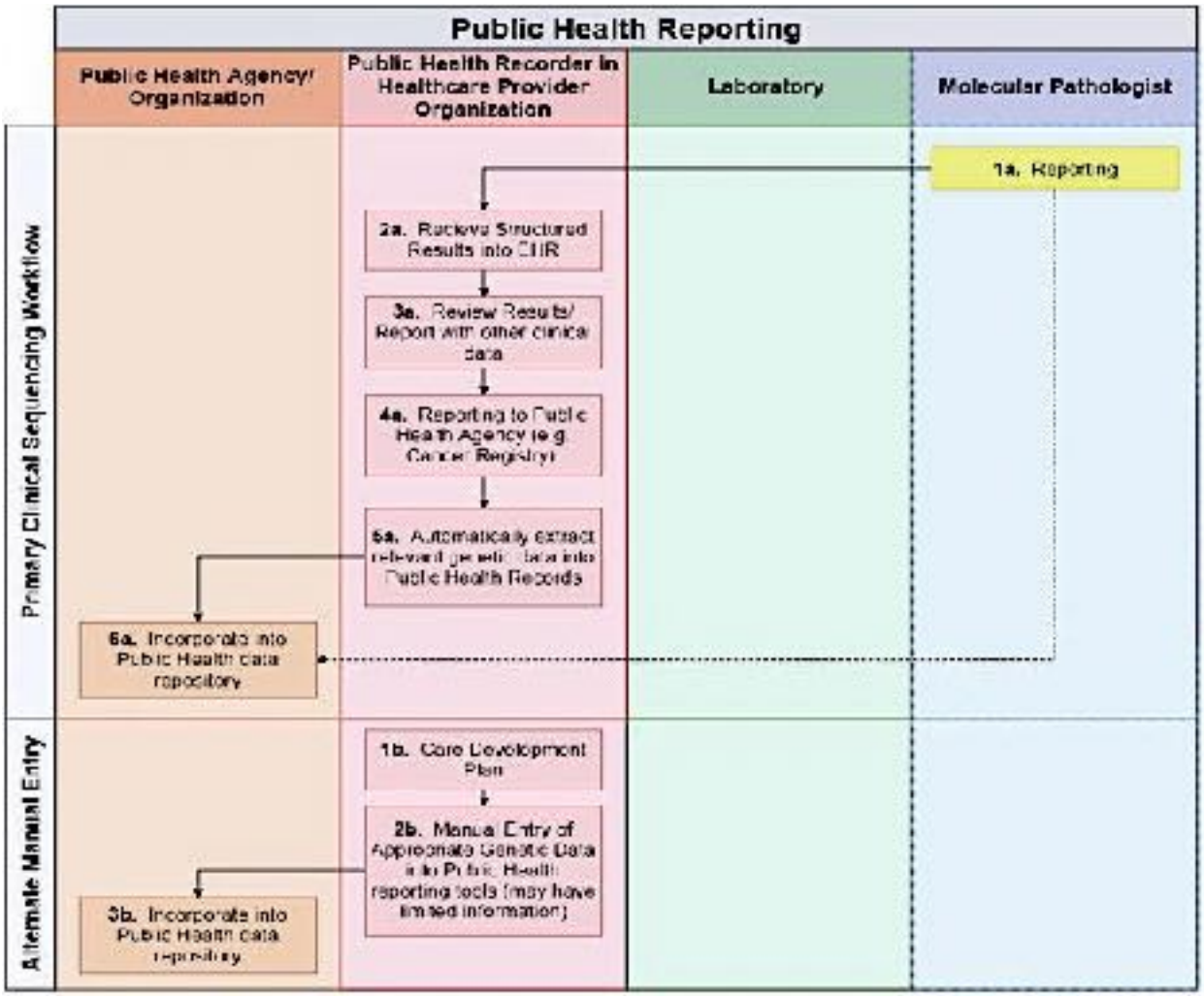


Figure 5.5-1: Public health recording of relevant genetic data. **Automatic**: 1a. Genetic report from molecular pathologist is shown (an alternate route shows data being automatically integrated from the report into the Public Health repository). 2a. The structured results are integrated into the EHR of the patient. 3a. The results are reviewed by the clinician, as well as being compiled with other clinical data of the same cohort. 4a. Results are reported to public health agencies like the cancer registry. Currently, it takes about 6 months where the report is manually reviewed by professionals due to differing standards and then entered. 5a. The pertinent data is automatically extracted into public health records. 6a. The data enters a public health data repository where it can be stored and used in the future. **Manual**: 1b. The care development plan is created. 2b. The appropriate genetic data is manually

entered into Public Health reporting tools. This step is very time consuming and is currently being done. 3b. The data is then incorporated into the public health repositories.

5.5.2. Cancer Registry workflow

Cancer registrars perform patient chart review translating and summarizing clinical information into public health reporting systems. Currently, the requirements for reporting genetic/genomic information are minimal but are likely to rapidly expand. There are numerous challenges associated with this process of genetic reporting to cancer registrars.

Genetic test results are inconsistently reported, and these inconsistencies are due to a number of factors. Frequently there is a lack of adherence to the guidelines of medical professional organizations like the College of American Pathologists (CAP) and American College of Medical Genetics and Genomics (ACMG). Compounded with the lack of adherence is the fact that the granularity of results is tied to the specific testing platform, and there is no known mapping that 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
- Sanger 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 dependent 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.

5.6. Scenario 6: Clinical and Research Data Warehouses

Electronic health records (EHR) 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 does not meaningfully support these activities (see Figure 5.6-1).

Health data warehousing should persist data in standardized formats, while allowing users to export subsets of the data for specific use cases, analyses, or reporting needs. Warehouse data should be represented in the richest form possible using generic standards, while each data subset is optimized for a specific use case, e.g. clinical research, public health registrars, or even EHRs. In this way, all different ‘views’ of the data are based on the same standardized semantics, thus achieving consistency and interoperability while avoiding data loss through transformations and duplication of data.

Additionally, as many applications of genetic data are designed for research applications that utilize data structures, such as variant call files (VCFs), the data stored within the data warehouse should be convertible to these structures for the broadest potential secondary

use. If the clinical genetic data cannot be converted directly, tools should be available that can convert it to other data structures.

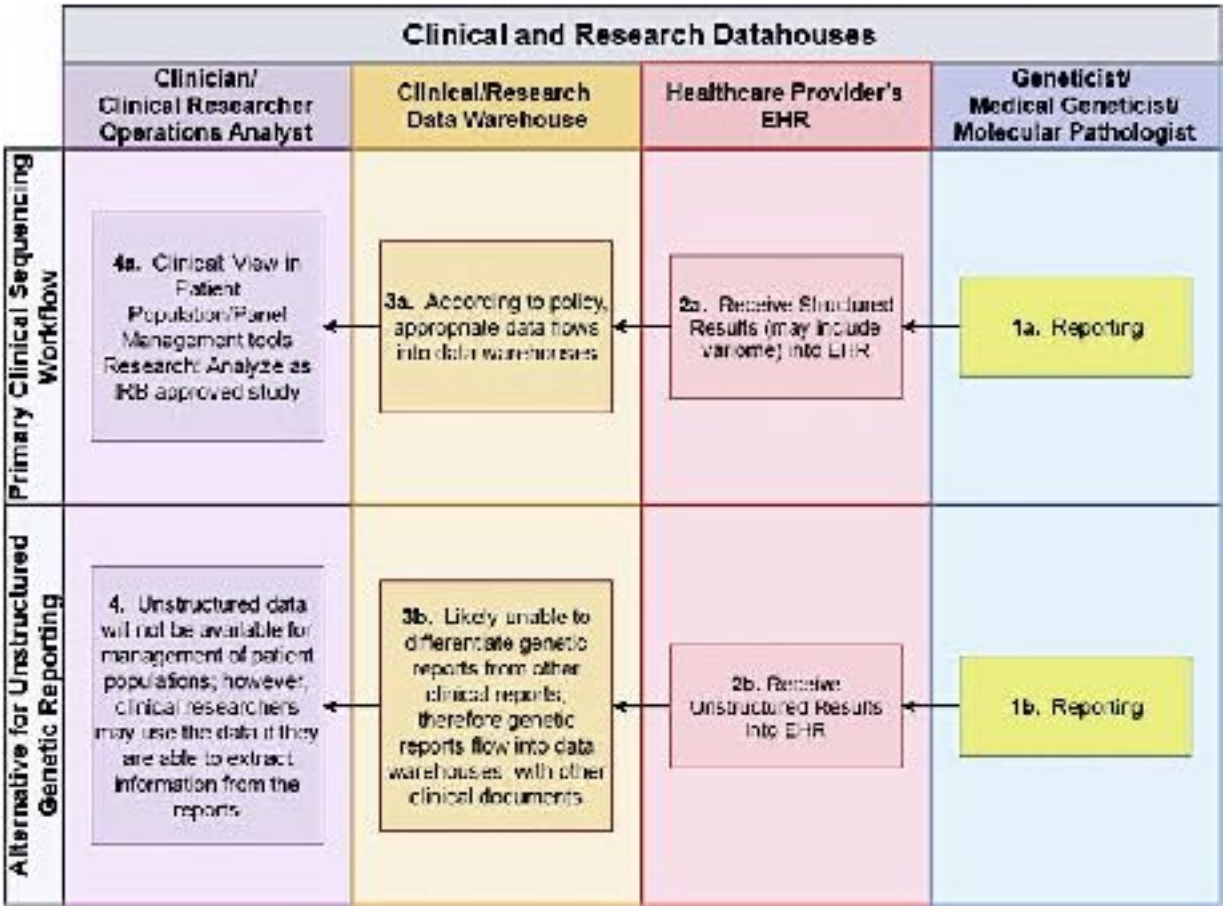


Figure 5.6-1: Example of structured and unstructured genetic reporting influence data warehouse incorporation. After reporting (1), results can be received into the EHR as structured (2a) or unstructured (2b) data. In the structured scenario: 3a. the appropriate data is relatively easily pulled from the EHR and flows into the research or clinical warehouse. 4a. The structured data then allows ease and enhancements for patient care through the use of management and organizational tools, or in research can be used to set up or approve a study. It should be noted here that this workflow is an example and may be different depending on the consumer and use case. In unstructured: 3b. the clinical or research warehouse is unable to pull the necessary genetic information which becomes entangled with other data and documents. 4b. The unorganized data can no longer be utilized in patient management or for research purposes.

5.7. Scenario 7: Cytogenetic Marker identification via sequencing

Cytogenetic testing, often referred to as karyotyping, investigates numerical and/or structural chromosome abnormalities during cell metaphase. Serving as the standard for genetic testing, cytogenetic testing should ultimately flow into the sequence pipeline. Cytogenetic testing serves as a traditional approach that is FDA approved and well established, and will be used long-term. Detailed methods include the following: Tissue samples are sent from the clinician to the laboratory for chromosome harvesting, banding, microscopic analysis, and karyotype production. Several methods are utilized in molecular cytogenetic testing, such as metaphase, interphase, dual-color/fusion, and dual-color/break-apart (cytogenetic methods and information). These methods use fluorescence microscopy to assess for the presence, absence, relative positioning, and/or copy number of specific DNA segments.

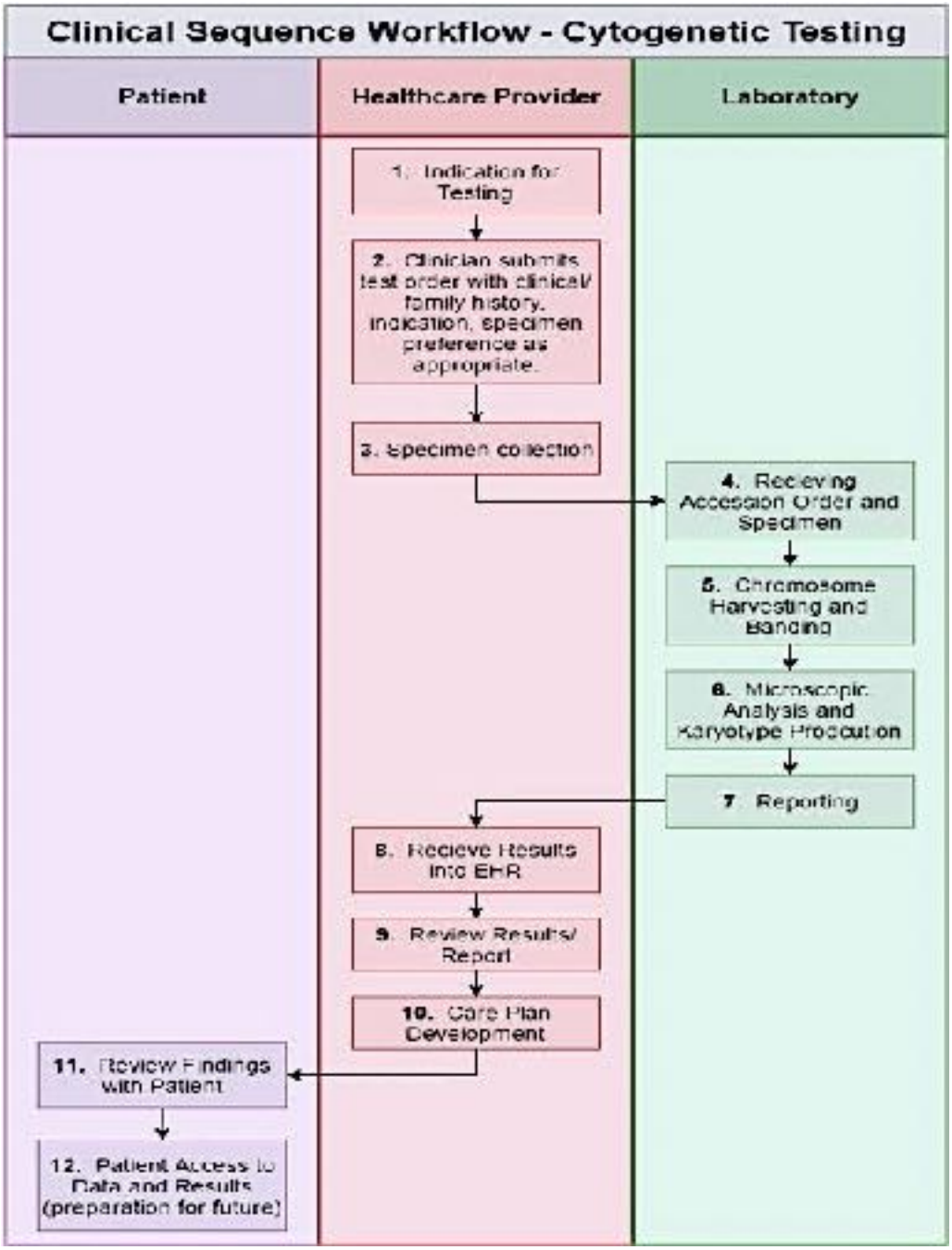


Figure 5.7-1: Cytogenetic testing. 1. Clinician recognizes an indication for cytogenetic testing. 2. The clinician orders testing with appropriate information for the laboratory. 3. Cell samples are taken and sampled for cytogenetic testing. 4. The order is processed at the laboratory with specific instructions. 5. The chromosome is harvested from the sample and banding (staining) is done for further analysis. 6. Analysis of the banded chromosome and karyotype for number or structural abnormalities. 7. A report is compiled on the findings. 8. The results are received into the EHR. 9. The report is reviewed by the physician and 10. A care plan is developed. 11. The physician reviews the results with the patient/parents. 12. The results are available via a web portal.

5.8. Scenario 8: Pharmacogenomics

5.8.1. Description of Scenario

Pharmacogenomics (PGx) is the study of the ways in which genes will impact an individual's response to drugs. In order to determine how a patient will react to a medication, researchers are currently studying genetic differences and how these differences will affect the body's response to drugs (NIH PGx Information). Given a patient's disease, doctors may utilize pharmacogenomic information to determine which drug to prescribe (Moen, Godley, Zhang, & Dolan, 2012; Farrugia and Weinshilboum, 2013; Bielinski, Olson, Pathak, Weinshilboum, et al, 2014). For example, the FDA currently recommends genetic testing before the administration of the chemotherapy drug mercaptopurine. In addition to mercaptopurine, the FDA has a table of over 200 approved drugs that have known genetic interactions, which lists with their associated therapeutic areas and biomarkers (FDA Table). Thus, after diagnoses, doctors will transmit patient information to the geneticist, or clinical pharmacologist, who will analyze genomic data. After the analysis, the doctors will be able to effectively prescribe the drug with minimal risk of adverse side effects (de Jesus Castillejos-Lopez et al. 2006).

The Clinical Pharmacogenetics Implementation Consortium (CPIC) has been developing clinical guidelines from the large amount of information on pharmacogenomic results to facilitate the clinical implementation of pharmacogenomics (CPIC Guidelines). In pharmacogenomics, the use of sequencing assays is increasing, and with it there is an increased need for messaging standards to accurately move information around and to transform raw data into application and site specific assorted data structures.

Furthermore, pharmacogenomic research and development has moved on from early entrepreneurial investors to clinical organizations, the implications of which mean that pharmacogenomics and its tools must be tightly integrated into the work of physicians. In order to integrate fully with practicing physicians:

1. Genomic test results must include both discrete data and a human-readable text report. At this time most EMRs are not sophisticated enough to be able to store discrete genomic data, but EMR capabilities will expand in the future.
2. Decision support alerts, recommendations, and educational material should be designed to minimize the learning curve for those that do not have a background in genetics or pharmacogenetics
3. Once the data is processed, there must be a way to reinterpret the results over time.
4. There should be a protocol or system to see if previous genetic results are currently relevant to a physician, either automatically or at the physician's discretion. The system should enable application providers to determine if a genetic test is necessary and what biomarkers or genes are to be analyzed for a specific drug.

Going forward, there are two important aspects that should be harmonized to make progress. The first is to provide a link to go back to variants that have been tested and sequenced (i.e. it must be possible to uniquely and unambiguously identify a genetic variant/allele). The other is to ensure that the linking of a gene, or combinations of genes, to recommendations is consistent across standards and platforms.

5.8.2. Pharmacogenomics – Somatic Profiling

Pharmacogenomics can be particularly useful in the development and implementation drugs that target variant specific biomarkers. A test of a patient's genome reveals the presence of biomarkers for the pharmacist, which are included in the EHR and are available for the pharmacist to view. From the report of the biomarkers, a prescription recommendation can be made for the specific patient and filled.

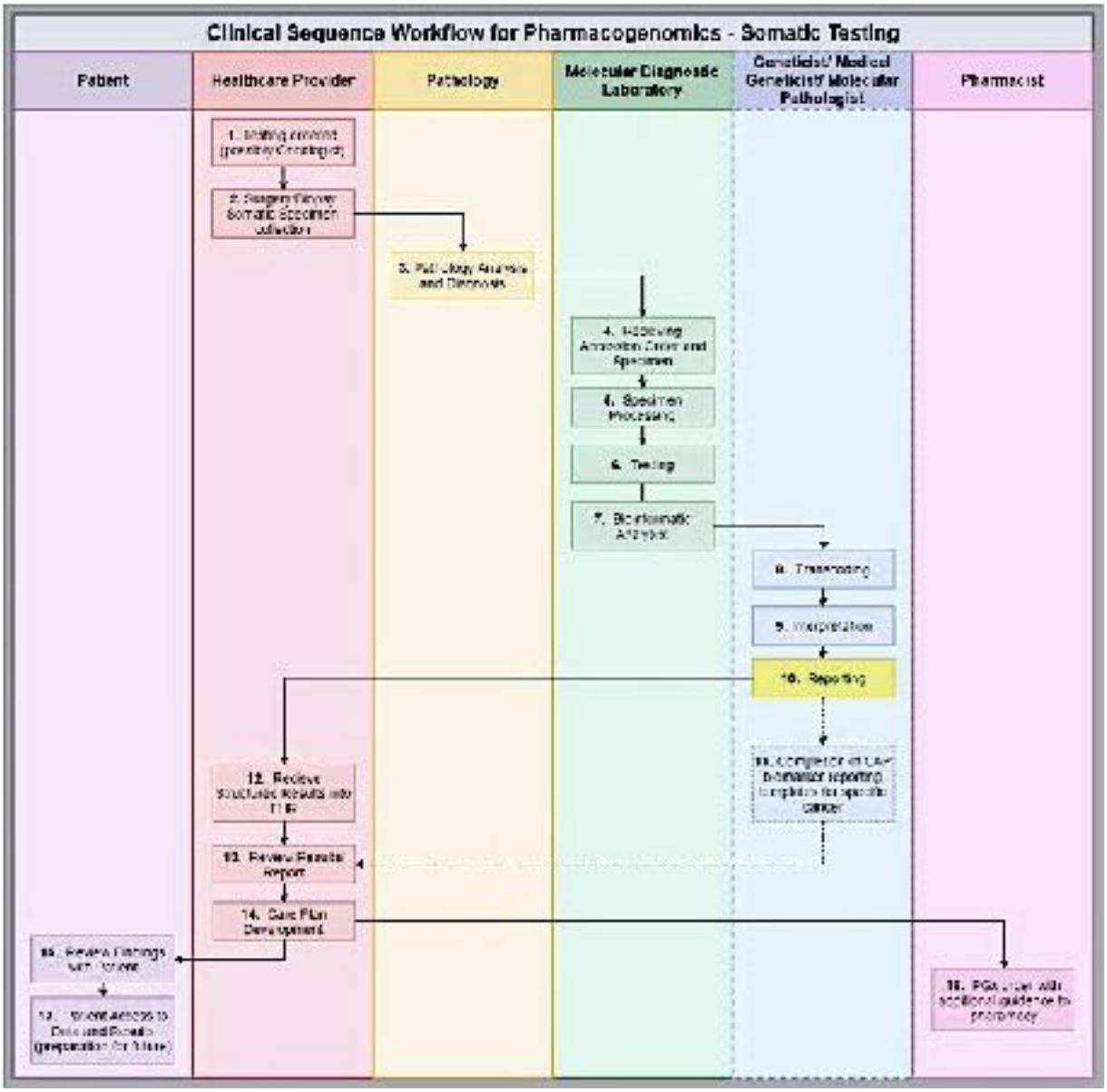


Figure 5.8-1: Pharmacogenomics workflow for somatic testing. Pharmacogenomics can be used to identify biomarkers in tumorigenic cells that can be targeted by certain drugs. Overall the workflow is very similar to the primary somatic testing workflow (Figure 5.3-1). Pharmacogenomics specific steps: 16. An order with detailed instructions is sent to the pharmacist to prepare. Additionally, the pharmacist is able to view the EHR and the biomarkers.

5.8.3. Pharmacogenomics – Germline

5.8.3.1. Primary Germline Pharmacogenomics Germline Testing Workflow

Germline testing for pharmacogenomics is similar in that there is a test for specific biomarkers which can be used to guide prescription decisions either at present or in the future.

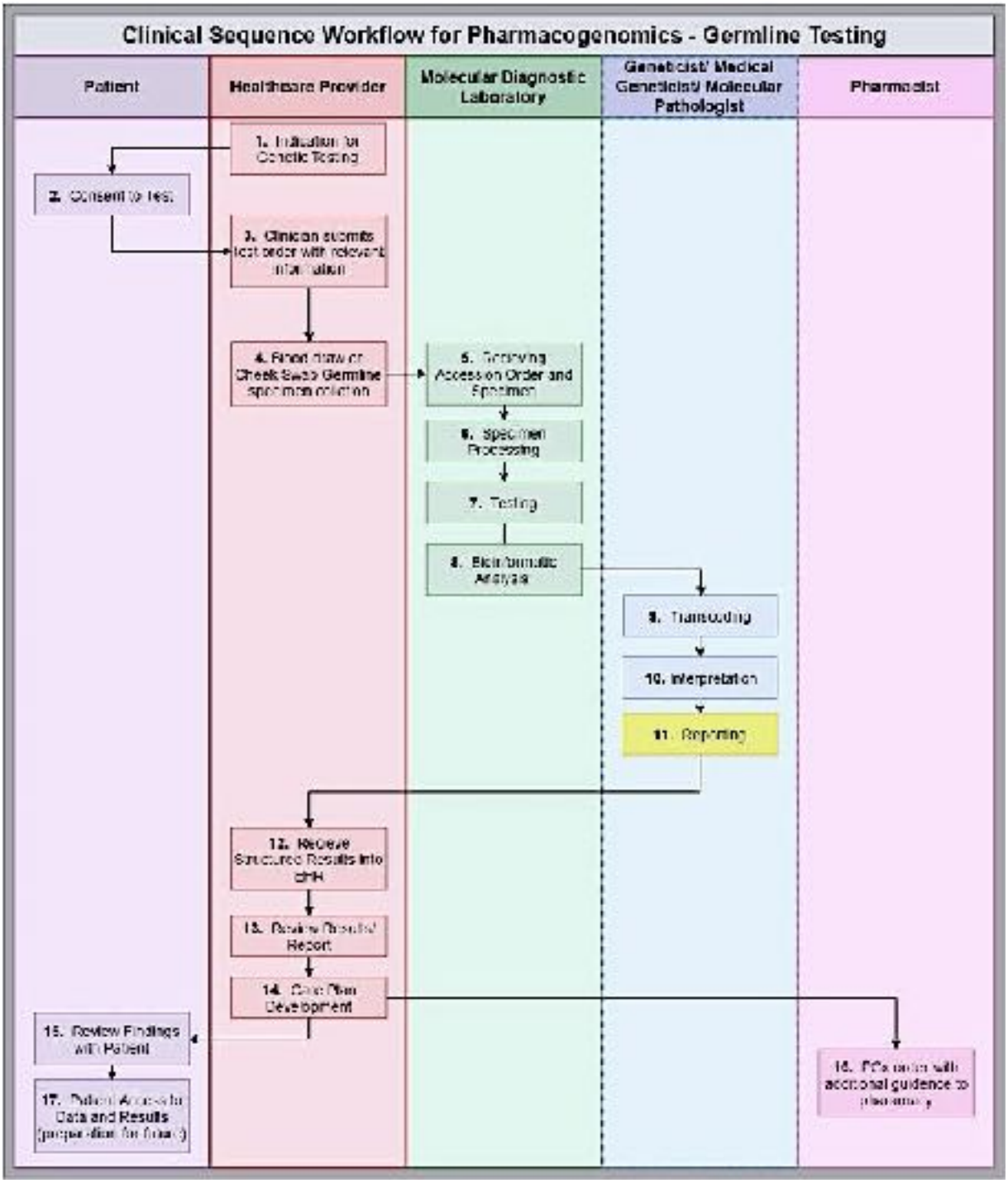


Figure 5.8-2: Pharmacogenomics workflow for germline testing. Overall the workflow is similar to germline testing (Figure 5.2-1). Pharmacogenomic specific steps: 16. An order with detailed instructions is sent to the pharmacist to prepare. Additionally, the pharmacist is able to view the EHR and the biomarkers.

5.8.4. Alternate Germline Pharmacogenomic Workflow – Pharmacist Involvement

There are scenarios in pharmacogenomics where a pharmacist may assist in the production of a pharmacogenomic recommendations for the general report, where the structured data

can be seen by the pharmacist. This is not necessarily typical, but may be relevant to pharmacists trained in pharmacogenomics.

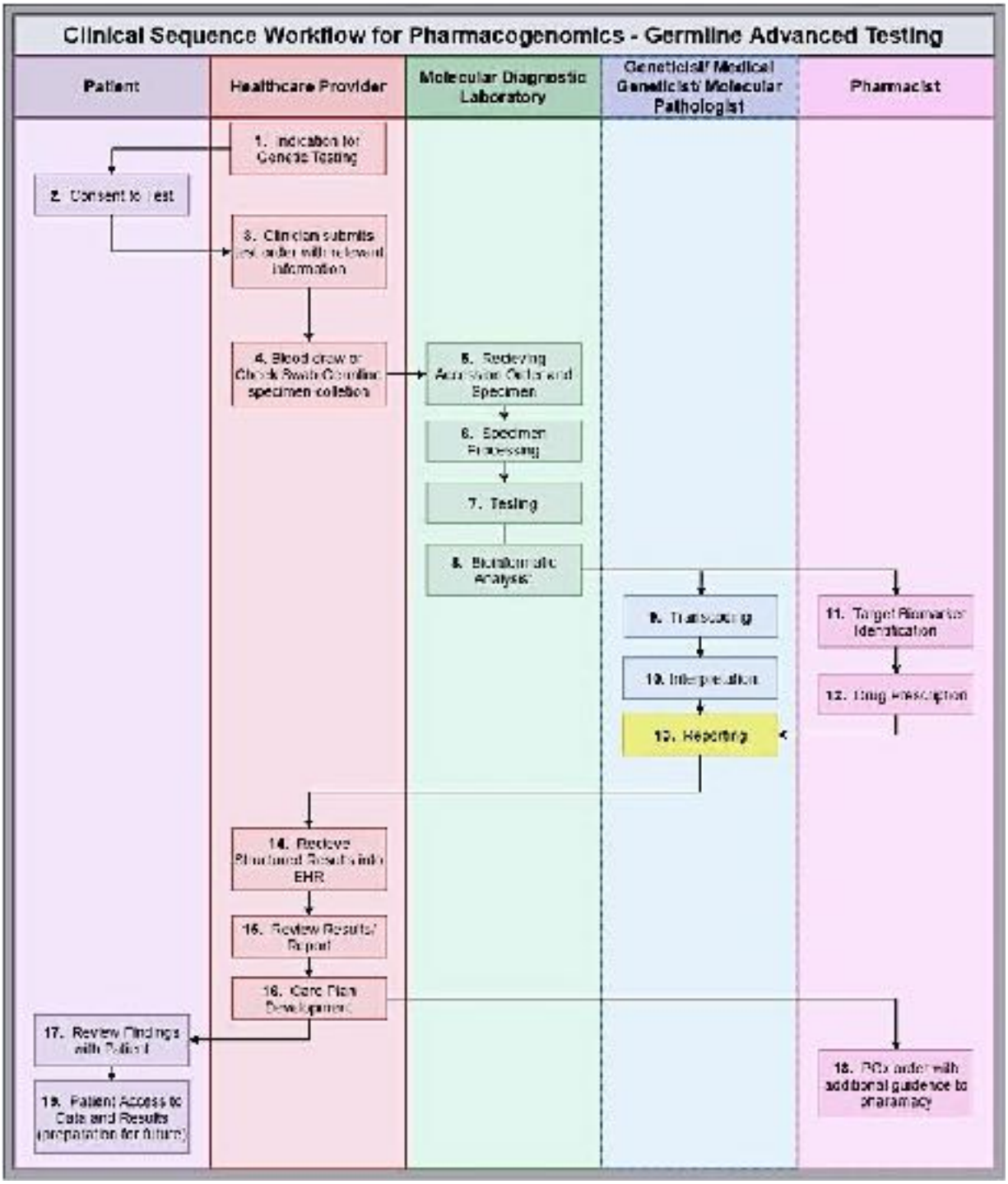


Figure 5.8-3: An advanced pharmacogenomics workflow. Most of the workflow is similar to germline testing (Figure 5.2-1). Pharmacogenomics specific steps: 11. A pharmacist trained in pharmacogenomics, using the germline sequence, identifies possible biomarker targets for drugs. 12. The pharmacist then makes a drug prescription recommendation for the report. 18. An order is sent to the pharmacist for a prescription with instructions and access to biomarker data.

5.10. Scenario 9: State & Regional Health Information Exchanges (HIE)

State and regional Health Information Exchanges (HIE) are becoming an important part of the healthcare ecosystem by improving accurate exchanges of information across a network of organizations. As utilization of cloud-hosted software-as-a-service (SaaS) solutions becomes an integrated part of the healthcare business model, there is an interesting possibility for hosting systems to increase data interoperability of genetic/genomic data. That is, we have to accommodate expanded interoperability architectures other than standard messaging from point A to point B.

5.11. Scenario 10: Human leukocyte antigen (HLA) Typing

5.11.1. Summary of Challenges

Unlike standard genomic testing, human leukocyte antigen (HLA) typing offers unique challenges:

1. Genomic regions of interest are not included within a genome build; therefore, using only a genome build and chromosome in conjunction with genomic location does not support HLA typing. Typically, NGS based typing of HLA is based on a combination of local assembly and alignment to reference alleles and/or genomes.
2. Clinical genetic standards for communicating a variant (e.g. HGVS) do not support the complexity of HLA typing; therefore, efforts were made to come up with adequate standards, e.g. in the US the National Marrow Donor Program has developed their own standard. These include a combination of domain specific nomenclature (IMGT/HLA), string based reporting of genotyping with full ambiguity (Tissue Antigens. 2013 Aug; 82(2):106-12. doi: 10.1111/tan.12150), and a XML based message structure called Histoimmungenetics Markup Language (HML).
3. Marrow donor nomenclature is based on allele naming and continues to evolve as more is understood and technology platforms are capable of increased detailed detection.
4. Systems must support different versions of the marrow donor nomenclature and various degrees of ambiguity, for backwards compatibility.

5.11.2. Background on NMDP

The National Marrow Donor Program (NMDP) is a nonprofit organization that matches patients needing life-saving stem cell transplants with potential donors. To achieve this mission, it operates the Be The Match donor registry that currently stores tissue typing data from more than 11 million potential donors. The most important factor in matching a patient with a donor is HLA.

HLA Nomenclature and official allele designations are assigned by the HLA Informatics Group (HLA Information), on behalf of the WHO Nomenclature Committee for Factors of the HLA System, the KIR Nomenclature Committee and the nomenclature committees set up by the International Society for Animal Genetics (ISAG) (http://hla.alleles.org). This work is overseen by the Comparative MHC Nomenclature Committee and is supported by ISAG and the Veterinary Immunology Committee (VIC) of the International Union of Immunological Societies (IUIS). The HLA Informatics Group has registered an OID node (2.16.840.1.113883.13.252) for developing External Value Set for these allele designations and is currently working out the details for OID association with each HLA allele. A specialist

database for HLA sequences has been established (http://www.ebi.ac.uk/ipd/imgt/hla/) and includes the official sequences for the WHO Nomenclature Committee For Factors of the HLA System. The IMGT/HLA Database is part of the international ImMunoGeneTics project (IMGT).

Recently, the Immunogenomics community has gathered to develop standards for recording and reporting NGS based genotyping of HLA (ngs.immunogenomics.org). The goals of these meetings have been to identify the **M**inimum **I**nformation for **R**eporting **I**mmunogenomics **N**GS **G**enotyping, aka MIRING (HLA and KIR genotyping with NGS) based on the principles of MIBBI (Minimum Information for Biological and Biomedical Investigations). The MIRING identifies ten principles for NGS based genotyping of immunogenomics data. These include:

1. Sample Annotation
2. Reference Context
3. Full Genotype
4. Consensus Sequence
5. Unreferenced Sequences
6. Novel Polymorphisms
7. Sequence of Regions Targeted
8. Read Metadata
9. Primary Data
10. Platform Documentation

Items 1-6 are considered method independent and dynamic with each report. Items 7-10 are static in nature, and dependent on the specific NGS methodology employed, and could be externally referenced through a resource such as the NCBI Genetic Testing Registry or Sequence Read Archive.

5.11.3. HML and HL7

While the MIRING provides principles and guidelines, it doesn't provide a technical specification for the message. Together in collaboration with vendors and the immunogenomics community, NMDP is enhancing HML (Histoimmunogenetics Markup Language) to meet the principles of the MIRING (HML Information). While HML has been developed outside of HL7, it currently serves the purposes of the immunogenomics community. However, the community recognizes the need for interoperability with the larger healthcare community and the potential to interface with EMR systems. In light of this, the possibility to encapsulate HML in HL7 messages or structured documents is being explored, working closely with the HL7 Clinical Genomics Work Group.

6. Additional Use Case Scenarios

The following use cases should be considered in standards development and implementations. These additional use cases will be more fully described in future releases.

6.1. Comprehensive Pathology Report

For an increasing number of specimens (e.g. a bone marrow aspiration and biopsy) the specimen will undergo a series of tests, such as morphology, immunohistochemistry, flow cytometry, cytogenetics, fluorescence in situ hybridization [FISH], and molecular testing [e.g. NGS]). These tests will provide genetic information in the form of a comprehensive report,

created by pathologists. Challenges include integration of findings across multiple testing platforms and interpretation of these findings in creation of the comprehensive report. See Seegmiller et al. 2013 Am J Clin Pathol.

6.2. Rare/Undiagnosed Diseases

The diagnoses of rare diseases often can happen long after symptoms are first evident within affected individuals. Because of limitations on physician and patient access to the most up-to-date information about rare diseases, testing can be expensive and cumbersome. NGS is enabling rare and de novo variants to be found and associated with previously undiagnosed diseases. Comparison across several organizations (potentially spread across the country or world) is often needed to obtain enough power to associate variants with clinical findings.

6.3. Preimplantation Testing

Preimplantation genetic diagnosis (PGD) is utilized to test for conditions which could cause termination of the pregnancy or major genetic disorders. PGD is usually reserved for those with high genetic risk factors or those who are undergoing in-vitro fertilization (IVF), even with low genetic risk factors. Oocytes or early stage embryos are screened for major inherited conditions such as Huntington's disease, familial predisposition to cancer, Li-Fraumeni syndrome, or aneuploidy before continuing with IVF (Sermon et al. 2004; Boyle and Savulescu, 2001). Preimplantation screening is regulated in many countries ensuring screening for major, not minor, genetic conditions and to prevent sex-selective screening.

Globally, preimplantation genetic screening (PGS) accounts for more than all the other indications for PGD added together (Minasi et al. 2017). The combination of PGD and PGS improves the capability of detecting embryos with high potential to implant while simultaneously avoiding the transfer of embryos that, although free of monogenic disease or chromosomal rearrangements, carry chromosomal aneuploidies.

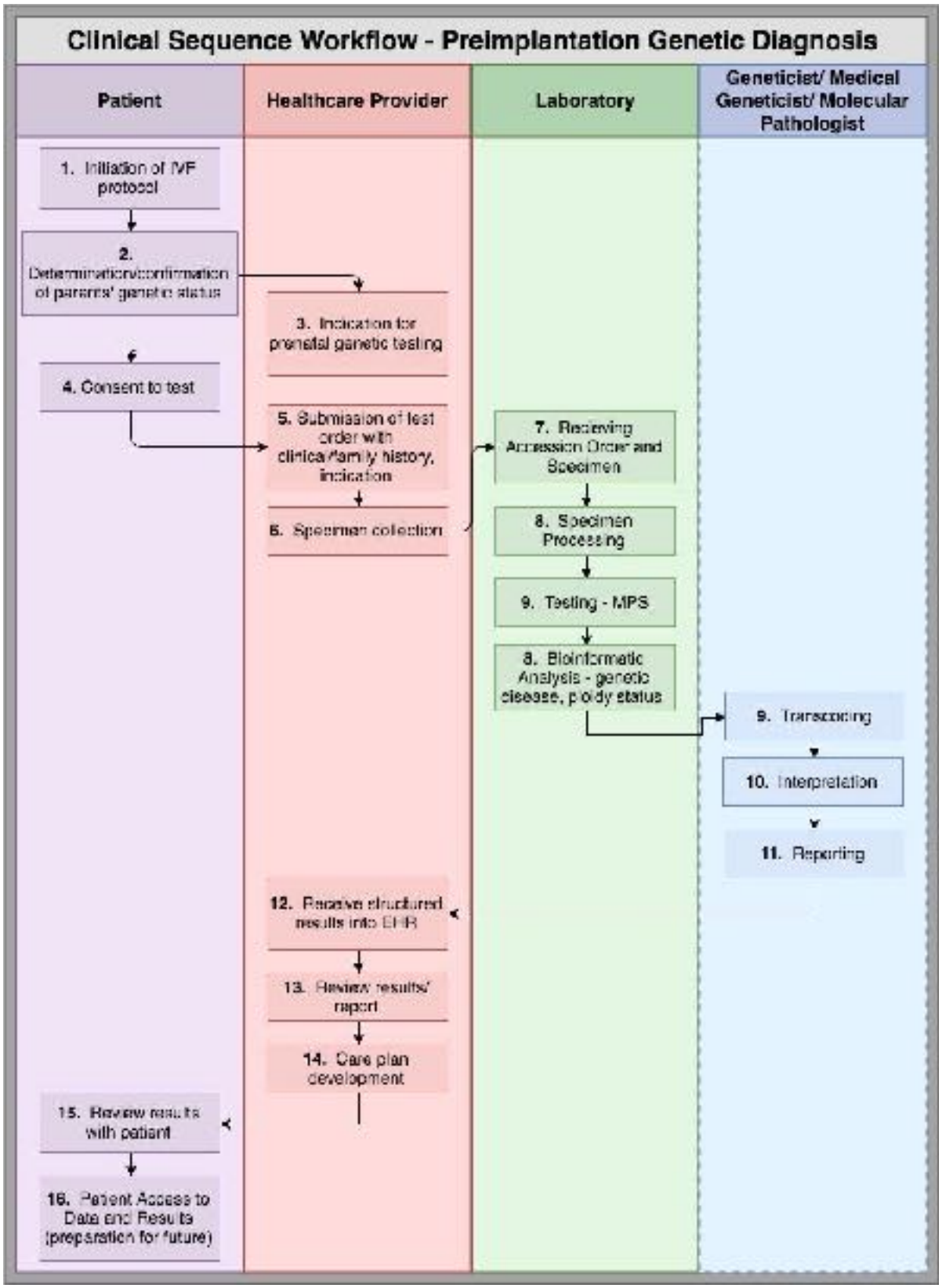


Figure 6.3-1: PGD workflow. **2.** Parental genetic analysis can identify mutant alleles in current and for future gestations. **6.** The use of NGS for PGD alleviates issues arising from embryonic chromosomal mosaicism. **8.** Embryo vitrifi-

cation and blastocyst biopsy. **9.** Array-comparative genomic hybridization (aCGH) which analyzes specific amplifications and deletions of the genome in high resolution regions.

6.4. Cell-free Fetal DNA (cffDNA) Based Noninvasive Prenatal Testing

Next Generation Sequencing can offer a method for non-invasive prenatal diagnosis of genetically inherited conditions such as β-thalassaemia and congenital adrenal hyperplasia, by enabling testing of cell-free fetal DNA (cffDNA) circulating in the maternal plasma. The test works by distinguishing cffDNA from maternal DNA by identifying paternally inherited alleles. Currently, cffDNA is used primarily to test for aneuploidy in high risk patients where it has been shown to be more reliable than serum testing (Bianchi et al. 2014). Further tests are still under development, but cffDNA testing provides the possibility of replacing invasive procedures like amniocentesis or chorionic villus sampling, which are risky for the patient and fetus. See Figure 6.4-1 for a workflow of this testing.

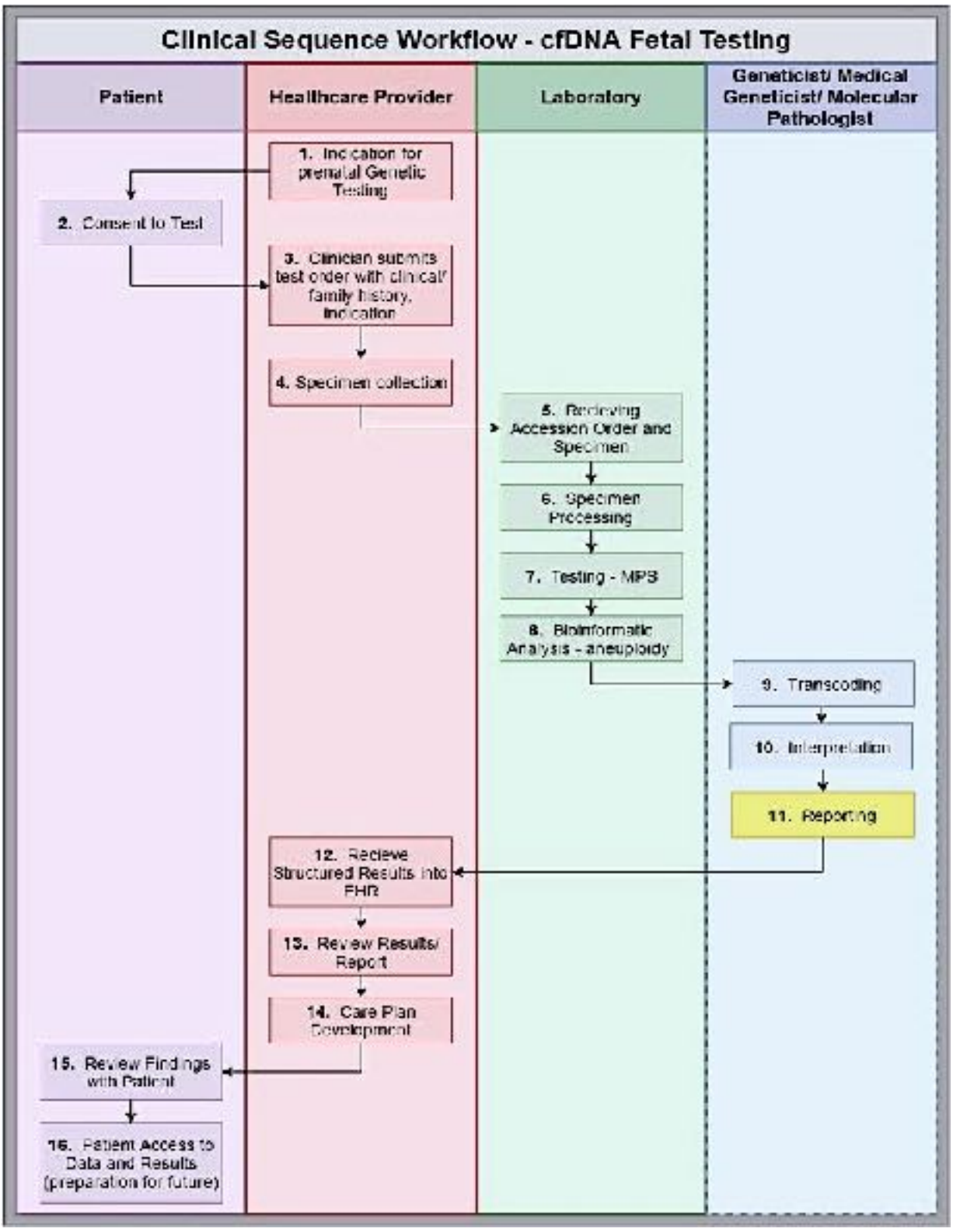


Figure 6.4-1: cffDNA fetal testing workflow. **7.** Massively parallel sequencing (MPS) is utilized. **8.** cffDNA must be differentiated from maternal DNA by recognizing paternal DNA sequences in cffDNA. MPS analyzed for aneuploidy as well as genetic disorders.

6.5. Newborn Screening

6.5.1. Current Newborn Screening

Newborn screening practices test newborns for certain harmful or fatal disorders that are not apparent at birth. The initial screening technique utilizes tandem mass spectrometry which can screen for 23 inherited metabolic disorders (e.g. maple syrup urine disease) with a single drop of blood (Schulze et al, 2003). It should be noted that screening through dried blood spots is primarily a profile of amino acids and acylcarnitines in the newborn's blood as opposed to genetic information. In the case of a positive result on the screening, a secondary metabolic screening (either another blood spot test or chromatography of a urine sample) and/or genetic sequencing can confirm a diagnosis.

However, the actual process by which newborn screening is carried out is logistically messy. Each state or region has its own testing done in a state run laboratory. In this case, the blood spot is taken by the hospital and sent to the state run testing laboratory. From there the results are faxed back to the hospital's pediatrician as opposed to the newborn's pediatrician. The flow of information from the state laboratory, to the hospital, to the pediatrician is far from an ideal system, however the relative rarity of a positive test result means that the system is not completely flooded. Storing residual dried blood spots and the data from newborn screening can cause further complexities with the public health authority (it is stored in the state laboratories currently), and there is an unresolved issue of locating the testing that was not ordered by the patient and its timing.

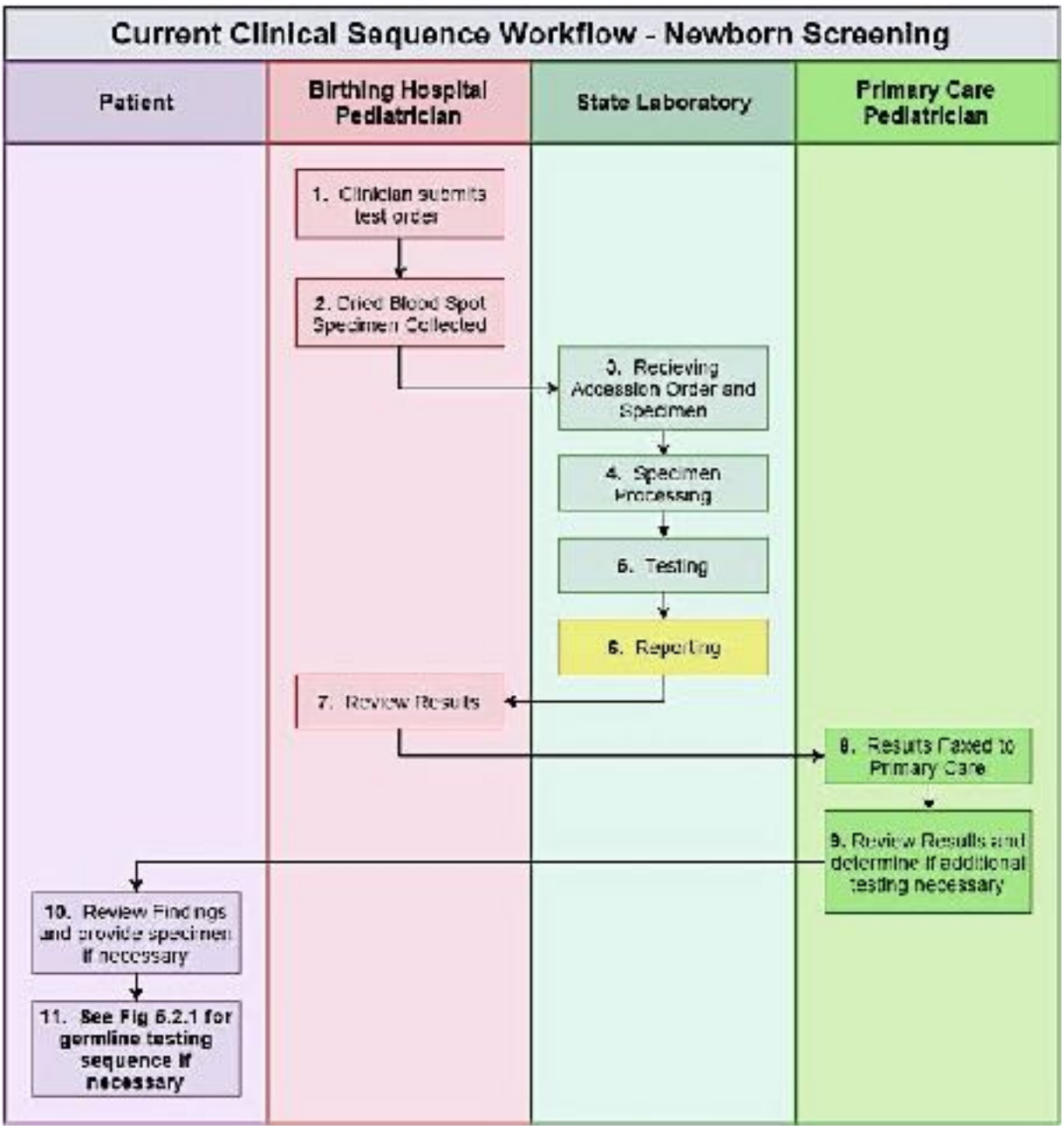


Figure 6.5-1: The clinical sequence workflow for newborn screening by dried blood spots. 1. A mandatory test is ordered by the pediatrician upon birth. 2. The dried blood spot (DBS) is collected from the heel, finger, or toe of the infant. 3. The specimen and order are received by the state testing facility. 4. The specimen is then processed within the specific state's system. 5. The blood is then tested by HPLC for amino acids, etc. 6. A report is compiled and sent to the birthing hospital via fax. 7. The hospital will review the results and then send the report to the patient's pediatrician. 8. The pediatrician will be sent the results via fax as well. The transfer of information from 6 to 7 and 7 to 8 can be problematic, and is not electronic based. Many times a hospital will not have the information of the primary care pediatrician. 9. The pediatrician will review the results and determine if additional testing is necessary. 10. The pediatrician will then review results with the parents and possible further testing would be used to confirm a diagnosis.

6.5.2. Alternative Research-Based Newborn Screening

The blood-spot based newborn screening can be a long and disorganized process. As an alternative to this process, some researchers are currently employing sequencing of newborns for their screening. Sequencing provides data that can be more readily stored, transferred, and interacted with than blood spot testing. The sequences can also be utilized in later care as well, assuming no spontaneous variants.

It should be noted here that this test is not mandatory as the DBS testing is, and the information is not sent to or stored at a state laboratory. If this method becomes adopted in the future, the workflow would change to accommodate state regulations.

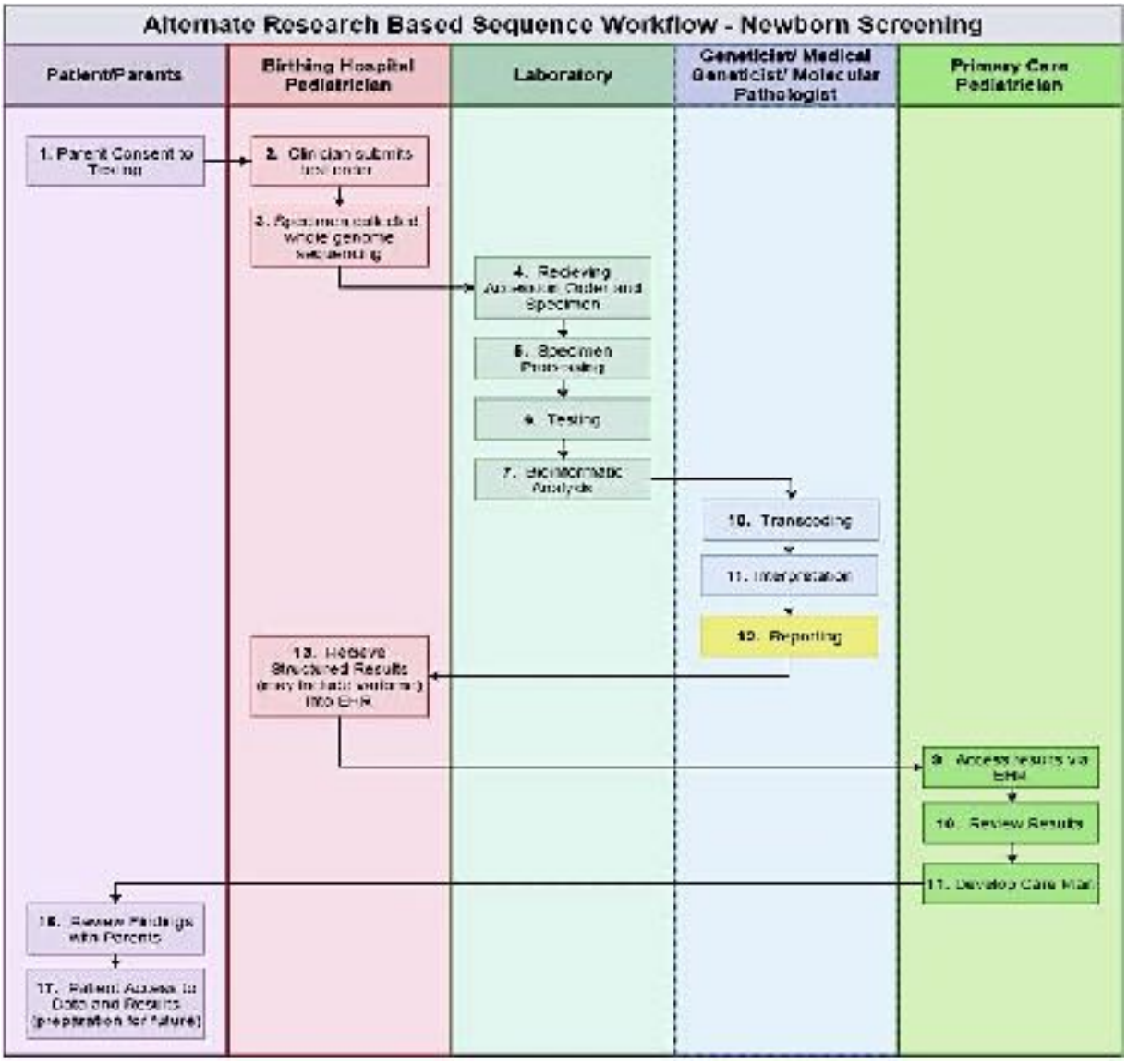


Figure 6.5-2: Alternative flow based on a whole genome sequencing screening as opposed to blood spot testing. The workflow itself is similar to a germline flow (see Figure 5.2-1), however information is distributed via EHR to the patient's pediatrician for care (steps 9-11).

6.6.Newborn Genome and Targeted Panel Testing

For additional information, the healthcare provider can conduct further testing after birth and into adulthood. For example, today the qualitative assay CytoScan is used to detect chromosomal copy number variants (CNV) in genomic DNA postnatally. After whole blood testing, the genomic DNA obtained is referred for chromosomal testing based on clinical presentation. Another emerging area with both near and long term clinical benefits is targeted sequencing. Targeted panel testing is currently in use to test for certain cancers or diseases and to develop targeted therapies. Targeted panel testing is of great use to pharmacogenomics (see Section 5.8).

As the cost of genome sequencing continues to go down, whole genome sequencing at birth will eventually become economically feasible and crucial component of individualized care. This type of testing will then supersede the testing described in this section and Section 6.5.1.

6.7. Whole Exome Sequencing

The exome, or protein-coding portion of the genome, makes up ~1-2% of the entire genome. Despite this, variations in the DNA sequence of the exome are much more likely to be associated with a particular phenotype of a Mendelian disorder. Whole-exome sequencing (WES) is a technique used to analyze the DNA sequence of the exome and has begun to show promise, specifically in the genetic work-up of patients who present with a challenging constellation of phenotypic features that has sent both clinicians and patients on a ‘diagnostic odyssey.’

Whole-exome sequencing (WES) represents a significant breakthrough in the field of human genetics. This technology has largely contributed to the identification of new disease-causing genes and is now entering clinical laboratories. WES is particularly relevant concerning rare diseases, which proves particularly difficult for physicians to diagnose. However, the complexity of this technology renders its applicability onto the clinical setting uncertain. In several small series, prenatal whole exome sequencing (WES) approaches have identified genetic diagnoses when conventional tests (karyotype and microarray) were not diagnostic. Data regarding the clinical utility and interpretative challenges from the clinician's perspective are lacking.

The Trio Whole Exome Sequencing test is a novel, composite test recently developed for the purpose of identifying changes in a patient’s DNA that are cause for medical concern. The Trio Whole Exome test uses next-generation sequencing techniques to concurrently analyze the coding regions, or exons, of thousands of genes.

The purpose of the test is to sequence the exomes of a patient to the individual nucleotide—a level of detail necessary to compile a consensus sequence with high accuracy. This highly accurate sequence is then compared to available references and standards, and most importantly—the parental WES data. Upon interpretation by a board-certified clinician or laboratory science and examination of reference sequences, variations in the patient’s DNA sequence can be readily determined and related back to the patient’s clinical examinations in an effort to identify the cause of a genetic or medical disorder.

Whole Exome Sequencing Platform Tests (Wong, Lee-Jun C., ed. "Next Generation Sequencing Based Clinical Molecular Diagnosis of Human Genetic Disorders." (2017).)		
Test Name	**Time**	**Parents**
WES Trio	10 weeks	Required
Critical WES Trio	3 weeks	Required
Proband WES	15 weeks	Recommended
Proband WES + CMA	15 weeks for WES	Recommended
BluePrint Proband WES	10 weeks	Recommended
Prenatal WES Trio	3 weeks	Required
Adult Screening Exome	15 weeks	Recommended

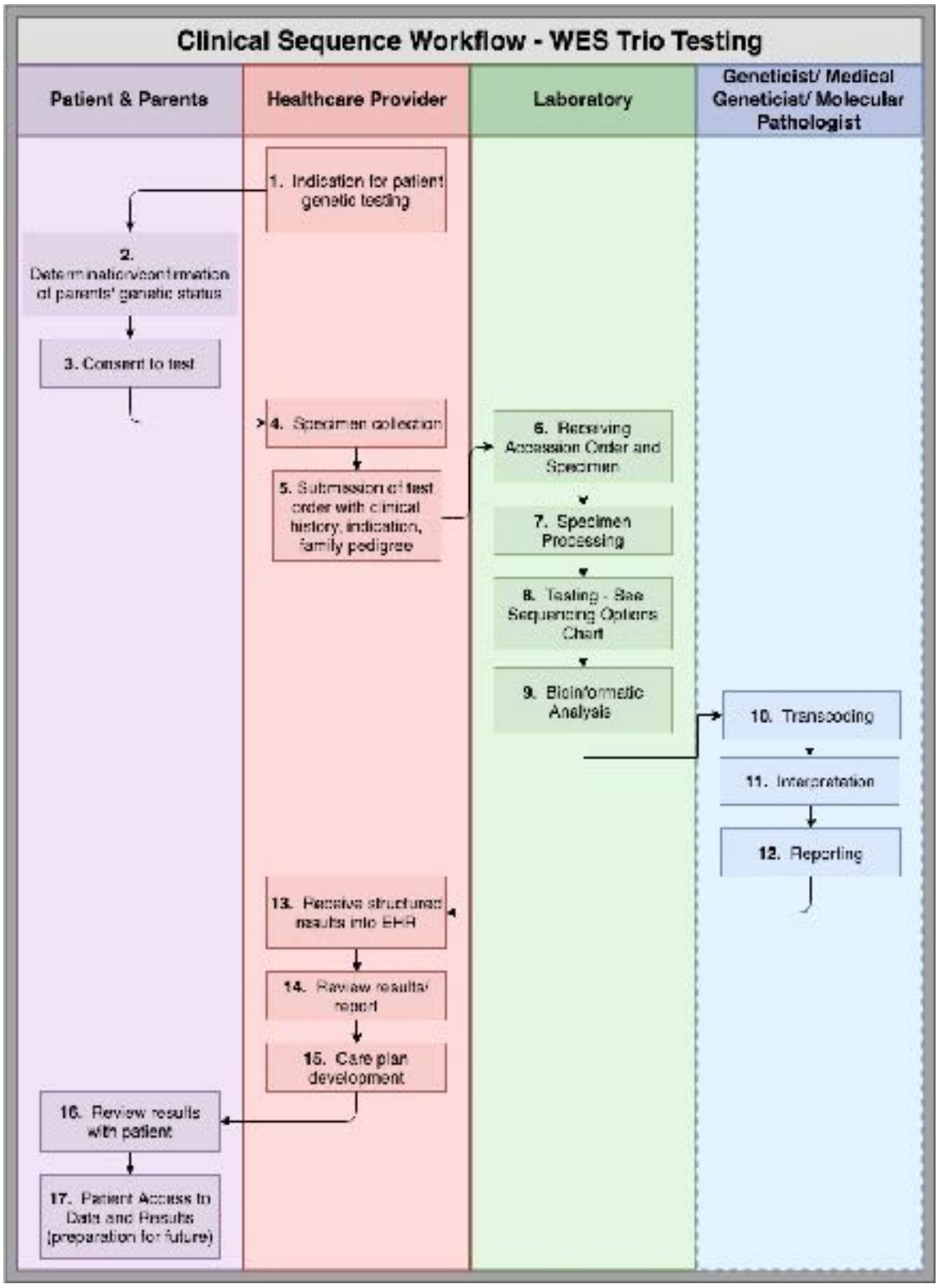


Figure 6.7.2 Workflow WES Trio Testing

6.8. Public Health Testing – Microbial

Whole genome sequencing of pathogenic bacteria, viruses, and fungi offers multiple tools for clinicians, researchers, epidemiologists, and public health officials to combat disease (Köser et al. 2012). In this situation, the specimen collection method would vary for different pathogens. After the proper specimen is collected and later isolated, sequencing and traditional phenotypic tests and analysis would be done to determine the specimen's genotype and phenotypic traits respectively. The sequencing of pathogens provides a method to track the evolution of drug resistance with high genetic variability and rapid genotypic changes, and it also provides a better tool for investigators and researchers to determine transmission pathways and provide help in the case of outbreaks. In slow growing bacteria with stable genotypes, where phenotypic testing can take up to two weeks (e. g. *M. tuberculosis*), genome sequencing can provide a reliable alternative for drug susceptibility testing. A similar method of viral genotyping of HIV and other retroviruses can be used to determine retroviral drug resistance. In these cases, microbial genetic data from diagnostic laboratories and geneticists would be integrated into public health repositories and research databases (see Section 5.6). It should be noted that this data would be distinct from the human genetic data previously described (i.e. viral, bacterial, or fungal sequences) and would use different databases (see Section 7.1).

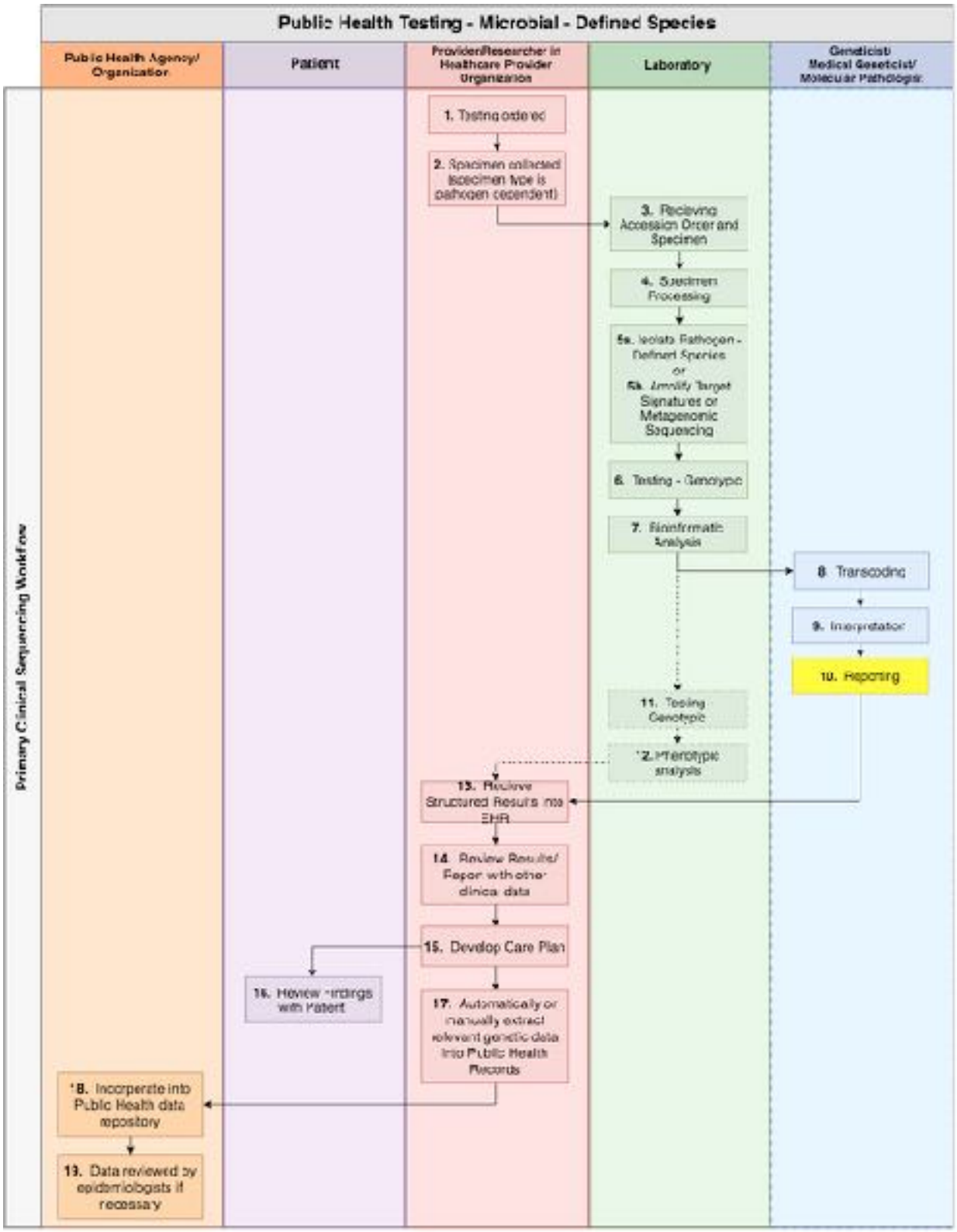


Figure 6.8-1: Public Health Testing Microbial. 1. The relevant sequencing and phenotypic based testing is ordered by the clinician. 2. The relevant specimen is collected from the patient. Depending on suspected pathogen type/infection site, different specimen types will be collected. 3. The specimen order and the specimen itself is sent to the lab with pertinent information. 4. The specimen is processed and prepped for genotypic and possibly phenotypic testing.

5a. If the pathogen is a defined species, then it will be isolated. 5b. If the pathogen is not defined, then target signatures (e.g. HIV, HCV viral genotyping) or metagenomics sequences will be amplified. Currently there are no CLIA-level applications for metagenomics though. 6. The specimen is tested and analyzed for genotype. 7. Genotypic data from the instrument passes through a bioinformatics pipeline for data processing: alignment and identification of sequence variants. 8. Raw genomic data is transformed from bioinformatics format into healthcare IT data standards. 9. The genotypic results are interpreted. 10. A genotypic report is compiled from the interpretation. 11. Phenotype is not always specifically called from genomic data. It depends upon assay. Phenotypic tests include different plating techniques and assays. 12. Phenotypic results are analyzed and reported to the clinician. 13. The clinician then receives the genotypic and phenotypic data. 14. The genotypic and phenotypic results are reviewed along with other data. 15. A care development plan is made for the patient dependent on pathogen. 16. The results reviewed by the patient with the clinician. 17. The pertinent genotypic and phenotypic information is then added to health organizations' databases automatically or manually by chart review. 18. The information is then integrated into the public health data repository for possible future use or in evaluating outbreaks/clonality. 19. The data is then possibly reviewed by epidemiologists or those studying/researching an outbreak or pathogen if necessary.

6.9. Defined Genetic Testing vs. Expanding Genetic Tests

Different business models are evolving within the genetic testing field, which will have implications for information systems needed. 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 variants. If **new** regions are found to be associated with cardiomyopathy, the patient's DNA may not be retested without a new clinical requisition. The burden to identify new genetic tests may fall to the genetic counselor or doctor caring for the patient. Clinical decision systems which support identification of patients needing follow-up testing or reinterpretation of results would be ideal. 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.

6.10.Patient Panel Management – Analytics for Care Quality

HL7 Clinical Genomics is looking for a partner to help inform this use case.

In addition, the HL7 Clinical Genomics workgroup will be collaborating with the Clinical Quality Information Workgroup.

Panel management can be defined as set of tools and processes for population care that are applied systematically at the level of a primary care panel. It typically involves non-physicians who utilize chronic disease registries, electronic health records, and data support tools to identify missed opportunities for unmet preventative care and chronic disease care, and who communicate recommendations from the provider to patient. Teams may utilize panel management to reach out to every patient at each visit, called in-reach, or to address unmet needs between patient visits, called outreach.

The use of medical assistants or nurses to conduct panel management has been associated with improved care process outcomes such as improved rates of vaccination, health care proxy designation, and bone density screening in the care of elderly patients (Loo TS, Davis RB, Lipsitz LA, et al. Electronic medical record reminders and panel management to improve primary care of elderly patients). Practices may use panel management to ask, "Have all of our patients between 50 and 75 years of age received colorectal cancer screening at the appropriate time intervals? Have all of our patients with diabetes had laboratory tests for HbA1c, LDL cholesterol, and urine microalbumin at the appropriate times?"

Electronic decision-support systems appear to enhance care, but improving both tools and work practices may optimize outcomes. Successful panel management programs need to be supported by computerized clinical support systems that provide relevant care reminders at the point of care, flexible data registries, and performance feedback. Even in systems with a panel management support tool, there are barriers to its utilization, such as insufficient time, competing demands, and suboptimal staffing. Facilitators of panel management include strong team roles, leadership support and training for tool implementation, and dedicated time for use.

The workflow below details implementation of a panel management program at a primary care practice.

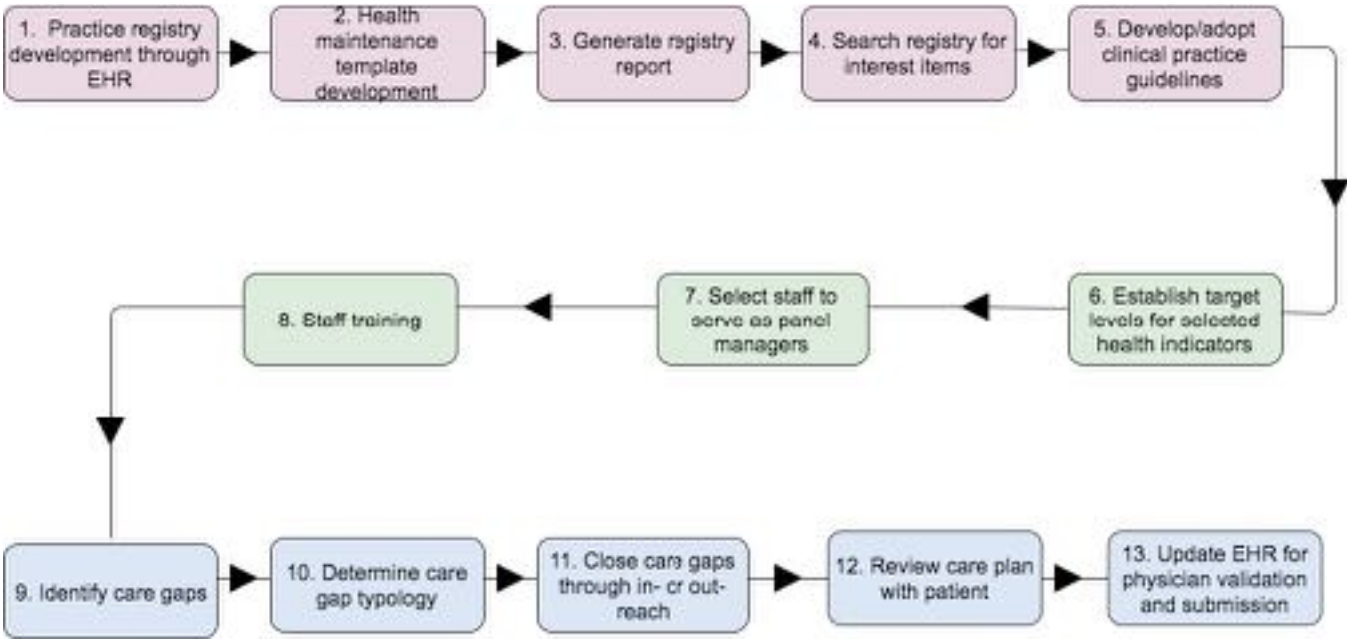


Figure 6.10.1 Module for panel management implementation. (1) Red boxes detail pre-implementation EHR configuration. 1. A registry is a database with medical information about immunizations, cancer screenings, etc. Many EHR systems contain a registry function, but it is more common in patient panel management systems to use a separate registry program. (2) Green boxes detail practice/staff training protocols for panel management. 5. Many practices use evidence-based national guidelines for establishing target levels for selected health indicators. (3) Blue boxes detail clinical practices in a panel management system. 10. Types of care gaps include process care gaps, which include when a patient is overdue for a service that should be done periodically, and outcome care gaps, which include when a patient is not meeting a goal range for a particular condition.

FDA Scenarios in Genomics Testing/Reporting

The FDA created precisionFDA, a cloud-based community research and development portal that engages users across the world to share data and tools to test, pilot, and validate existing and new bioinformatics approaches to NGS processing. Precision FDA is a live website at precision.fda.gov and it is a community platform for NGS assay evaluation and regulatory science exploration.

PrecisionFDA seeks to advance regulatory science by supplying the genomics community with a cloud-based platform in which participants can securely share and access datasets, analysis pipelines, and bioinformatics tools. It does not serve a regulatory role-- instead, the goal of precisionFDA is to allow participants to benchmark their approaches and contribute to knowledge informing regulatory pathways and decision-making in the field of genomics.

To accomplish this goal, precisionFDA participants can conduct genome analyses and comparisons against reference materials, and publish and communicate results, materials, and tools. Participants include, but are not limited to, FDA and other government agencies, researchers, genome test or software providers, standards-making bodies, and biotechnology companies.

The workflow below serves as a guide for user interaction with the precisionFDA platform. It is important to note that the platform is currently still in beta, and features and capabilities are expected to evolve over time.

6.11.CMS and/or Payor Scenarios

HL7 Clinical Genomics is looking for specific representatives as partners to help inform this use case.

Centers for Medicare and Medicaid Services (CMS) and commercial payers are determining rates of reimbursement for genetic and genomic tests that are impacting patient care, physicians' abilities to order tests, and abilities to receive adequate reimbursement for tests performed.

In 2014, the American Medical Association (AMA) issued new current procedural terminology (CPT) codes for genomic testing. These codes range from 81410 to 81471 (28 codes in total) and cover testing using targeted panel sequencing (5-50 genes), whole exome sequencing and whole genome sequencing. While the existence of such codes is a prerequisite for genomic testing reimbursement, health insurance payers do not automatically cover these tests.

There are many CPT codes for testing of individual genes or pairs of genes and all such tests are reimbursed (see below), with reimbursement levels ranging from $58.31 to several thousand dollars. As clinical-grade annotated genetic variant information becomes more broadly publicly available (see for example ClinVar), it is likely that the clinical utility of multi-gene genomic testing will become more obvious.

Below are a selection of CPT codes relevant to the use cases described in this document (AMIA 2016):

CPT Code	Procedure Description	2016 CMS Fee ($)
81415	Exome (eg, unexplained constitutional or heritable disorder or syndrome); sequence analysis	0
81416	Exome (eg, unexplained constitutional or heritable disorder or syndrome); sequence analysis, each comparator exome (eg, parents, siblings) (List separately in addition to code for primary procedure)	0
81417	Exome (eg, unexplained constitutional or heritable disorder or syndrome); re-evaluation of previously obtained exome sequence (eg, updated knowledge or unrelated condition/syndrome)	0
81420	Fetal chromosomal aneuploidy (eg, trisomy 21, monosomy X) genomic sequence analysis panel, circulating cell-free fetal DNA in maternal blood, must include analysis of chromosomes 13, 18, and 21	0
81425	Genome (eg, unexplained constitutional or heritable disorder or syndrome); sequence analysis	0
81426	Genome (eg, unexplained constitutional or heritable disorder or syndrome); sequence analysis, each comparator genome (eg, parents, siblings) (List separately in addition to code for primary procedure)	0

81427	**Genome (eg, unexplained constitutional or heritable disorder or syndrome); re-evaluation of previously obtained genome seq**	**0**
81432	**Hereditary breast cancer-related disorders (eg, hereditary breast cancer, hereditary ovarian cancer, hereditary endometrial cancer); genomic sequence analysis panel, must include sequencing of at least 14 genes, including ATM, BRCA1, BRCA2, BRIP1, CDH1, MLH1, MSH2, MSH6, NBN, PALB2, PTEN, RAD51C, STK11, and TP53**	**0**
81433	**Hereditary breast cancer-related disorders (eg, hereditary breast cancer, hereditary ovarian cancer, hereditary endometrial cancer); duplication/deletion analysis panel, must include analyses for BRCA1, BRCA2, MLH1, MSH2, and STK11**	**0**
81434	**Hereditary retinal disorders (eg, retinitis pigmentosa, Leber congenital amaurosis, cone-rod dystrophy), genomic sequence analysis panel, must include sequencing of at least 15 genes, including ABCA4, CNGA1, CRB1, EYS, PDE6A, PDE6B, PRPF31, PRPH2, RDH12, RHO, RP1, RP2, RPE65, RPGR, and USH2A**	**0**
81437	**Hereditary neuroendocrine tumor disorders (eg, medullary thyroid carcinoma, parathyroid carcinoma, malignant pheochromocytoma or paraganglioma); genomic sequence analysis panel, must include sequencing of at least 6 genes, including MAX, SDHB, SDHC, SDHD**	**0**

81438	**Hereditary neuroendocrine tumor disorders (eg, medullary thyroid carcinoma, parathyroid carcinoma, malignant pheochromocytoma or paraganglioma); duplication/deletion analysis panel, must include analyses for SDHB, SDHC, SDHD, and VHL**	**0**
81440	**Nuclear encoded mitochondrial genes (eg, neurologic or myopathic phenotypes), genomic sequence panel, must include analysis of at least 100 genes, including BCS1L, C10orf2, COQ2, COX10, DGUOK, MPV17, OPA1, PDSS2, POLG, POLG2, RRM2B, SCO1, SCO2, SLC25A4, SUCLA2, SUCLG1, TAZ, TK2, and TYMP**	**0**
81445	**Targeted genomic sequence analysis panel, solid organ neoplasm, DNA analysis, and RNA analysis when performed, 5-50 genes (eg, ALK, BRAF, CDKN2A, EGFR, ERBB2, KIT, KRAS, NRAS, MET, PDGFRA, PDGFRB, PGR, PIK3CA, PTEN, RET), interrogation for sequence variants and copy number variants or rearrangements, if performed**	**597.91**
81450	**Targeted genomic sequence analysis panel, hematolymphoid neoplasm or disorder, DNA analysis, and RNA analysis when performed, 5-50 genes (eg, BRAF, CEBPA, DNMT3A, EZH2, FLT3, IDH1, IDH2, JAK2, KRAS, KIT, MLL, NRAS, NPM1, NOTCH1), interrogation for sequence variants, and copy number variants or rearrangements, or isoform expression or mRNA expression levels, if performed**	**648.40**

81455	**Targeted genomic sequence analysis panel, solid organ or hematolymphoid neoplasm, DNA analysis, and RNA analysis when performed, 51 or greater genes (eg, ALK, BRAF, CDKN2A, CEBPA, DNMT3A, EGFR, ERBB2, EZH2, FLT3, IDH1, IDH2, JAK2, KIT, KRAS, MLL, NPM1, NRAS, MET, NOTCH1, PDGFRA, PDGFRB, PGR, PIK3CA, PTEN, RET), interrogation for sequence variants and copy number variants or rearrangements, if performed**	**0**

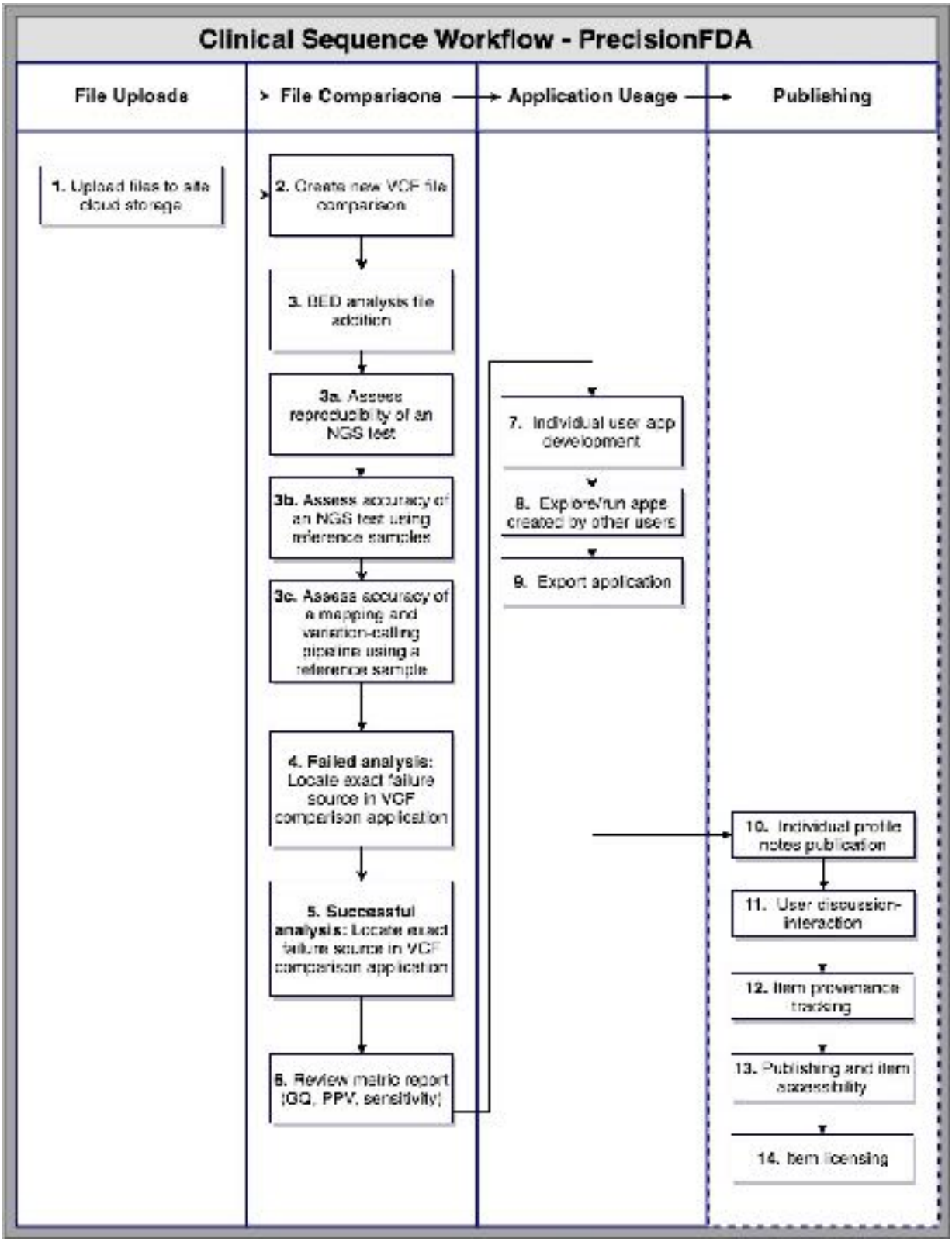
Figure 6.12-1 PrecisionFDA workflow. 1. Files may be uploaded to precisionFDA according to size. Smaller files may be uploaded directly through the user browser and larger files may be uploaded through a precisionFDA secure URL. 2. PrecisionFDA's comparison method conceptually resembles comparison method #3 of GA4GH benchmarking definitions. 11. Additional interactions include commenting on and upvoting other users' notes and discussion threads. 12. Tracking is a precisionFDA feature that allows a user to investigate the provenance of an item and to generate a provenance graph. 14. The licensing feature of precisionFDA allows users to protect datasets or software by requiring that users agree to a license agreement before they can access items.

Mutation and Type Specific Prognosis – Cancer Testing

After mutation type (missense, frameshift, etc) is significantly analyzed through somatic and tumor based genotyping (see Section 5.3), clinicians will be able to make better informed prognoses based on the mutation type. The data elements necessary to inform better prognoses such as cancer subtype, date of diagnosis, date of death, and the specific mutation that was detected, can be compiled and further analyzed statistically to produce a more accurate, patient-by-patient based prognosis. In the future as more sequencing data becomes available, this technique could be extended to variant specific prognoses in addition to mutation specific.

6.12.Clinical Trial Ascertainment and Feasibility

Readily available genomic sequencing and improved interoperability of patient health records and data can facilitate increased specificity and scope in a clinical trial for a drug.



In the case of a specific cancer, the biopsied tumor tissue will yield a somatic sequence which can be entered into the National Cancer Institute's Molecular Analysis for Therapy

Choice (NCI's MATCH). In the case of MATCH, the somatic sequencing data will be analyzed for 143 actionable mutations which have targeted drug therapies available in clinical trials, then organized into cohorts and administered the treatment as part of the trial. In MATCH the drugs list is being updated, with drugs being added and removed, and the sequencing is carried out by the NCI itself. However, similar clinical trial databases can be compatible with germline and somatic sequencing already part of the patient's EHR going forward. In addition to sequencing being used to find compatible patients for clinical drug trials, patient sequencing databases can be used to determine the feasibility of a trial for specific bio-markers. Analysis of variant type and frequency can guide drug development by solidifying the relationship between cancer type and mutation type and then designing a clinical trial around the presence or absence of a relationship.

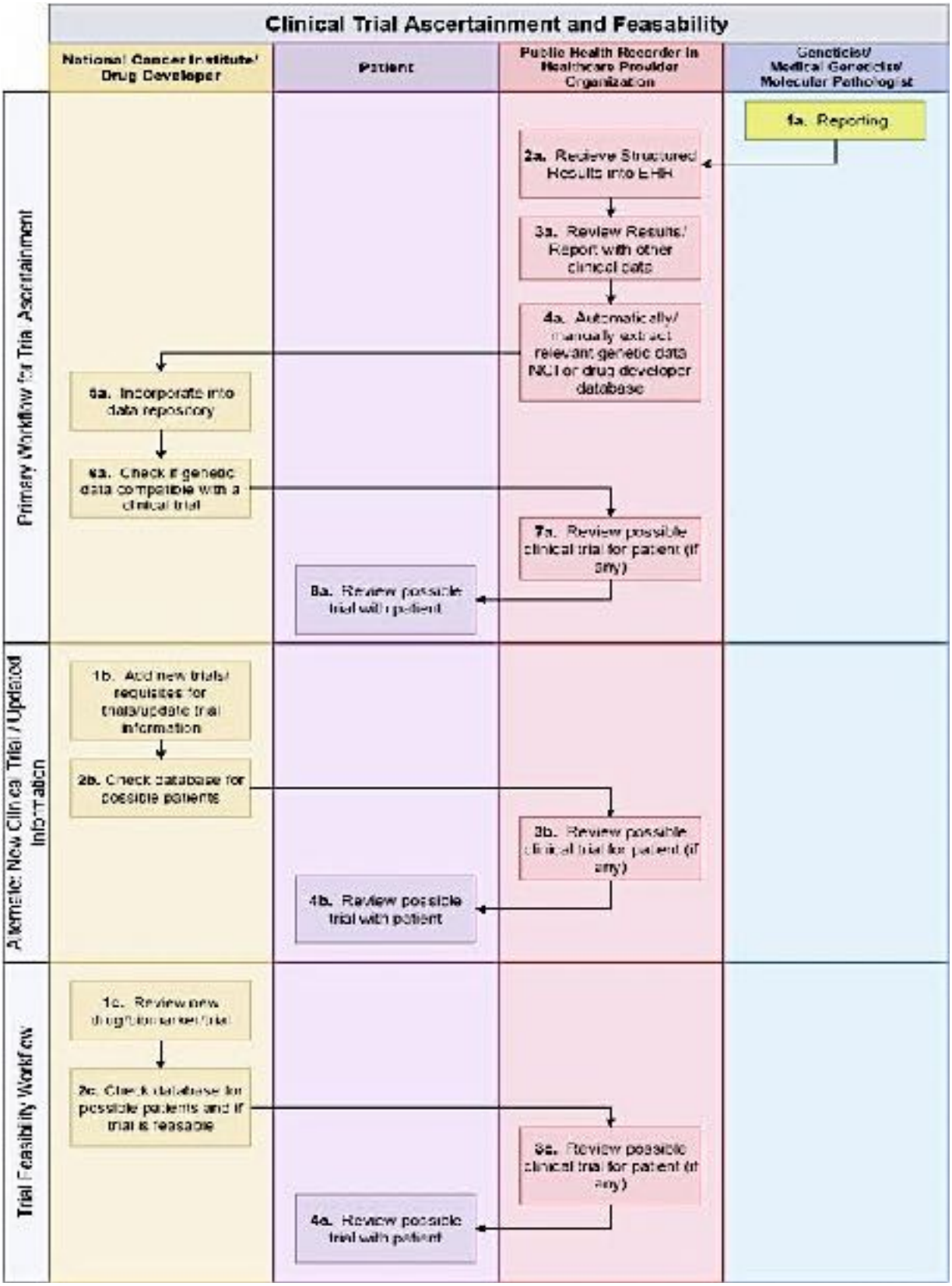


Figure 6.13-1: Clinical Trial Ascertainment and Feasibility. 1a. The genetic report is completed. 2a. The clinician receives structured results into the EHR (this could include the variome in the future). 3a. The clinician then reviews

the report with other clinical data. 4a. The relevant biomarker data is then either automatically extracted from the report or is manually entered into the relevant clinical trial database (either government backed like NCI or private). 5a. The patient data is then incorporated into the relevant database. 6a. The data is checked to determine if patient's genome is compatible with current or projected clinical trials. 7a. If there is a trial, this information is sent back to the clinician for review of the trial. 8a. The patient can then review the trial before making a decision. **Alternate trial ascertainment**: 1b. A new or an updated trial is added to the system with new/updated biomarkers and trial information. 2b. Previously input data is then screened to see if compatible subjects for the trial. 3b. Clinician is alerted if a patient qualifies for a new trial. 4b. The patient can review the trial before entering the trial. **Trial feasibility determination:** 1c. A new biomarker is considered for development/trial. This biomarker is entered into the system. 2c. The database is screened to determine if a trial would be feasible for the developer and if there are enough subjects that qualify to get the trial. If so the trial is then added. 3c. The clinician is alerted that the patient now qualifies for a new trial. 4c. The patient may review the trial before entering the trial.

6.13.Genome-Directed Treatment and Dosing (Cancer)

In the case of specific cancers, different treatments may be more or less effective depending on the genotype of tumor or somatic cells (see Section 5.3 for somatic genotyping workflow). The case of genome directed treatment dosing is similar to that of the drug dosage calculator (see Section 5.4), but the dosage of specific drugs is affected by the presence or absence of specific genes or variants. In a study on the effectiveness of imatinib on chronic eosinophilic leukaemia (CEL) and hypereosinophilic syndrome (HES), the drug dosage was different for patients with two distinct, molecular abnormalities – i.e. the presence or absence specific fusion genes (Metzgeroth et al. 2008). Another study developed a drug dosage algorithm for warfarin using genetic polymorphism of two specific genotypes in addition to age and height (Sconce et al. 2005). Different variants can result in different active sites or even a resistance to a drug and can be taken into account when prescribing a drug and drug dosage.

Neoantigens and Immunotherapy Response (Cancer)

A relatively novel and exciting treatment for certain types of cancer is the use of antibodies to target antigens, eliciting an immune response in a patient that attacks the cancer cells. This treatment along with certain vaccinations and immune checkpoint inhibitors are grouped together as types of immunotherapy. Tumor cells can acquire clonal neoantigens, i.e. present throughout all tumors, that can promote and be affected by a T cell immune response (McGranahan et al. 2016). Clonal neoantigens are determined by genome/exome sequencing, and the type and frequency relative to a specific cancer can be recorded through an antigen database similar to HLA typing. Using a the neoantigen data of a patient, personalized vaccines and cell therapies can be developed and specific immunoreactivity can be modeled and predicted

6.14.Proteomics

Mass spectrometry–based proteomics has emerged as the leading method for detection, quantification, and characterization of proteins. Proteogenomics enables the detection of proteomic variations and can be defined, broadly, as the use of nucleotide sequences to generate candidate protein sequences for mass spectrometry database searching. Proteogenomics is experiencing

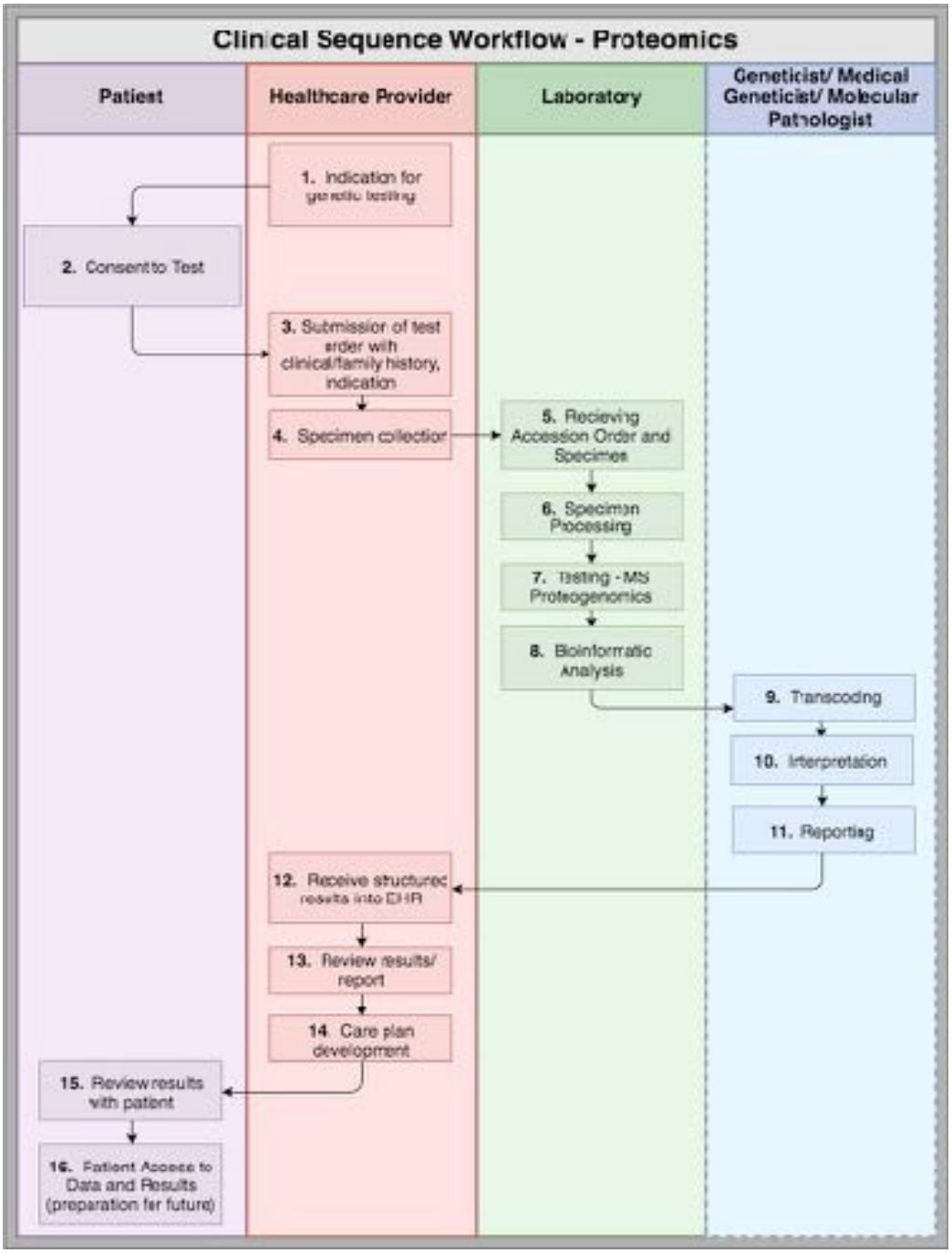


Figure 6.17-1: Clinical Proteomics Workflow

heightened significance due to two developments: (*a*) next-generation sequencing techniques, and (*b*) the revelation of the tremendous complexity of the human proteome as expressed at the levels of genes, cells, tissues, individuals, and populations. In the clinical laboratory, the method(s) used will depend upon the goal of the analysis, taking into consideration the costs and benefits associated with each methodology.

Furthermore, as clinical proteomics is increasingly developed as a field, more complex analysis techniques are emerging. Here we put forth a sample workflow for the analytical, or "Testing" portion of the general clinical proteomics workflow.

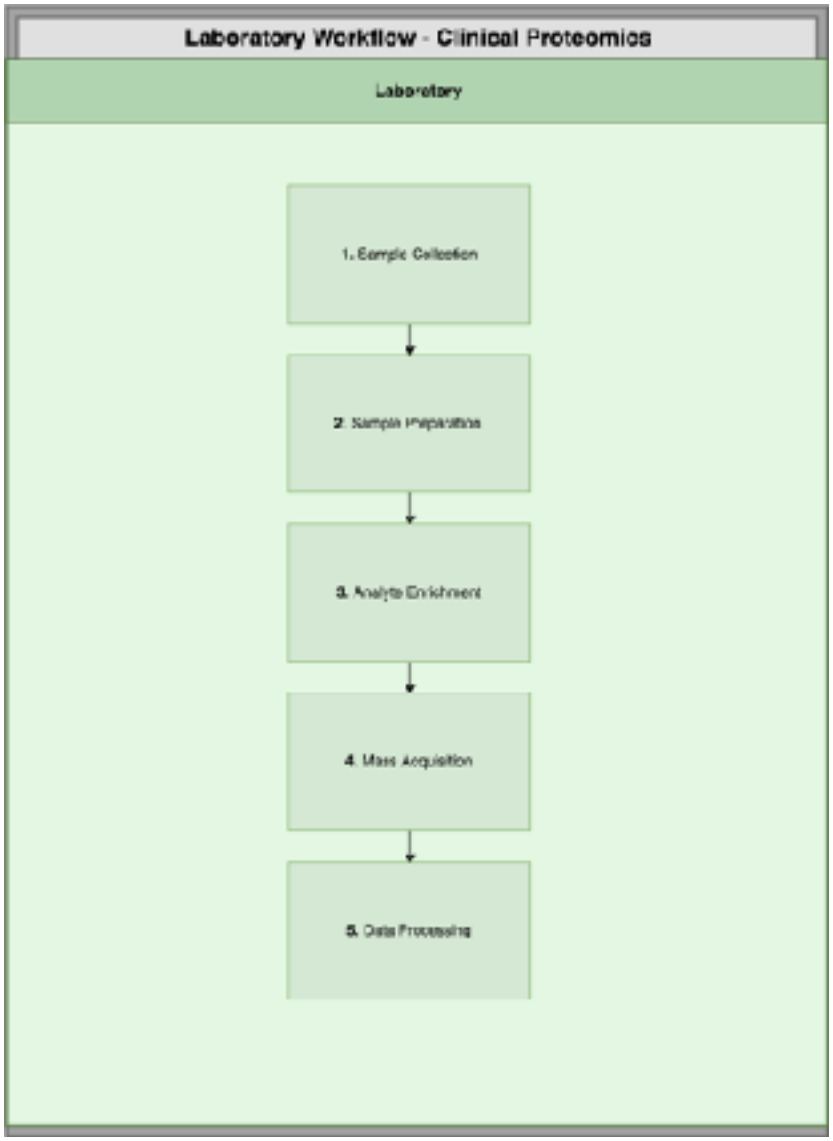


Figure 6.17-2: Proteomics Laboratory Workflow**. 1** Samples are usually depleted of complexity/enriched for a specific analyte of interest prior to analysis. **2** Extraction and digestion by organic solvents for protein quantitation. **3** By chromatographic separation and affinity purification. **4** Various mass spectrometric methods have been developed for clinical proteomics, such as MALDI-TOF MS. **5.** Novel technologies for clinical proteomics data processing include SRM, MRM for relative or absolute quantitation and PASSEL for meta-analyses.

6.15. RNA-Sequencing

With the emergence of novel RNA sequencing (RNA-seq) technologies, RNA-based biomolecules hold expanded promise for their diagnostic, prognostic and therapeutic applicability in various diseases, including cancers and infectious diseases. Detection of gene fusions and differential expression of known disease-causing transcripts by RNA-seq represent some of the most immediate opportunities. However, it is the diversity of RNA species detected through RNA-seq that holds new promise for the multi-faceted clinical applicability of RNA-based measures, including the potential of extracellular RNAs as non-invasive diagnostic indicators of disease. Ongoing efforts towards the establishment of benchmark standards, assay optimization for clinical conditions and demonstration of assay reproducibility are required to expand the clinical utility of RNA-seq.

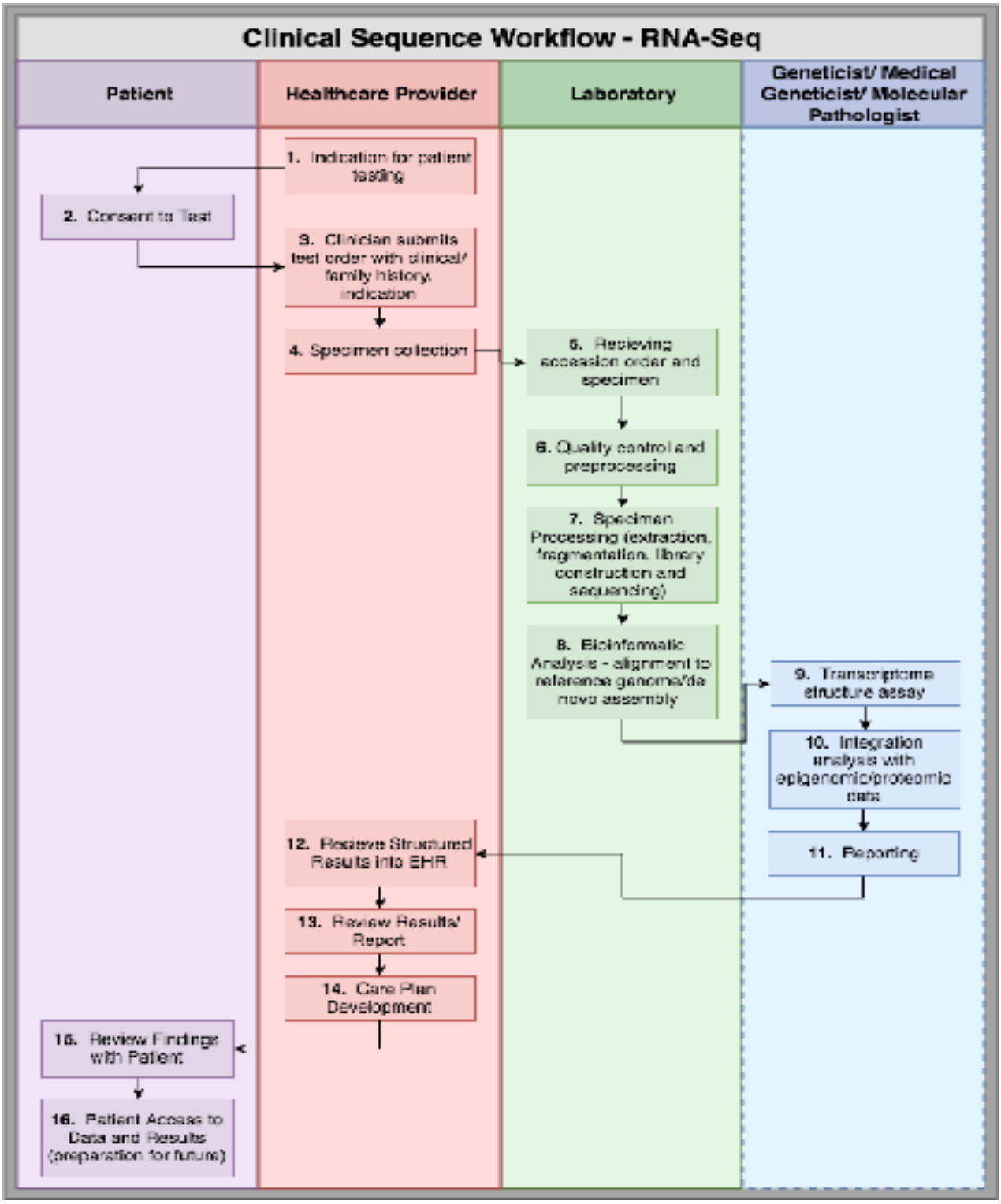


Figure 6.17-1 RNA-Seq workflow.

7. Clinical Genomic Standard Use Cases and Applications

A majority of the use cases outlined above are fairly general and ubiquitous across platforms. It is particularly important for expanded interoperability that a genomic standard and applications have semantic, if not syntactic, similarities. For example, in terms of describing diplotypes, there can be different types of data – the way one refers to the original list of variants and possibly about a specific allele associated with a diplotype – which must be consistent across the platforms. The following describes some scenarios that a clinical genomic standard may support.

7.1. Scenario 1: Clinical DecisionMaking Alerts

As the technology develops and clinical genomics becomes more widely implemented, a variety of applications will be developed to aide in clinician care. One of these applications is the clinical decision making alert system. The alert system would provide the clinician with a 'pop-up' message with suggestions for the clinician's consideration prior to the ordering of tests, development of treatment, or prescribing of drugs. The alert system would recommend new tests as they become available and as new genes are discovered for conditions on the patient's problem list (e.g. cardiomyopathy). The application would advise clinicians of FDA approved/required companion diagnostics for specific drugs (e.g. EGFR test for TKA inhibitors in NSCLC). The application would also aid in safe prescription of drugs, notifying the clinician if there is a discrepancy between diagnosis, drug prescription, and drug dosage and offering a recommendation based on previous similar cases.

7.2. Scenario 2: Search

Clinical genomics should facilitate the ability to search sequences, cohorts, and medical data with ease. The ability to easily go back or search into files is important in streamlining precision medicine. In order to implement robust searches across multiple platforms, genomic data needs to be standardized and structured.

7.3. Scenario 3: Data Aggregation

Clinical genomics should facilitate the congregation of relevant genetic data with other medical data on a patient-by-patient basis. A genomic standard should enable applications to combine the results of multiple genetic tests for a patient into an accessible and comprehensive file. Further, different tests and information from pathology, surgery, and radiology can be combined with the genetic data into a single view.

8. Variant Classification

8.1. Variant Type by Genomic Source – Germline, Somatic, Prenatal/Fetal, Microbial

As noted in the discussion of specimen, variants need to be clearly defined as germline, somatic, prenatal, microbial, or unknown origin, when reporting into the electronic health record. In this way, variants will be appropriately contextualized for use.

8.2. Variant Type by Size and Characteristics

Supports the reporting of DNA variants identified within a gene, by sequencing or genotyping technology, with or without interpretation.

8.3. Structural Variants/Rearrangements

HL7 Clinical Genomics standards support reporting of these variants using ISCN (International Standard of Cytogenetic Nomenclature). These standards will be extended for identification of rearrangements using NGS technologies; however, to our knowledge the field has not yet adopted a uniform representation. Some options for consideration include dbVar (*dbVar Variant Call Submission Format Guidelines*) and HGVS extensions (Taschner & den Dunnen 2011).

8.4. Copy Number Change

HL7 Clinical Genomics standards will be extended for identification of copy number variants using NGS technologies; however, to our knowledge the field has not yet adopted a uniform representation.

8.5. Biomarkers

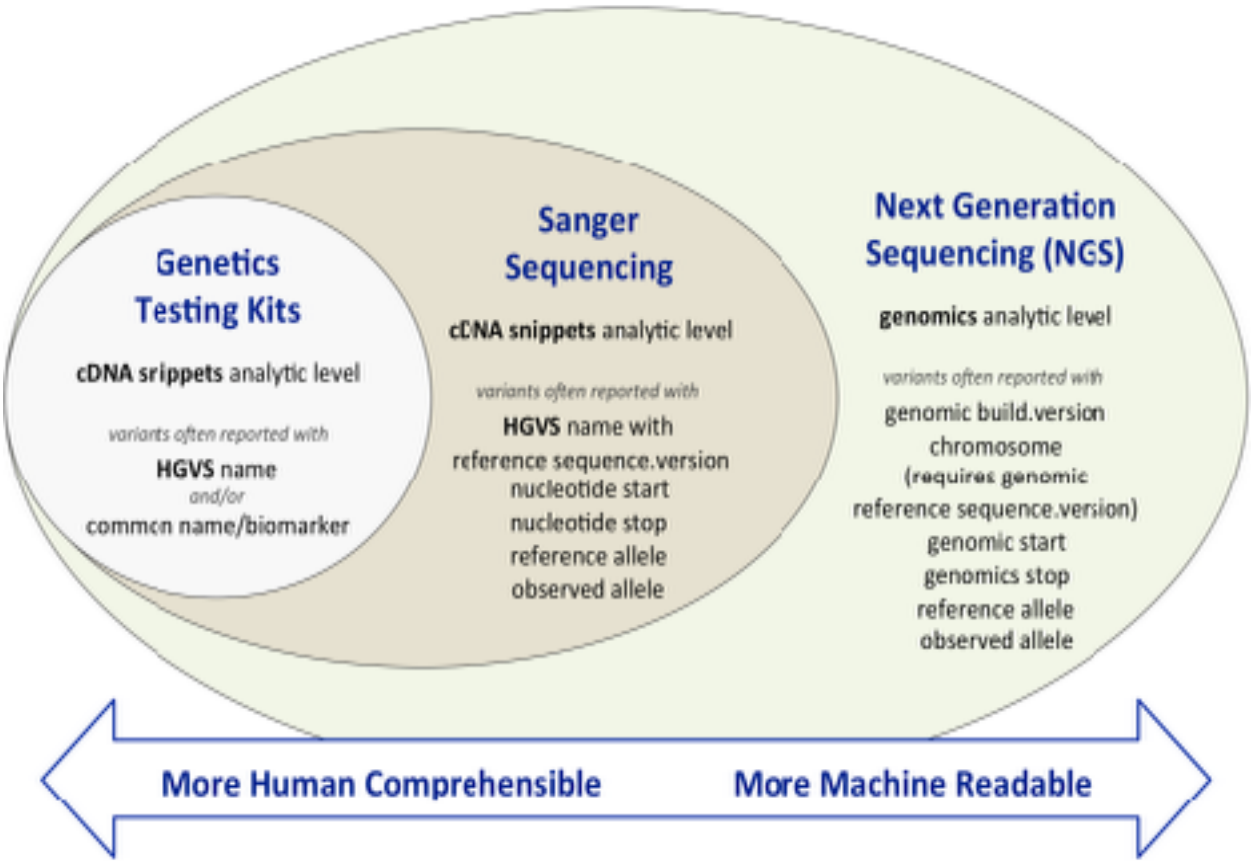
These standards will be extended for machine-readable coding of Biomarkers with mapping to genomic coordinates; however, to our knowledge the field has not yet adopted a uniform approach. A likely solution would be the encoding of Biomarkers in MedGen (NCBI's medical concept database; http://www.ncbi.nlm.nih.gov/medgen) with mapping to variants in ClinVar and dbVar.

9. Testing Platforms and Variant Representation

9.1. Relative Number of Next Generation Sequencing Tests

NCBI's Genetic Test Repository (NCBI's GTR at http://www.ncbi.nlm.nih.gov/gtr/) currently lists (as of December 2016) over 48,000 tests for 10,600 conditions and 16,200 genes from 483 laboratories. Of these, 1711 clinical tests (for 3718 conditions) in 54 laboratories are based on Next-Generation (NGS)/Massively parallel sequencing (MPS) methods. This represents approximately 9% of all registered tests. That is, the vast majority of genetic tests listed in the GTR are performed on other genetic testing platforms. By comparison, 14304 tests for 3590 conditions in 176 laboratories use Sanger Sequencing Analysis, representing approximately 75% of all listed tests.

9.2. Different Testing Platforms



Extension of HL7's clinical genomic reporting standards need to support interoperability across testing platforms, drive translation of machine readable formats into those readily understood by clinicians, and guide implementers in how to most fully and unambiguously represent genetic/genomic data. For a more comprehensive discussion on the topic, see Section 4.

9.2.1.DNA Variant Detection Approaches

Technology	Analytic Level	Use Case	Common Methods of Variant Definition	Notes
Cytogenetics	Chromosome	Prenatal testing	• Chromosome • Banding pattern using ISCN Nomenclature based on reference 'map' of normal	
Genetic Testing Kits	Nucleotide snippets, often with cDNA context	Most of current clinical genetic testing	• HGVS at cDNA level or • Biomarker	Testing context is aligned with clinical understanding, making results more actionable to a greater number of clinicians
Sanger Sequencing	Regional variant investigation using cDNA, or genomic reference	Smaller targeted sequencing tests	• RefSeq (cDNA, genomic) • Start/stop • Reference nucleotide • Observed nucleotide *Additional*: • Biomarker (optional)	Start, stop, and nucleotide information is denoted following HGVS nomenclature. Current software does not denote genomic coordinates. RefSeq is commonly used in the U.S.A, but EBI-based identifiers may be used in Europe.
NGS	Genomic, or regional (genomic contig, also cDNA)	Germline / somatic	• Genomic build.version • Chromosome • Start • Stop • Reference nucleotides • Observed nucleotides • Allelic fraction and subclonality *Additional*: • HGVS nomenclature at cDNA level • Biomarker (optional)	NGS is used for whole genome, whole exome, large gene panels, or single gene or region
NGS	Genomic, or regional (genomic contig,also cDNA)	HLA	• Genomic reference sequence.version • With enhanced variant detection against assembled cDNA • Start • Stop • Reference nucleotides • Observed nucleotides *Additional*: • HLA specific nomenclature	HLA regions are not included in the genome build, so locus-specific reference sequences must be used

9.2.2.

9.3.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 bioinformatics representation, to support linking to research and clinical information systems.

11. Genetic/Genomic Standards in Healthcare IT

The following subsections list recommendations for specific nomenclatures (e.g. HGVS), field standards (e.g. reference sequences), public repositories and knowledge bases, 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 are listed here as well as indication to whether this should required or optional.

11.1. Genes

HGNC ID (required)

Table 10-1 - HGNC	
Code sets, vocabularies, terminologies and nomenclatures that need to be constrained	HGNC
OID	2.16.840.1.113883.6.336
Minimum attributes of the component	Gene ID
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/. HGNC carries the Gene ID, gene symbol, and full name, however gene symbols and names are subject to change overtime so the Gene ID is used.

11.2.Sequence Variations

HGVS (optional, recommended)

Table 10-2 - HGVS

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
Other Comments	Human Genome Variation Society (HGVS) Nomenclature standards for the description of sequence variations are maintained at http://varnomen.hgvs.org/. This standard is well accepted by the clinical genetic community and is extended on an ongoing basis to support genetic findings. Several freely available tools and libraries exist to manipulate HGVS-formatted variants. While HGVS may be preferred for human readability, it should not be relied upon for computability or as primary identifier in EHRs.

dbSNP (optional, highly recommended)

Table 10-3 - 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.)

COSMIC (optional)

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

Table 10-4 - 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 Variants In Cancer (COSMIC) serves as a repository for somatic variants identified in specific cancer specimens. These variants are recorded associated with structured description of the specimen. Available at: http://www.sanger.ac.uk/genetics/CGP/cosmic/

11.2.1.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.

RefSeq

Table 10-5 - 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

LRG

Table 10-6 - 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

11.2.2.Publicly Available References (for clinical and translational genomics)

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 variants.

Table 10-7 - 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	Knowledgebase for genes, variants/variants and genetic based phenotypes. Note this information includes somatic or acquired variants/variants and phenotypes and is not limited to inherited variants/variants and phenotypes. Available at: http://www.omim.org/ and through NCBI at http://www.ncbi.nlm.nih.gov/omim Additionally, dbSNP contains links to variants in OMIM.

PubMed (optional)

Coding of references may include PubMed IDs to peer-reviewed medical literature.

Table 10-8 - 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	"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." Available at: http://www.ncbi.nlm.nih.gov/pubmed/

PharmGKB (optional)

PharmGKB IDs to community curated information on emerging pharmacogenomic associations.

Table 10-9 - 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/

ClinicalTrials.gov (optional)

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

Table 10-10 - 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

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. As part of this constraining process, message fields were bound to LOINC codes (see the Appendix for examples). Also, the Clinical Document Architecture (CDA) was constrained, resulting 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 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 coding system ID, name and precise version when binding is done at instantiation time (with the assumption that the coding system is well controlled and maintained independently).

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 variant.

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

Review of Existing HL7 Clinical Genomics Specifications

11.3.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 at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=23

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:

Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model, Release 1 (US Realm)

11.4.Fast Healthcare Interoperability Resources (FHIR) Genomics

FHIR Genomics is a subset of FHIR maintained by the HL7 Clinical Genomics Workgroup to cover clinical genomics. An implementation guidance document describes this here: http://hl7.org/fhir/2016Sep/genomics.html

Rather than messages or documents, FHIR contains discrete resources (as well as profiles and extensions) to capture information and make it available via an API. It uses standard web-based technologies (HTTP-based RESTful protocol, HTML and Cascading Style Sheets for user

interface integration, and JSON/XML. It was designed to capture the use cases described in this document.

11.5.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

11.6.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)

11.7.Sequence Variations / Chromosomal change

11.7.1.Small Genetic Variations within a Gene

HL7 Clinical Genomic standards support the reporting of small genetic variants/mutations identified within a gene using v2.5.1 Implementation Guide for Laboratory Reporting

HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model, Release 2

http://www.hl7.org/implement/standards/product_brief.cfm?product_id=23

And the v3 CDA Reporting specification:

HL7 IG for CDA R2: Genetic Testing Reports, Release 1 – GTR

http://www.hl7.org/implement/standards/product_brief.cfm?product_id=292

11.7.2.Structural Variations

HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Cytogenetic Model, Release 1

http://www.hl7.org/implement/standards/product_brief.cfm?product_id=364

HL7 Encapsulation of Raw Genomic Data

With a growing stream of raw data in research and clinical environments, it is important to develop approaches to extract subsets that have clinical relevance. Types of data include medical imaging information, health sensor data, and DNA sequences (especially clinically significant variants found in such). Each data type typically has a common format developed by its respective professional community. These data should be encapsulated using such formats in medical records so that they can be referenced as evidence supporting analysis results and be reassessed when needed.

Accordingly, the clinical genomics standard specifications should 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 may be the variants found in these DNA sequences that are associated with responsiveness to drugs relevant to the treatment for that patient.

Clinical Grade-Genomic Data File Standards

There is a lack of adopted standards for clinical Next Generation Sequencing (NGS) based representation of sequence variants and haplotypes in bioinformatics format. To address the lack of data content standards, the US Centers for Disease Control (CDC) together with other federal partners (FDA, NIST, NCBI) established the Clinical-Grade Variant File Specification Workgroup (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5417043/) that includes informaticians, platform and software developers, clinical laboratory directors, translational researchers, representatives from a laboratory accrediting body and the HL7 Clinical Genomics workgroup. The Domain Analysis Model will continue to be informed by and inform these efforts.

Gaps & Extensions

11.8.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 variants and risk assessment. While many laboratories currently use electronic ordering, some laboratory orders are still paper- or PDF-based. However, as genetic analysis becomes a standards part of clinical care, paper-based order entry will not scale.

12. Future Considerations

1. Will electronic health records (EHR) incorporate a genomic repository housing a patient's genome/variome for access on demand, much as images are stored in PACS (picture archiving and communication system), or will EHRs contain a pointer to a centralized repository?
2. Will laboratories continue to sequence a patient's DNA repeatedly for each time a test is ordered, or will a sequence be performed once and many conclusions drawn from the one sequence?

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.

Glossary

aCGH: analyzes specific amplifications and deletions of the genome. aCGH enables the detection of gains and losses of genetic material in specific high resolution regions of genome (10 to 100 times higher than traditional techniques), depending on the analysis platform used. It also provides a more accurate assessment of chromosomal anomalies and gene deletions or rearrangements than FISH.

Biomarker: Short for a biological marker, a site that indicates a specific/distinct biological function that plays a role in infection, cancer, or disease of some kind. Provides a target for drugs.

Genome: Nucleic acid component of the genetic material of an organism. For many organisms it is DNA, but can be RNA in certain viruses.

Germline: Related to cells whose DNA may be passed to the next generation in reproduction.

HLA: Human Leukocyte Antigen – a group of genes that code for surface proteins responsible for a proper immune response. HLA typing is the sequence and check for compatibility of those antigens.

Metabolism: the cellular and molecular processes involved in processing materials and energy to maintain the organism's living state.

NGS: Next Generation Sequencing – methods of sequencing faster and more efficient than traditional Sanger sequencing. Includes massive parallel sequencing (MPS or "shotgun" sequencing) among other techniques. Requires software to decode.

Sequence Variation: General variation from a common DNA reference sequence and synonymous with variant.

Somatic: Related to non-germline cells such that DNA material is not passed on to next generation.

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

Variant: A single change in the typical DNA sequence. Commonly SNPs (single nucleotide polymorphism – or a sequence with one nucleotide different) or mutations.

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

References

1. The 1000 Genome Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393.
2. Alterovitz G., Warner J., Zhang P., Chen Y., Ullman-Cullere M., Kreda D., Kohane I.S. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. Journal of the American Medical Informatics Association : JAMIA. 2015.
3. Warner J.L., Rioth M.J., Mandl K.D., Mandel J.C., Kreda D.A., Kohane I.S.,…& Alterovitz G. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. Journal of the American Medical Informatics Association: JAMIA. 2016;23(4):701-10.
4. Bielinski, S. J., Olson, J. E., Pathak, J., Weinshilboum, R. M., Wang, L., Lyke, K. J., ... & Takahashi, P. Y. (2014, January). Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time—using genomic data to individualize treatment protocol. In Mayo Clinic Proceedings (Vol. 89, No. 1, pp. 25-33). Elsevier.
5. Bianchi DW, Parker RL, Wentworth J, et al. DNA Sequencing Versus Standard Prenatal Aneuploidy Screening. Obstetrical & Gynecological Survey. 2014;69(6):319-321. doi: 10.1097/01.ogx.0000451481.18231.1b.
6. Boyle RJ, Savulescu J. Ethics of using preimplantation genetic diagnosis to select a stem cell donor for an existing person. BMJ: British Medical Journal. 2001;323(7323): 1240-1243.
7. Byron, Sara A., et al. "Translating RNA sequencing into clinical diagnostics: opportunities and challenges." *Nature Reviews Genetics* 17.5 (2016): 257-271.
8. Cherradi, N. (2015). microRNAs as Potential Biomarkers in Adrenocortical Cancer: Progress and Challenges. *Front Endocrinol (Lausanne), 6*, 195 (2015)
9. Caudle, K. E., Dunnenberger, H. M., Freimuth, R. R., Peterson, J. F., Burlison, J. D., Whirl-Carrillo, M., ... & Relling, M. V. (2016). Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genetics in Medicine*.
10. de Jesus Castillejos-Lopez, M., Garcia-Sancho, M. C., Torres-Espindola, L. M., & Perez-Padilla, J. R. (2006). [Contributions of pharmacogenomics to enhance the efficiency of randomized clinical trials]. *Rev Invest Clin, 58*(5), 512-524. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17408112
11. Farrugia, G., & Weinshilboum, R. M. (2013). Challenges in implementing genomic medicine: the Mayo Clinic Center for Individualized Medicine. *Clinical Pharmacology & Therapeutics*, *94*(2), 204-206.
12. Gekas, J., Langlois, S., Ravitsky, V., Audibert, F., van den Berg, D. G., Haidar, H., & Rousseau, F. (2016). Non-invasive prenatal testing for fetal chromosome abnormalities: review of clinical and ethical issues. *Appl Clin Genet, 9*, 15-26. doi:10.2147/TACG.S85361
13. Kirkup BC, Benjamin C., Mahlen S, Kallstrom G. Future-generation sequencing and clinical microbiology. Clinics in Laboratory Medicine. 33.3 (2013) 685-704. doi: 10.1016/j.cll.2013.03.011.
14. Köser CU, Ellington MJ, Cartwright EJP, et al. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. PLoS Pathogens. 2012;8(8). doi:10.1371/journal.ppat.1002824.

15. McGranahan N, Furness AJS, Rosenthal R, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016;351(6280):1463-1469. doi:10.1126/science.aaf1490.
16. Metzgeroth G, Walz C, Erben P, et al. Safety and efficacy of imatinib in chronic eosinophilic leukaemia and hypereosinophilic syndrome – a phase-II study. British Journal of Haematology. 2008;143(5):707-715. doi:10.1111/j.1365-2141.2008.07294.x.
17. Moen, E. L., Godley, L. A., Zhang, W., & Dolan, M. E. (2012). Pharmacogenomics of chemotherapeutic susceptibility and toxicity. *Genome Med, 4*(11), 90. doi:10.1186/gm391
18. Minasi, Maria Giulia, et al. "Genetic diseases and aneuploidies can be detected with a single blastocyst biopsy: a successful clinical approach." *Human Reproduction* (2017): 1-8.
19. Mischak, Harald, et al. Clinical proteomics: a need to define the field and to begin to set adequate standards. *PROTEOMICS-Clinical Applications* 1.2 (2007): 148-156.
20. New, Maria I., et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. *The Journal of Clinical Endocrinology & Metabolism* 99.6 (2014): E1022-E1030.
21. Schulze A, Lindner M, Kohlmuller D, Olgemoller K, Mayatepek E, Hoffmann GF. Expanded Newborn Screening for Inborn Errors of Metabolism by Electrospray Ionization-Tandem Mass Spectrometry: Results, Outcome, and Implications. Pediatrics. 2003;111(6):1399-1406. doi:10.1542/peds.111.6.1399.
22. Sconce EA, Khan TI, Wynne HA, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. Blood. 2005;106(7):2329-2333. doi:10.1182/blood-2005-03-1108.
23. Seegmiller, A. C., Kim, A. S., Mosse, C. A., Levy, M. A., Thompson, M. A., Kressin, M. K., ... & Sinkfield, K. J. (2013). Optimizing personalized bone marrow testing using an evidence-based, interdisciplinary team approach. *American journal of clinical pathology*, *140*(5), 643-650.
24. Sermon, Karen, André Van Steirteghem, and Inge Liebaers. "Preimplantation genetic diagnosis." The Lancet 363.9421 (2004): 1633-1641.
25. Taschner PE, Dunnen JTD. Describing structural changes by extending HGVS sequence variation nomenclature. Human Mutation Hum Mutat. 2011;32(5):507-511. doi:10.1002/humu.21427.
26. Wang, E., Batey, A., Struble, C., Musci, T., Song, K., & Oliphant, A. (2013). Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. Prenatal diagnosis, 33(7), 662-666.

Appendix – LOINC Codes

See Section 12 for a description of LOINC Code use. Additional LOINC codes by CPIC can be found in PMID 27441996 (supplementary material, tables S4 and S5).

12.1.LOINC Code Examples

LOINC	Name	Description/Comments
51963-7	Medication Assessed	A coded medication accessed in a pharmacogenetic test (recommend RxNorm).
51964-5	Drug Efficacy Analysis Overall Interpretation	Overall predicted phenotype for drug efficacy for all DNA Sequence Variations identified in a single case. LOINC Answer List values can be seen in table below.
51967-8	Genetic disease assessed	A coded disease which is associated with the region of DNA covered by the genetic test (recommend SNOMED).
51969-4	Genetic analysis summary report	Narrative report in disease diagnostic-based format, which is used for pharmacogenomic reporting as well and disease risk or diagnosis. These reports currently follow the same formatting recommendations.
51971-0	Drug metabolism analysis overall interpretation	Overall predicted phenotype for drug metabolism for all DNA Sequence Variations identified in a single case. LOINC Answer List values can be seen in table below.
53039-4	Genetic Disease Analysis Overall Carrier Interpretation	Carrier Identification interpretation of all identified DNA Sequence Variations along with any known clinical information for the benefit of aiding clinicians in understanding the results overall. LOINC Answer List values can be seen in table below.

12.2.

12.2.LOINC Answer Lists

LOINC	Sequence	Answer text	LOINC Answer Code
51964-5	1	Responsive	LA6677-4
	2	Resistant	LA6676-6
	3	Negative	LA6577-6
	4	Inconclusive	LA9663-1
	5	Failure	LA9664-9
51971-0	1	Ultrarapid metabolizer	LA10315-2
	2	Extensive metabolizer	LA10316-0
	3	Intermediate metabolizer	LA10317-8
	4	Poor metabolizer	LA9657-3
	5	Inconclusive	LA9663-1
53039-4	1	Carrier	LA10314-5
	2	Negative	LA6577-6
	3	Inconclusive	LA9663-1
	4	Failure	LA9664-9

12.3.

12.3.LOINC Pharmacogenetic Interpretation Codes

Source: Caudle et al. 2016

LOINC	LOINC Component	Answer List
50956-2	HLA-B*57:01	Positive vs. negative
57979-7	HLA-B*15:02	Positive vs. negative
79711-8	HLA-B*58:01	Positive vs. negative
79712-6	HLA-A*31:01	Positive vs. negative
79713-4	TPMT gene product metabolic activity interpretation	Metabolizer status
79714-2	CYP2C19 gene product metabolic activity interpretation	Metabolizer status
79715-9	CYP2D6 gene product metabolic activity interpretation	Metabolizer status
79716-7	CYP2C9 gene product metabolic activity interpretation	Metabolizer status
79717-5	CYP3A5 gene product metabolic activity interpretation	Metabolizer status
79718-3	UGT1A1 gene product metabolic activity interpretation	Metabolizer status
79719-1	DPYD gene product metabolic activity interpretation	Metabolizer status
79720-9	CYP2B6 gene product metabolic activity interpretation	Metabolizer status
79721-7	CYP4F2 gene product metabolic activity interpretation	Metabolizer status
79722-5	SLCO1B1 gene product functional interpretation	Functional status

12.4.LOINC Answer Lists for Pharmacogenetic Interpretation Codes

Source: Caudle et al. 2016

Answer List	Answer ID	Answer (CPIC Phenotype Term]
Positive vs. negative		
	LA6576-8	Positive
	LA6577-6	Negative
Metabolizer Status		
	LA10315-2	Ultrarapid metabolizer
	LA25390-8	Rapid metabolizer
	LA25391-6	Normal metabolizer

	LA10317-8	Intermediate metabolizer
	LA9657-3	Poor metabolizer
Functional Status		
	LA25392-4	Increased function
	LA25393-2	Normal function
	LA25395-7	Decreased function
	LA25394-0	Poor function