Lynn Bry, MD, PhD, FCAP
lbry@partners.org

Clinical use cases for Whole Genome Analyses –

Purpose: Describe clinical use cases for whole genome analysis (WGA), including points in workflow where standards can improve processes to enhance test performance and integrated reporting of results with other Pathology and clinical information.

Background:
Clinically used genetic tests frequently employ one of the following methods:

- Probe-based detection to identify single nucleotide polymorphisms (SNPs) or genetic rearrangements (Factor V Leiden, t9:22 in CML).
- Multiplex gene panels or arrays (cystic fibrosis testing, arrayCGH, SnapShot or Sequenom cancer panels).
- Targeted gene sequencing to identify clinically significant variants (developmental delay, cardiomyopathy).
- Whole Genome Analysis: Sequencing of large portions of the human genome, whether in the exome or full genome.

WGA leverages NextGen sequencing technologies to sufficiently cover an individual's genome. These platforms enable testing in a timeframe that may take days to weeks to complete, from sample receipt to reporting. Costs ever decrease but currently run in the high thousands of dollars per genome.

Nonetheless, NextGen technologies are still quite new, particularly when considering their clinical use within CLIA laboratories. Platforms have not been designed nor optimized for clinical testing. Furthermore, the bioinformatics tools used in post-analytic and interpretive pipelines have largely been developed for research purposes, not for clinical testing. CLIA laboratories thus spend significant time and effort validating research tools and developing de novo internal pipelines to handle aspects of quality control, sequence interpretation, and clinical decision support.

With time, these issues should improve. To facilitate clinical use of WGA existing standards for communication of genetic test orders and results should be leveraged. New metadata elements may be needed to document how sites technically perform sequencing and handle bioinformatic analyses. Reporting should incorporate information supporting the interpretation of clinical variants, and means to integrate genetic data with other Pathology and clinical information from the sending site.

At the current time, institutions may use different approaches to perform clinical WGA based on available expertise and resources. Approaches range from performing all aspects of WGA at the institution to outsourcing the sequencing and some or all of the interpretation to a CLIA-certified reference lab.

Clinical use cases for whole genome analyses

(1) Cancer
(2) Germline/constitutional disorders
(3) Infectious diseases
(4) Virtual genotype, re-evaluate existing genetic data
(5) Research methods that may have eventual clinical application (analyses of the microbiome and immunome).

Items in ***bold and italic*** highlight issues where new standards may be required.

(1) Cancer

Uses: Identify genetic variants causing tumor formation, and assist with selection of therapy and management. Current clinical applications for WGA in cancer involve detection of variants not readily identified with existing SNP, cytogenetic, or multiplex/array-based methods.

Workflow: Over the next 3-5 years, WGA will likely remain a deep node within the decision trees that are guided by histopathological evaluation of tumors. Use of WGA will also vary relative to the type of cancer.

From the sending laboratory that received the case:

- Accession the case into a clinical LIS; generate a unique caseID linked to the patient MRN.
    - Cases may include multiple sample types and derivates from primary samples (blocks, DNA, etc.).
    - Each sample in a case should have its own sampleID. However, this standard may vary from LIS to LIS.
- Initial workflow in Pathology includes gross and microscopic analysis, and may include other testing modalities such as immunohistochemistry for defined markers on slide sections, cytogenetics and molecular diagnostic testing to identify presence or absence of targeted mutations, or flow cytometry to phenotype cancer cells.
- Consider WGA for cases where genetic information will greatly assist with diagnosis and/or selection of appropriate therapy and follow-up – currently few applications but the area continues to evolve.

Data to forward to the laboratory performing WGA

- Case/accession information: Sending laboratory's patientID, caseID, sampleID(s) for cancer and germline source tissues. Some testing may request peripheral blood for detection of germline vs. tumor-specific somatic mutations.
- Test(s) ordered
- Indications for WGA.
- Additional information from the sending laboratory:
    - Local histopathological findings.
    - ***Cellularity of the tumor*** (can influence depth of coverage needed).
    - Other local findings regarding tumor phenotype or presence/absence of genetic markers - may influence Tier 1 & 2 bioinformatics and algorithms used to call variants.
    - Other pertinent items from the patient's clinical history.
- ***Clinical and/or research consent to test:*** Many WGA laboratories have their own clinical consent forms and may also include a research consent under an IRB protocol for additional activities conducted at the receiving laboratory.

Reference or Internal WGA laboratory receives the case and evaluates material for adequacy to perform requested testing. If materials pass pre-analytic QC, the laboratory performs sequencing, assembles and/or aligns the sequence data, makes variant calls, identifies signatures associated with cancer vs germline cells, and provides an interpretation.

As use of NGS expands, the business models for receiving laboratories may develop so that some only undertake technical sequencing to a certain stage (assembled genome, variant calls, through to full interpretive report), and allow the

sending laboratory to retrieve the processed data to render a final interpretation. However, given the complexity, costs, and knowledge required in this area, it is unlikely in the next 3 years that anything short of an interpretive report will be requested by sites outsourcing WGA for cancers.

Report to the sending institution should include:

- Case/accession information to link reported data within the sending laboratory's pathology report or from one LIS to another LIS within the same institution, such as the Laboratory Medicine LIS (e.g. Soft, Sunquest or Cerner Millennium) to the Anatomic Pathology LIS (e.g. CoPath, Powerpath).
- Metadata regarding the following:
    - Pre-analytic information: For example, if sample failed QC at the receiving laboratory and could not be tested.
    - Analyses prior to WGA: if the testing laboratory incorporates cytogenetic studies, targeted gene arrays (OncoMAP) or other multiplex analyses as part of WGA to assist with identification of genetic variants or assembly of genome sequence and identification of copy number variants (CNVs).
        - Platform, method(s) used, array version.
    - Sequencing:
        - Platform, version of reagents, method used
    - Parameters relating to "Tier 1 Bioinformatics":
        - Processing, filtering, QC, assembly, alignment: version of the software pipeline or of individual tools used.
    - Parameters relating to "Tier 2 Bioinformatics" (calling of variants)
        - Depth of coverage, reference genome used, version of the reference/curated database(s) used.
- Identified variants: The variant and data supporting its call. Supporting information can refer to elements in curated databases or PubMed reference IDs. Variants should be reported using standard genetic terminology. Pathologists and Medical Geneticists may review newly identified variants and provide case-specific input to assess their clinical significance.
    - Of note - standards for WGA in a CLIA setting may eventually incorporate ***quality parameters regarding strength of the findings***. Development of quality metrics is an active area of research. ***If implemented, structured reports may include these parameters with significant variants.*** The following types of quality parameters may be considered:
        - (1) A parameter reflecting depth of coverage where a clinically significant variant was identified. Coverage may need to be 1000X+ for cancer testing to insure adequate depth to detect mutations, particularly when tumor cells represent a minor population of a sample's cellularity. Based on input cellularity and tumor type, different quality metrics could be applied to ascertain appropriate depth of coverage.
        - (2) A probability score incorporating coverage and additional computational algorithms to indicate strength of the finding.
        - (3) Simplest may be an indication that the variant passed minimum standards for reporting, presuming labs will not report variants that fail internal metrics.
- Interpretive report provided by a qualified Pathologist.
- Genome sequence: We consider it highly unlikely that DNA sequence will be communicated via HL-7. However, ***HL-7 metadata structures could provide***

fields for the testing lab to send sequenceIDs or other information for the sending laboratory to retrieve sequence data from a laboratory server, HIPAA-compliant cloud, or track a uniqueID/hardware signature for information physically shipped to the sending laboratory.

Sending laboratory receives variants and/or interpretive report and integrates with local pathology data to provide an integrated report to the ordering physician(s) and into the patient's EHR and PHR.

(2) Germline/constitutional disorders

Clinical Uses: Traverses medical specialties. Clinical indications continue to evolve. Current applications include identification of highly penetrant, single gene mutations causing a prominent clinical phenotype, particularly where targeted testing is not available, or has not identified the underlying cause. Examples include newborn syndromic disorders and rare inborn errors of metabolism. Examples in pediatric and adult medicine often associate with specific diseases such as genetic cardiomyopathies, developmental delay, and severe inflammatory bowel disease associated with underlying immunodeficiency. Emerging areas include broader use for risk-assessment of complex diseases such as diabetes and neurodegenerative diseases, drug responsiveness (pharmacogenomics), evaluation of infertility and recurrent spontaneous abortions. We can expect broader clinical adoption as sequencing platforms improve, costs decrease, computational tools improve to enable multi-gene risk assessments, and insurers reimburse for testing and interpretation.

Workflow: In most cases a blood sample will be collected to obtain germline DNA. For some testing, particularly in pediatric patients, samples from both parents and the child may be obtained. ***Mechanisms to link samples and the pedigree information need to be implemented.*** The sample(s) may receive a unique accession number at the sending laboratory, linking it to the patient's MRN. For some sending labs, a unique case ID is assigned, and each sample from each family member receives a unique sampleID.

Data to forward to the reference or internal laboratory performing WGA

- Case/accession information: Sending laboratory's patientID, accessionID or caseID.
- Testing ordered
- Indication for WGA. Some or all of the following information may be included.
    - Clinical reasons for testing.
    - Clinical case information.
    - ***Pedigree information.***
    - Other pertinent items from the patient's clinical history.
    - Previous genetic tests and results.
- ***Clinical and/or research consent to test:*** Many WGA laboratories have their own clinical consent forms and may include a research consent under an IRB protocol for additional activities conducted at the receiving laboratory.

Reference laboratory evaluates incoming material for adequacy of ordered testing, performs sequencing, makes variant calls and generally provides an interpretation.

Report to sending institution or LIS includes:

- Case/accession information to link to local information in the patient's record (e.g.

glucose and HbA1c results in type II diabetes; ECG, echocardiogram or laboratory testing used to evaluate cardiomyopathy with likely genetic etiology).

- Metadata regarding the following:
  - Pre-analytic information: For example, if sample failed QC at the receiving laboratory and could not be tested.
  - Analyses prior to WGA: if the testing laboratory incorporates GWAS, arrayCGH or other multiplex testing modalities as part of WGA to assist with assembly of genome sequence or call copy number variants (CNVs).
    - Platform, array version, methods used.
  - Sequencing:
    - Platform, version of reagents used, method used.
  - Parameters relating to "Tier 1 Bioinformatics":
    - Processing, filtering, QC, assembly, alignment: version of the software pipeline, or specific tools used.
    - Receiving laboratory may choose to indicate the SOP for the bioinformatics rather than detailed documentation of its individual components.
  - Parameters relating to "Tier 2 Bioinformatics" (calling of variants)
    - Depth of coverage over regions where variants called, reference genome used, version of the reference/curated database(s) used.
  - Variants identified using standard genetic terminology: The variant, data supporting its call. Supporting data can refer to elements in the curated databases or PubMed reference IDs.
    - ***Quality metrics indicating the strength of the call or that it passed minimum requirements for clinical reporting.***
- Interpretation from a Medical Geneticist, Pathologist, or other qualified medical professional, based on variant findings.
- Genome sequence - We consider it highly unlikely that DNA sequence will be communicated via HL-7. However, ***an HL-7 message could provide metadata structures for the testing lab to send sequenceIDs or other information for the sending laboratory to retrieve sequence data*** from a laboratory server, HIPAA-compliant cloud, or track a uniqueID/hardware signature for information physically shipped to the sending laboratory.

Sending laboratory or LIS receives data and may integrate with additional data in an integrated report to the ordering physician(s) and to the patient's EHR and PHR.

(3) Infectious disease testing

Uses: Sequencing all or portions of the viral genomes of *in vivo* viral populations to identify drug resistance. For HIV and HIC this testing is standard of case as an adjunct to viral load testing. Genome analyses of bacteria, fungi or parasites have utility for epidemiology and infection control purposes but are not standard of care for individual patient testing at this time. Initial use of whole bacterial genomes for individual patient care may be to assist with identification and resistance testing of difficult-to-culture species such as *M. tuberculosis*, malarial species, and certain fungi.

Section (5) describes additional uses of NextGen sequencing to evaluate genetic signatures in complex populations. Two examples include study of the microbiome and immunome, both of which are research-use-only applications at this time.

Workflow: Primary sample received at the sending laboratory could be material submitted for culture, a derivate from a clinical case (paraffin block), or a pure isolate. Sample has obtained a unique accession# at the sending institution, and is linked to the patient's MRN. The accession may have additional culture, stain or other supporting information linked to it.

Data to forward to the reference or internal laboratory performing WGA on pathogens:

- Case/accession information: Sending laboratory's patientID, accessionID
- Testing ordered
- Indications for testing. Some or all of the following information may be included.
    - Clinical reasons for testing.
    - Clinical case information including results from preliminary testing or phenotypic analyses of resistance.
    - Other pertinent items from the patient's clinical history.
- A separate clinical and/or research consent to test could be requested but is not common, particularly as viral genotype assays have been standard of care for many years.

Reference laboratory evaluates incoming material for adequacy of ordered testing, performs sequencing, makes variant calls and provides an interpretation.

Report to sending institution includes:

- Case/accession information to link to accessionID from sending laboratory.
- Metadata regarding the following:
    - Pre-analytic information: For example, if sample failed QC at the receiving laboratory and could not be tested.
    - Sequencing:
        - Platform, version of reagents used, method used.
    - Parameters relating to "Tier 1 Bioinformatics":
        - Processing, filtering, QC, assembly, alignment: version of the software pipeline, or specific tools used.
        - Given the availability of clinical-grade software to make calls predicting resistance, testing laboratories may only need to report the version of the software used.
    - Parameters relating to "Tier 2 Bioinformatics" (calling of variants)
        - Depth of coverage over regions where variants called, reference genome used, version of the reference/curated database(s) used
    - Variants identified using standard genetic terminology: The gene or variant (e.g. *katG* or *inhA* for isoniazid resistance in TB), data supporting its call (can refer to elements in the curated database(s) or PubMed reference IDs).
- Interpretation from a Clinical Pathologist, Infectious Disease specialist, or ABMM-certified individual (American Board of Medical Microbiology).
    - Interpretive information may include a resistance report (HIV genotype or virtual phenotype) or strain typing/designation for pure isolates.
- Genome sequence: Unlikely that receiving laboratories will send genetic data via HL-7.

Sending laboratory receives data and may be incorporated with local data to provide an integrated microbiology report to the ordering physician(s) and to access for Quality

Improvement activities such as infection control monitoring or further reporting to local Departments of Public Health.

(4) Virtual genotype analyses

Use: Consider two clinical scenarios:

(a) A new clinical question arises in a patient who has had WGA performed. Example: A patient with new presentation of thrombophilia where the existing genome could be interrogated for presence of genetic causes (Factor V Leiden, Factor II 20210 A -> G) as well as genetic factors affecting responses to warfarin, clopidogrel or other drugs.

(b) Periodic re-analysis of genetic data, common for patients with genetic disorders in which a specific variant has not yet been identified as the cause of their disease, or as a routine component of the annual physical exam with a specified frequency such as every two years. Example: For patients with a likely genetic cardiomyopathy, the cardiologist may request a re-evaluation of their genetic data prior to future visits to determine if new scientific/medical information has identified significant variants, or changed the interpretation of previously identified variants.

Re-analysis also could be requested for cancer cases, either in the context of recurrence for comparison of the current cancer with the previous cancer, or in the setting of lack of response to chemotherapy. Given the rapidity at which viral genomes evolve *in vivo*, it is not likely to be used for infectious disease testing.

Workflow: Patient's physician requests the genetic workup, which may be a discrete orderable or as part of a "panel" for a given disease state. In the example of new presentation of thrombophilia, the genetic analysis would be included with PT/INR, PTT and other coagulating testing.

If previous WGA results are available, the information system may check if that information has previously been reported (e.g. WGA analysis that identified previously known risk factors), and if so, retrieve the result. For previously identified variants with weaker data regarding clinical application or lack of identified variants, an order is placed to include genome re-analysis relative to the new clinical question.

In this example, the laboratory or group performing the analysis may not have been the entity that performed the technical sequencing and tier 1 bioinformatics.

Data to forward to the laboratory performing the analysis:

- Case/request information: Sending laboratory's patientID, "requestID" (since there is no physical accession associated with a sample).
- Testing/analysis ordered
- Indications for testing. Some or all of the following information may be included.
    - Clinical reasons for the analysis.
    - Clinical case information.
    - Other pertinent items from the patient's clinical history.
- A separate clinical and/or research consent to test could be requested if the laboratory receiving the request has not handled the patient's data previously.

Receiving laboratory evaluates the request, performs some level of QC on the genomic sequence data to ascertain its use for clinical interpretation (e.g. depth of coverage, platform used, any issues with error rates or discriminating pseudogenes from coding

regions).

Report to the sending institution or LIS to include:

- Case-related data to link to local phenotypic data in the patient's record
- Metadata regarding the following:
  - Pre-analytic information: quality of patient's sequence data for analysis, particularly any errors if it failed, and why.
  - If performed - Parameters relating to "Tier 1 Bioinformatics."
  - Parameters relating to "Tier 2 Bioinformatics" (analysis): Version of the software pipeline, depth of coverage, reference genome used, version of the reference/curated database(s) used.
- Variants called using standard genetic terminology including supporting information for their clinical use in directing the patient's care.
- Interpretive report from a Medical Geneticist, Pathologist, or qualified medical professional, depending upon the clinical indication.

Sending site receives the report and may integrate with other local data before providing back to the patient's physician.

Note: "Continuous" reporting may also occur, where a testing laboratory forwards new, clinically-relevant findings on patients to the LIS or EHR of the sending site. ***The receiving site would need to define level of review of incoming updates, if they require sign-off by a specialist prior to forwarding to a physician responsible for the patient's care, and whether certain data may need to be treated as an alert value.***

(5) Research methods that may have eventual clinical application

Additional research applications exist that evaluate genetic signatures within complex populations. Some of these techniques may eventually be used in clinical diagnostics. These applications include 1) study of the microbiome – microbial signatures, commonly of the 16S rDNA gene, or of bacterial transcripts, that provide an indication of the population of microbes in a given host location, such as the bowel, skin or vaginal tract. Though used solely for research purposes, efforts are evaluating use of micrbiome data to predict risk for inflammatory processes (IBD, necrotizing enterocolitis), susceptibility to allergic or auto-immune processes (early bowel diversity to predict susceptibility to atopic diseases) or as an adjunct to assess metabolic capacity of the gut flora in the host.

A second population-based approach studies the "immunome" - deep sequencing across the recombined T cell receptor (TCR) locus, or IgH/antibody loci, Analyses may be used to define normal and/or pathogenic populations of lymphocytes, both to assist with diagnosis and/or follow responses to therapy.

Neither application is currently used diagnostically. Improvements in sequencing platforms (reduced error rates, increased read-lengths) are needed, as well as development of clinical-grade bioinformatics and computational tools, and curated databases to assist in identifying clinical relevant signatures.