The LIS – Molecular / Genomics

Alexis Carter, MD

Physician Informaticist

Department of Pathology and Laboratory Medicine

Children's Healthcare of Atlanta



Disclosures

Financial

- Paid Faculty for American Medical Informatics Association Clinical Informatics Board Review Course
- Employee of Children's Healthcare of Atlanta

Other disclosures

- Association of Molecular Pathology → Informatics Subdivision Chair,
 Board of Directors, Executive Committee
- College of American Pathologists → Member of Informatics
 Committee
- All data shown in this talk have been scrubbed of PHI
 - In the case of variant-level data, one "case" is actually showing variants from several cases to preserve privacy
 - Individual variants shown are shared between many different patients and not individually identifiable

Objectives

- By the end of this presentation, you should be able to describe...
 - The importance of informatics to molecular genetics and genomics information systems
 - Unique challenges for molecular and genomic testing present for information systems
 - Questions you need to ask your vendor BEFORE implementing new systems and software in your laboratory

The Future is Already Here

Critical juncture

- Genomics revolution is underway
- Increased demand for immediate access to data from anywhere

Contrast with

- Increasing security threats including ransomware attacks
- Genomic and other data privacy issues
- Increasing stringency in US Privacy and Security Law

Definitions

 Next generation sequencing uses BOTH bioinformatics and clinical informatics

Bioinformatics

Many definitions → derives knowledge from computer analysis of biological data (Institut Pasteur)

Bioinformatics

Clinical Informatics

 The application of informatics and information technology to deliver healthcare services

The American Medical Informatics Association (AMIA) – www.amia.org

Clinical

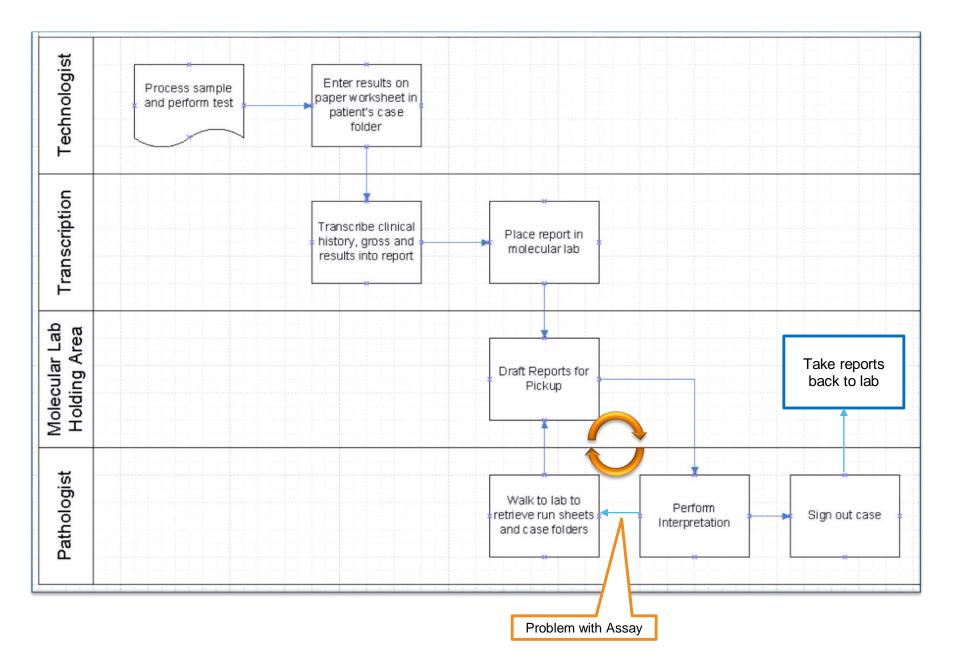
Informatics

Why do we care?

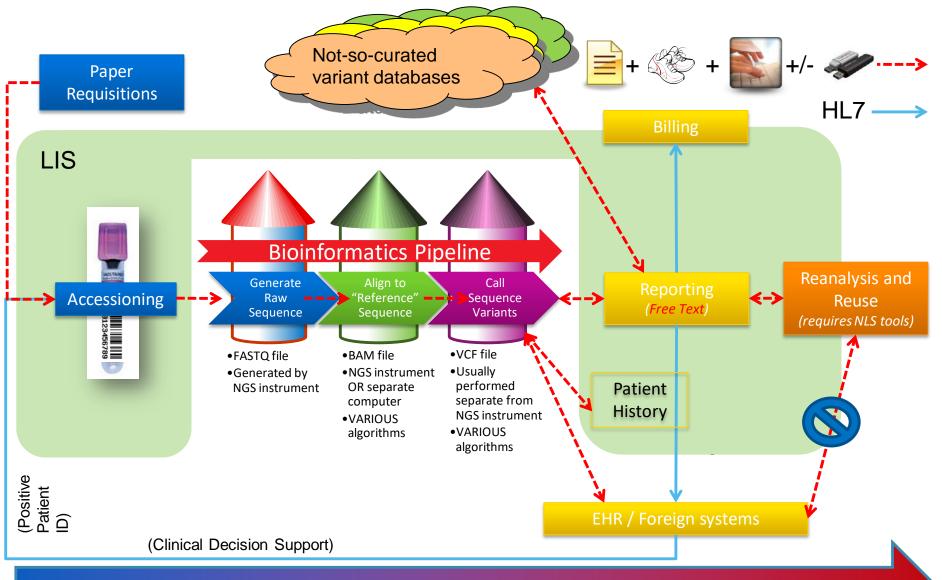
- To my knowledge, no full service* Laboratory Information System is able to consume VCF data without purchase of additional module(s)
 - Several highly specialized separate LISs can consume VCF and even BAM files but...
 - Typically less functional for HL7 interfaces, automation and billing
 - Discrete data may only be housed in the LIS, not in EHR
- How does data get recorded and stored?

* Full service: Able to service chemistry, hematology, microbiology, and transfusion at a minimum (preferably also anatomic pathology)





Where We Are Now



Sample Patient Populations

15. How are you recording nucleic acid quantities for those specimens in which nucleic acid is measured?	Create Chart
	Response Percent
We record the nucleic acid quantity in our LIS in a specific field meant for that purpose	23.3%
We record the nucleic acid quantity in our LIS in a field that is not specific for this purpose (e.g., a comment field)	3.3%
We write the information down on a paper worksheet	57.8%
We do not quantify nucleic acid for any of our studies	8.9%
Other (please specify) Show Responses	6.7%

Recorded on either paper or the technologically advanced thing called "Excel spreadsheet"

Myers C, Swadley M, Carter AB. Laboratory Information Systems and Instrument Software Lack Basic Functionality for Molecular Laboratories. Accepted for publication May 1, 2018, in *Journal of Molecular Diagnostics*. Publication pending.

Children's Healthcare of Atlanta

Molecular / Genomic Differences

- Molecular / Genomic tests are different
 - Physical characteristics of specimens
 - Data types
 - Functional requirements
 - Interpretation
 - Legal requirements

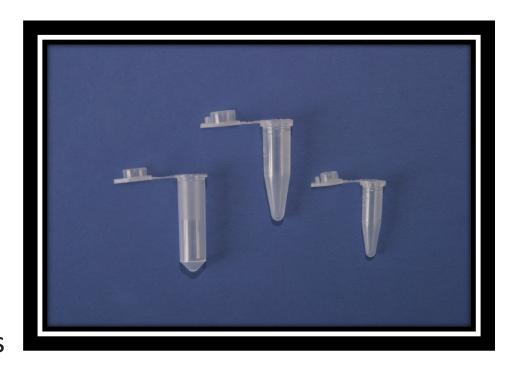
Physical Characteristics of Specimens

Preanalytic - Preparation of sample

- Long term storage of frozen specimens tracking and retrieval
 - Different specimen types may be stored (e.g. buffy coat, pure DNA, pure RNA, serum, plasma, etc.)
 - Specimens-per-unit-area of storage is higher and storage is longer (years) than for most other labs

Preanalytic - Preparation of sample

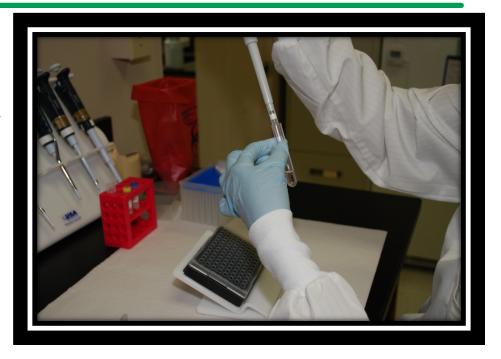
- Labeling of derivative containers (image)
 - Need small labels for microfuge tubes
 - 2.0 mL tubes (larger)
 - 1.5 mL (most common)
 - 500 μL tubes (small)
 - Can label sides or top (preferably sides)
 - Markers vs. sticky labels



Aliquots

- Labeling of derivative containers
 - Cannot cover all sides of tubes
 - Labels need barcodes
 - Too small to support 1D barcodes
 - 2D barcodes are better
 - Asset tracking is a problem





1D (Code 128; 0123456789)



2D more data smaller real estate more accurate, less scan failures

Data Types

Analytic - Targets

Human

- Germline mutations inherited disorders
 - Higher proportion of variant → Lower minimum depth of coverage
- Acquired mutations cancer and other diseases
 - Lower proportion of variant → higher minimum depth of coverage
 - Diagnostic, prognostic, therapeutic

Non-human

- Organism identification, quantification
- Comparison of normal flora vs. abnormal flora (e.g., gut microbiota)
- Mutation detection in organism for drug therapy,
 resistance testing

Preanalytic - Accessioning

- A place to list patient's genetic counselor
- May need to register patients that relate to other patients
 - Donor-recipient testing
 - Family studies for hereditary genetic disorders
 - Pedigrees
- Maintaining patient and specimen identity across all steps is not trivial

Preanalytic - Preparation of sample

- Cell counts performed prior to analysis
 - Some assays require minimum cell count (WBCs)
 - Ability to
 - Order the cell count
 - Get the results electronically
 - Alerts for unacceptable cell counts

Data Types

- Preanalytic
 - Cell counts
 - Tumor percentage
 - Pathology Review (yes/no)
 - Microdissection (yes/no)
 - Extraction data
 - Quantity/quality
 - Fixative used

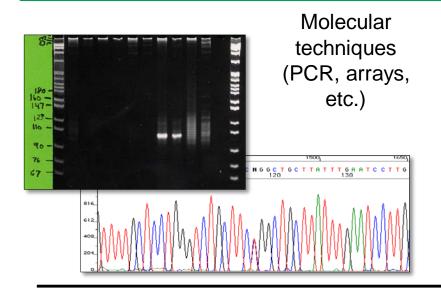
- One test has multiple...
 - Techs
 - Reagents
 - Data types
 - Instruments
 - Hand-offs

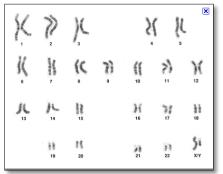
Analytic – Controls

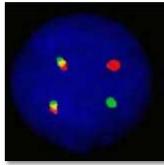
Population-Level

- Allele frequencies (germline testing)
- Run-level
 - Blanks AND negative controls
 - Positive controls (quant/qual)
- Sample-level
 - Internal amplification controls (ladders, housekeeping genes)
 - inhibition controls spiked into split samples
- Variant-level
 - Depth of coverage and others
- Enable flags for controls that fail

Molecular Testing – Many Data Types

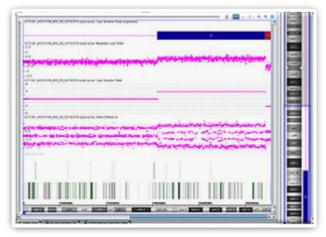




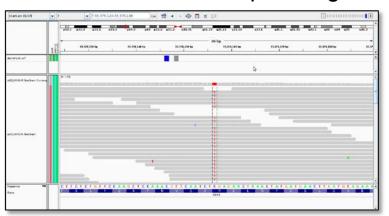


Chromosome analysis (Karyotyping), Fluorescence in situ hybridization (FISH), and other ISH

Microarrays

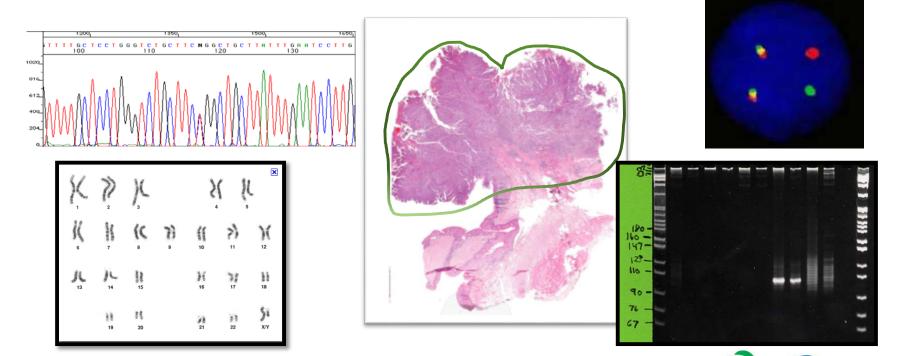


Next-Generation Sequencing



Analytic - Images

- Some LISs still do not accommodate images
- If image module exists, sometimes not associated with molecular module



Functional Requirements

Preanalytic - Ordering

- Consent
 - Did this patient get the appropriate consent for this test?
 - Do we have record of it in the system?
 - Critically important for exomes
- Billing → preauthorization for test?
- Germline genetic tests → no repeats (typically)
 - Catch at order entry rather than at lab accession
 - Complexity of determining duplicates (exome vs. microarray vs. gene panel vs. trinucleotide repeats vs. epigenetics)
- Complex tests often require assistance
 - FISH vs Karyotyping vs PCR vs NGS vs microarray
 - Depends on analytic vs. clinical sensitivity (desired vs. necessary)

Alerts

- Paraffin blocks
 - Paraffin blocks cannot be used for karyotyping
 - Alerts for unacceptable specimens & fixatives4
- Alerts for insufficient or poor quality nucleic acid

Analytic – Batches

- Batched testing for cost and FTE efficiency
 - Batch queues and worksheets
 - Maintaining patient/specimen identity is critical
- Special issue: run multiple targets on single run
 - e.g. human herpesvirus family analysis (HSV1, HSV2, EBV, CMV, VZV)
 - Very different from a single gene/organism target per run approach

Analytic – Calculations and QA

- Extraction completed
 - How much goes into reaction?
- Aliquot tracking
 - Do I have enough DNA/RNA for additional tests?
 - Alerts for when tests requested exceed amount of sample
- NGS has many, MANY performance metrics
 - Need to be tracked in the LIS
 - More on Thursday of this week...

Reporting - Calculations

- Complicating factors
 - Scientific notation
 - Log transformation

1 x 10⁻³ IU/mL

500000000000 IU/mL HCV

11.70 log IU/mL HCV

5000000000 IU/mL HCV

10.70 log IU/mL HCV



Getting Orders to Instruments

14. When your laboratory receives a specimen and logs it into your LIS, how does Create Chart the patient information and the test order get to the instrument that performs nucleic acid extraction?	
	Response Percent
The LIS sends the patient's information and test order <u>electronically</u> to the instrument.	7.8%
The <u>barcode label on the specimen</u> is scanned into the instrument.	25.6%
A <u>barcode label placed on a paper</u> <u>worksheet</u> is scanned into the instrument.	11.1%
Patient information is handwritten on a paper worksheet and <u>manually</u> typed into the instrument.	55.6%

Good infrastructure leverages additional uses for less cost

- Discrete and properly annotated variants in the EHR facilitate:
 - Drug-genome checks, particularly for chemotherapy
 - Alerts to providers when the clinical significance of a variant has been updated
 - Duplicate checking at genome level for germline tests
 - Verification of patient identity
 - Computer-assisted comparison of variants over time (e.g., pre- and post-chemotherapy)
 - Comparison of reporting between pathologists for the same variant

Interpretation

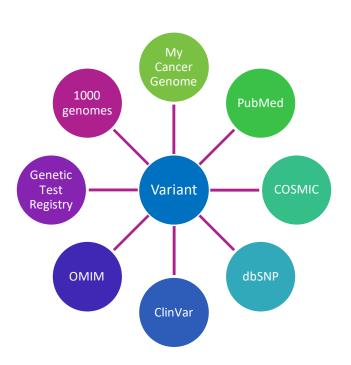
Interpretation and Annotation

- Interpretation is the assignment of clinical significance to the variant
 - In most cases, interpretation must be made by an advanced laboratory professional
 - May occur with or without assistance of other validated tools
 - Basic variant interpretations:
 - Artifact (false positive generated by sequencing process)
 - Benign polymorphism
 - Non-coding and synonymous variants
 - Known pathogenic variant
 - Variant of unknown significance



Annotation and Interpretation

- Only about 20% of variants have known significance
- Other 80% have to be researched



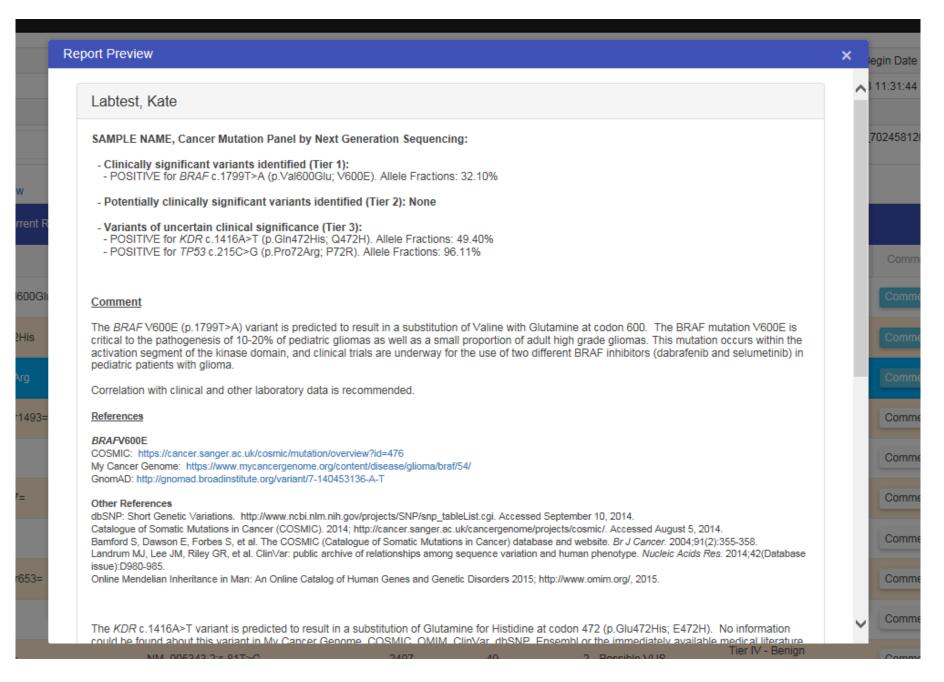
- Must always examine data from multiple sources to determine correct interpretation for variant in its context
 - Likely pathogenic vs. likely benign SNP
 - Germline vs. Acquired
 - Tissue and/or Tumor type
 - Patient's phenotype
- Filter out the "noise"
 - Variants reported as associated with disease when benign SNP has not been excluded
- Amount of noise proportional to amount of data (e.g., My Cancer Genome vs. COSMIC)

Takes a lot of time



Interpretation and Annotation

- Annotation is the labeling of a variant in the context of a particular clinical presentation (e.g., tumor type, tissue of origin, signs, symptoms) for future use in the analysis of other samples
 - Allows linkage of variant to online databases for that variant
 - Laboratories typically lack adequate tools to annotate variants and retrieve those annotations for future analysis



The *TP53* c.743G>A variant is predicted to substitute Arginine with Glutamine at codon 248. This variant has been reported in many tumors in COSMIC, a small fraction of which are thyroid tumors. This variant and other variants in *TP53* have been associated with more poorly differentiated carcinomas of the thyroid. However, this variant has also been reported in several families with Li-Fraumeni syndrome, a hereditary autosomal dominant syndrome that predisposes patients to a variety of cancers including thyroid cancer in some families (see references). This variant comprises approximately half of the alleles in this patient. It is not possible in this case to determine if the variant is germline (inherited) or somatic (acquired as a part of tumor development) in origin. If clinically indicated, genetic counseling followed by testing of germline (non-tumor) tissue can be performed to determine if this variant represents an inherited disorder. Somatic mutations in the *TP53* gene are the most common genetic changes found in human cancer, occurring in about half of all cancers. See results and notes for more information. Clinical correlation is recommended.

TP53 c.743G>A

- Pita JM, Figueiredo IF, Moura MM, Leite V, Cavaco BM. Cell cycle deregulation and TP53 and RAS mutations are major events in poorly differentiated and undifferentiated thyroid carcinomas. *J Clin Endocrinol Metab.* 2014;99;E497-507.
- Ito T, Seyama T, Mizuno T, et al. Unique association of p53 mutations with undifferentiated but not with differentiated carcinomas of the thyroid gland. *Cancer research*. 1992;52;1369-1371.
- OMIM Entry # 151623 LI-FRAUMENI SYNDROME 1; LFS1. 2015; http://www.omim.org/entry/151623. Accessed May 26, 2015.
- COSMIC: http://cancer.sanger.ac.uk/cosmic/mutation/overview?id=10662
- OMIM: http://omim.org/entry/191170#0010
- ClinVar: http://www.ncbi.nlm.nih.gov/clinvar/variation/12356/
- dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=11540652

References

- 1. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2014;42(Database issue):D7-17.
- 2. dbSNP: Short Genetic Variations. http://www.ncbi.nlm.nih.gov/projects/SNP/snp_tableList.cgi. Accessed September 10, 2014.
- 3. Catalogue of Somatic Mutations in Cancer (COSMIC). 2014; http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/. Accessed August 5, 2014.
- 4. Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. 2004;91(2):355-358.
- 5. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(Database issue):D980-985.
- 6. Online Mendelian Inheritance in Man: An Online Catalog of Human Genes and Genetic Disorders 2015; http://www.omim.org/, 2015.

CMP26-APC-c.3880C>T-colorectal CA

CMP26-APC-c.3921_3925delAAAAG-Colon CA

CMP26-APC-c.3949G>C-lung CA

CMP26-APC-c.3949G>C-GIST

CMP26-APC-c.4189G>T-colon CA

CMP26-APC-c.4285C>T-gastric CA

CMP26-APC-c.4321delC-colon CA

CMP26-APC-c.4336G>A-pancreaticobiliary CA

CMP26-APC-c.4588G>T-colon CA

CMP26-BRAF-c.1406G>T-lung CA

CMP26-BRAF V600E-Melanoma

CMP26-BRAF V600E-Colon Cancer

CMP26-BRAF-V600E-Lung Cancer

CMP26-BRAF-V600K-Melanoma

CMP26-CDH1-c.1849G>A-lung cancer

CMP26-CDH1-c.1849G>A-GIST

CMP26-CTNNB1-c.14-4A>G (intronic)-likely benign...

CMP26-CTNNB1-c.98C>G-lung cancer

CMP26-CTNNB1-c,121A>G-melanoma

CMP26-CTNNB1-c.122C>T-metastatic lung CA

CMP26-EGFR-exon 19 DELETION

CMP26-EGFR-exon 19 INSERTION

CMP26-EGFR-c.2573T>G (L858R)-lung cancer

CMP26-EGFR c.2127_2129delAAC

CMP26-EGFR-c.2239 2240delTTinsCC-metastatic lu...

CMP26-EGFR-c.2369C>T (T790M)

CMP26-EGFR c.2387G>T-lung CA

CMP26-ERBB2 (HER2)-exon 20 insertion-lung cancer

CMP26-FBXW7-c.1556A>G-intrahepatic cholangioc...

CMP26-FBXW7-c.1887T>G-lung cancer

CMP26-FGFR2-c.752G> A-melanoma

CMP26-FGFR2-c.766C>T-melanoma

CMP26-FGFR2-c.838G>A-melanoma

CMP26-GNAS-c.2530C>T-colon CA

CMP26-GNAQ-c.740G>A-melanoma

CMP26-BRAF V600E-Melanoma

The BRAF c.1799T>A mutation is predicted to result in a substitution of Valine with Glutamic Acid at codon 600. This mutation is associated with increased sensitivity to BRAF inhibitors.

References

BRAF c.1799T>A

My Cancer Genome: http://www.mycancergenome.org/content/disease/melanoma/braf/54/

CMP26-BRAF V600E-Colon Cancer

BRAF V600E mutations have been well described in colon cancer. This variant was present in approximately % of alleles in this specimen. The presence of the BRAF V600E significantly decreases the likelihood of germline mismatch repair gene mutations in patients with colon cancer, but its implication for therapy remains unknown. Several BRAF inhibitors are currently being investigated in clinical trials for cancer treatment, some of which include colorectal cancer patients. More information regarding the BRAF mutation identified in this specimen can be found at the My Cancer Genome web site (http://www.mycancergenome.org/content/disease/colorectal-cancer/braf/54/).

References

Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. Correlation of tumour BRAF mutations and MLH1
methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features
for MMR variant classification. J Med Genet. 2012;49:151-157.

CMP26-BRAF-V600E-Lung Cancer

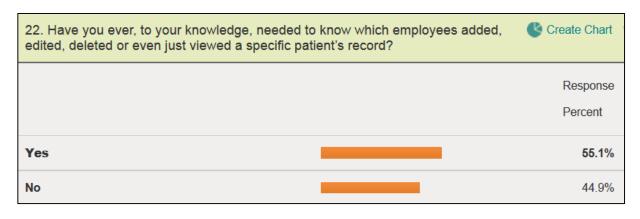
The cytopathologic diagnosis on this specimen is noted. Mutations giving rise to BRAF proteins with increased kinase activity, such as the V600E, have been associated with decreased sensitivity to gefitinib. It should also be noted that BRAF mutations are usually found in tumors wild type for EGFR, ALK, and other driver mutations. Please see http://www.mycancergenome.org/content/disease/lung-cancer/braf/54/ for more details.

26 gene panel = over 250 variant interpretations in word document

Legal Requirements

Information Security

- Why do I care about information security?
 - Patient privacy
 - Protect against identity theft
 - I don't look good in orange





Information Security

Federal Law	Compliance Date
HIPAA Privacy Rule	Apr 14, 2003
HIPAA Final Security Rule	Apr 20, 2005 (April 20, 2006 for small health plans)
Health Information Technology for Economic and Clinical Health Act (HITECH)	Feb 17, 2009 (effective; various compliance dates)
Genetic Information Non- discrimination Act (GINA)	May 21, 2009
HIPAA Omnibus Rule	Sep 23, 2013

HIPAA Final Security Rule

- Security standards for PHI which is transmitted, stored, maintained or used on any electronic media or transmission
 - Electronic PHI (ePHI)
- Three types of safeguards
 - Administrative Safeguards (Policies and Procedures)
 - Physical Safeguards (Hardware, Networks)
 - Technical Safeguards (Software)

HIPAA Final Security Rule

- Many cloud service providers advertise compliance with HIPAA
- Most of them are only compliant with the Physical Safeguards
- The entire system (software, hardware and procedures) must comply with all three categories of safeguards

GINA

- Genetic Information Nondiscrimination Act of 2008 (GINA)
- Generally prohibits employers, group health plans and health insurance issuers from
 - discriminating based on genetic information
 - requesting or requiring genetic testing
 - collecting of genetic information
- These insurance companies are excluded from the requirements:
 - Other forms of insurance like life insurance, long-term care, and disability insurance

GINA

- Defined "genetic information"
 - Genetic services: genetic tests, genetic counseling, or genetic education
 - Genetic tests: analysis of human DNA, RNA,
 chromosomes, proteins, or metabolites, if the analysis detects genotypes, mutations, or chromosomal changes
 - Does not include an analysis of proteins or metabolites directly related to a manifested disease, disorder, or pathological condition

HIPAA Omnibus Rule

- Requires "Genetic Information" (as defined by GINA) to be treated as PHI under HIPAA
 - HUGE implications for research on genetic material
 - Implications for data in the cloud
 - Controversy over whether "genetic information" definition excludes sequences/variants which are not individually identifiable (e.g., BRAF V600E)

Specific Risks to Genetic Information

- Identifying an individual from their genetic information seems far-fetched except...
 - Online genealogy databases are growing rapidly
 - Identification of the "Golden State Killer" in California
 - Studies have reidentified patients from these based on last name inference, phenotypic characteristics, etc.
 - Metadata often accompanies the genomic data in various file formats (gender, zip code)
 - 42% of patients were able to be accurately identified in the Personal Genome Project
 - Genetic data found in particular databases (e.g., drug abusers) can have additional inferences about the patient if the genetic data is reidentified

NHGRI Genomic Data Sharing Policy

- Became effective January 2015
- https://osp.od.nih.gov/wpcontent/uploads/NIH GDS Policy.pdf

4. Informed Consent

For research that falls within the scope of the GDS Policy, submitting institutions, through their Institutional Review Boards³⁴ (IRBs), privacy boards,³⁵ or equivalent bodies,³⁶ are to review the informed consent materials to determine whether it is appropriate for data to be shared for secondary research use. Specific considerations may vary with the type of study and whether the data are obtained through prospective or retrospective data collections. NIH provides additional information on issues related to the respect for research participant interests in its *Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications*.³⁷

For studies initiated <u>after</u> the effective date of the GDS Policy, NIH expects investigators to obtain participants' consent for their genomic and phenotypic data to be used for future research purposes and to be shared broadly. The consent should include an explanation about whether participants' individual-level data will be shared through unrestricted- or controlled-access repositories.

For studies proposing to use genomic data from cell lines or clinical specimens³⁸ that were created or collected <u>after</u> the effective date of the Policy, NIH expects that informed consent for future research use and broad data sharing will have been obtained even if the cell lines or clinical specimens are de-identified. If there are compelling scientific reasons that necessitate the use of genomic data from cell lines or clinical specimens that were created or collected after

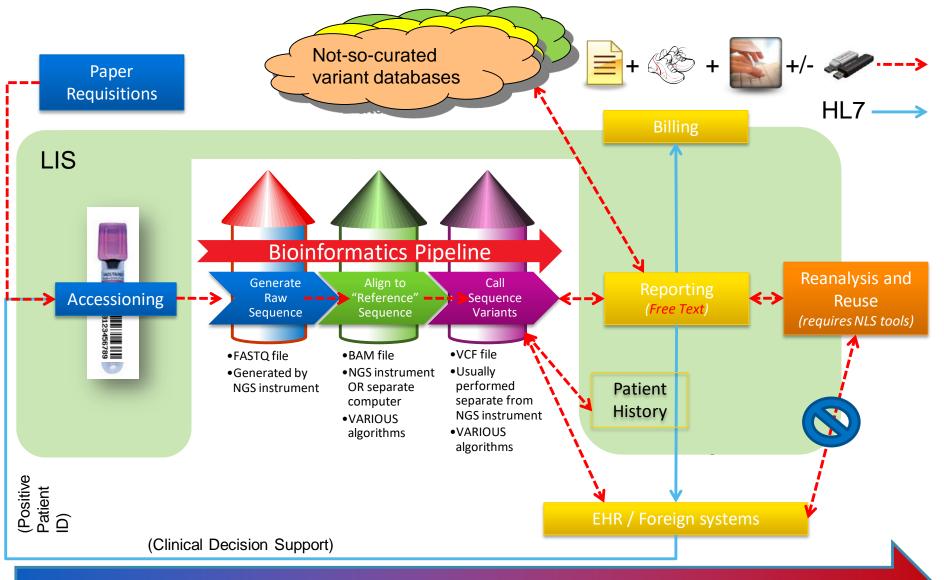
Cloud Services

- Computational requirements → Cloud Services
- Can get storage and processing power as needed
 - Server farms far larger than what an individual organization would have access to
 - Software which manages provisioning and releasing of resources based on various user needs
 - Very helpful for large sets of data (e.g., genomics)
- Charged on pay-per-use model
- Cloud service manages hardware, physical safeguards, and employees to manage systems
 - Very cost efficient when this is taken into account

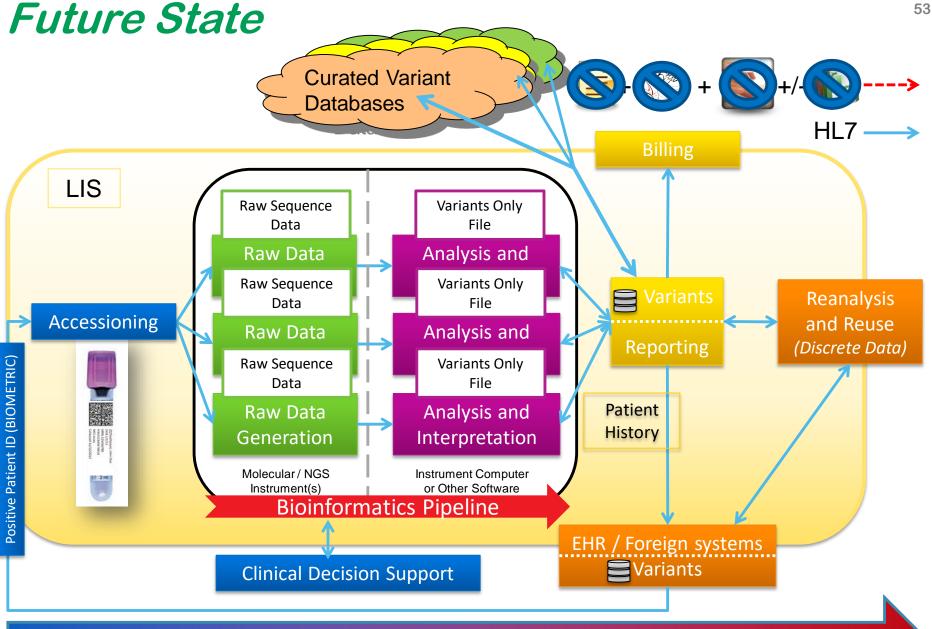
Cloud services

- Know your rights (OCR HIPAA Cloud Policy)
 - Your cloud service provider may NOT deny you access to your PHI ever
 - When your contract expires, Cloud provider must return or securely destroy PHI...they may not retain it
 - Cloud provider is only required to provide assurances that they are following HIPAA, but...
 - You can alter your BAA to require the cloud service provider to provide documentation and audits of compliance
 - Mobile devices accessing PHI on the cloud (or anywhere)
 must also be HIPAA-compliant
 - BAAs have a minimum set of standards (link to those in the references)

Where We Are Now



Sample Patient Populations



Sample **Populations Patient**

Acknowledgements

- Matthew Swadley, MD
- Charles Myers, MD

Questions?