

Genomic Data and Health Information Technology: The New Frontier

Ira M. Lubin, PhD, FACMG

**Laboratory Research and Evaluation Branch
Division of Laboratory Programs, Standards,
and Services, CSELS, OPHSS
Centers for Disease Control and Prevention**

April 17, 2014

Virginia Health Information Technology Standards Advisory Committee

The findings and conclusions in this presentation are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry

Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Agency for Toxic Substances and Disease Registry, the Public Health Service, or the U.S. Department of Health and Human Services

WHO WE ARE

Division of Laboratory Programs, Standards, and Services (DLPSS)

Improve the quality of laboratory testing and related practices in the U.S. and globally through the development and evaluation of innovative training, technical standards, practice guidelines and reference materials.

Some activities

- **Clinical Laboratory Improvement Advisory Committee**
- **Genetics Team**
 - **Next-Generation Sequencing**
 - **Reference materials**
 - **Test ordering / result reporting – test utilization**
- **Laboratory Health Informatics Team**
- **Clinical Laboratory Integration into Healthcare Collaborative**
- **Laboratory Medicine Best Practices**

What Capability Do We Want to Develop?

**Laboratory
quality assurance**

**Databases /
Registries**

**Clinical-decision
Support /
Re-assessment**

**Informatics
analysis**

Patient record

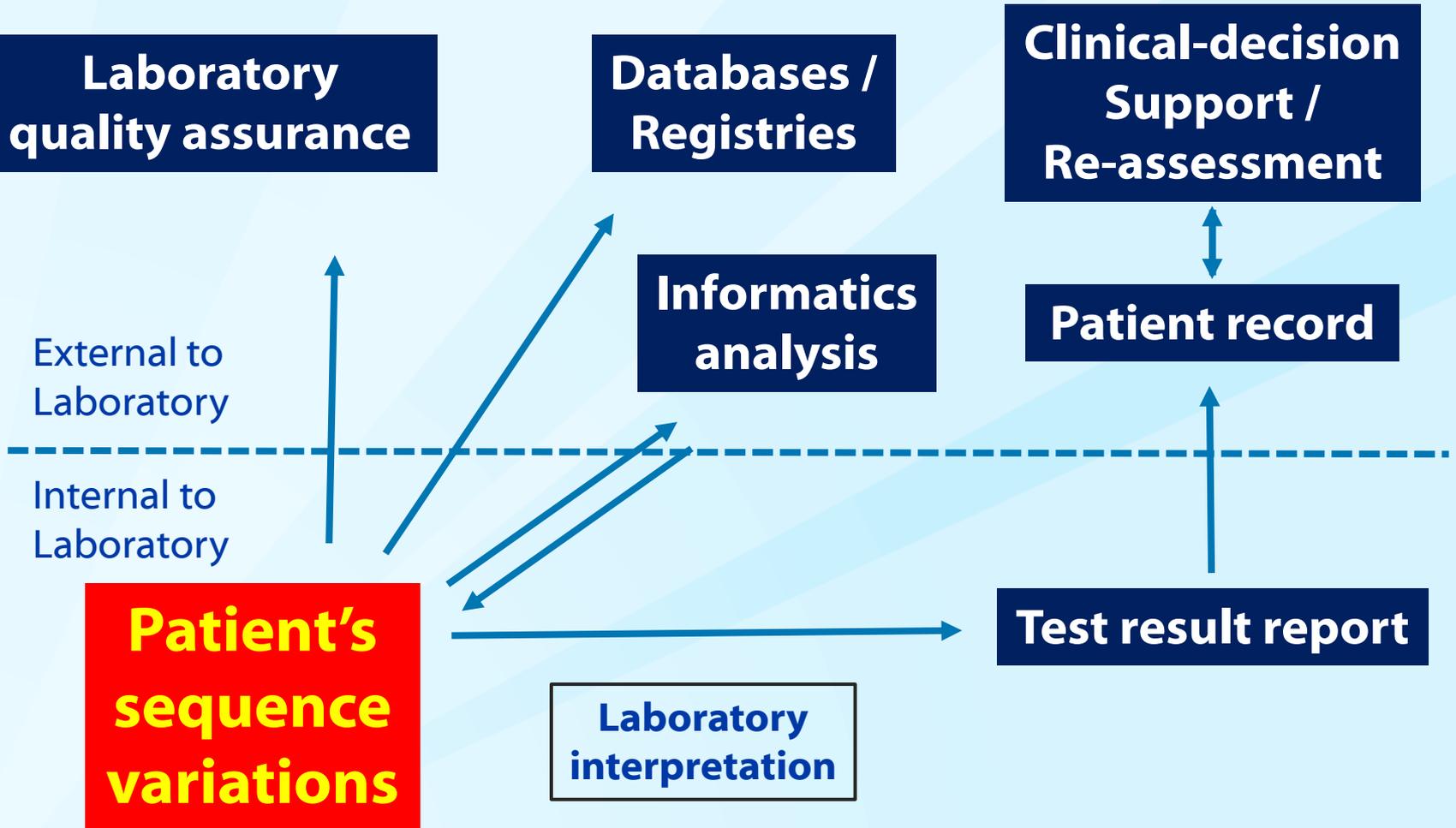
External to
Laboratory

Internal to
Laboratory

**Patient's
sequence
variations**

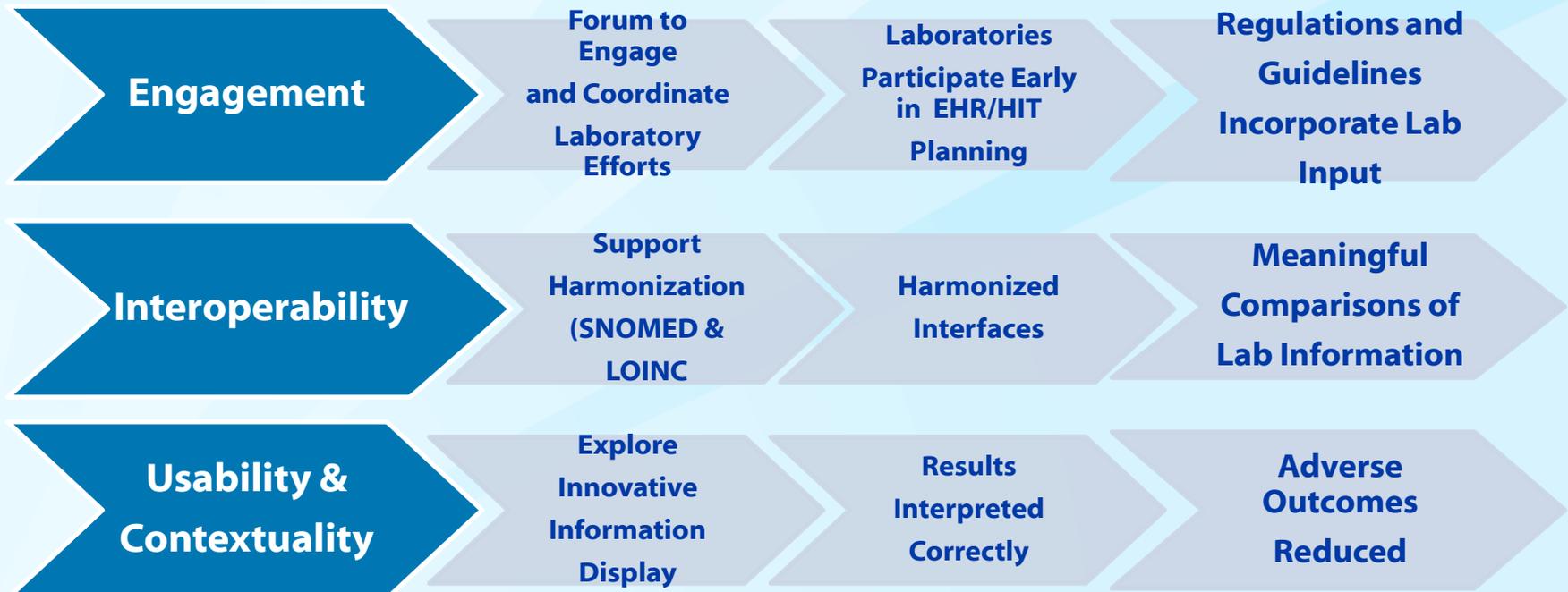
**Laboratory
interpretation**

Test result report



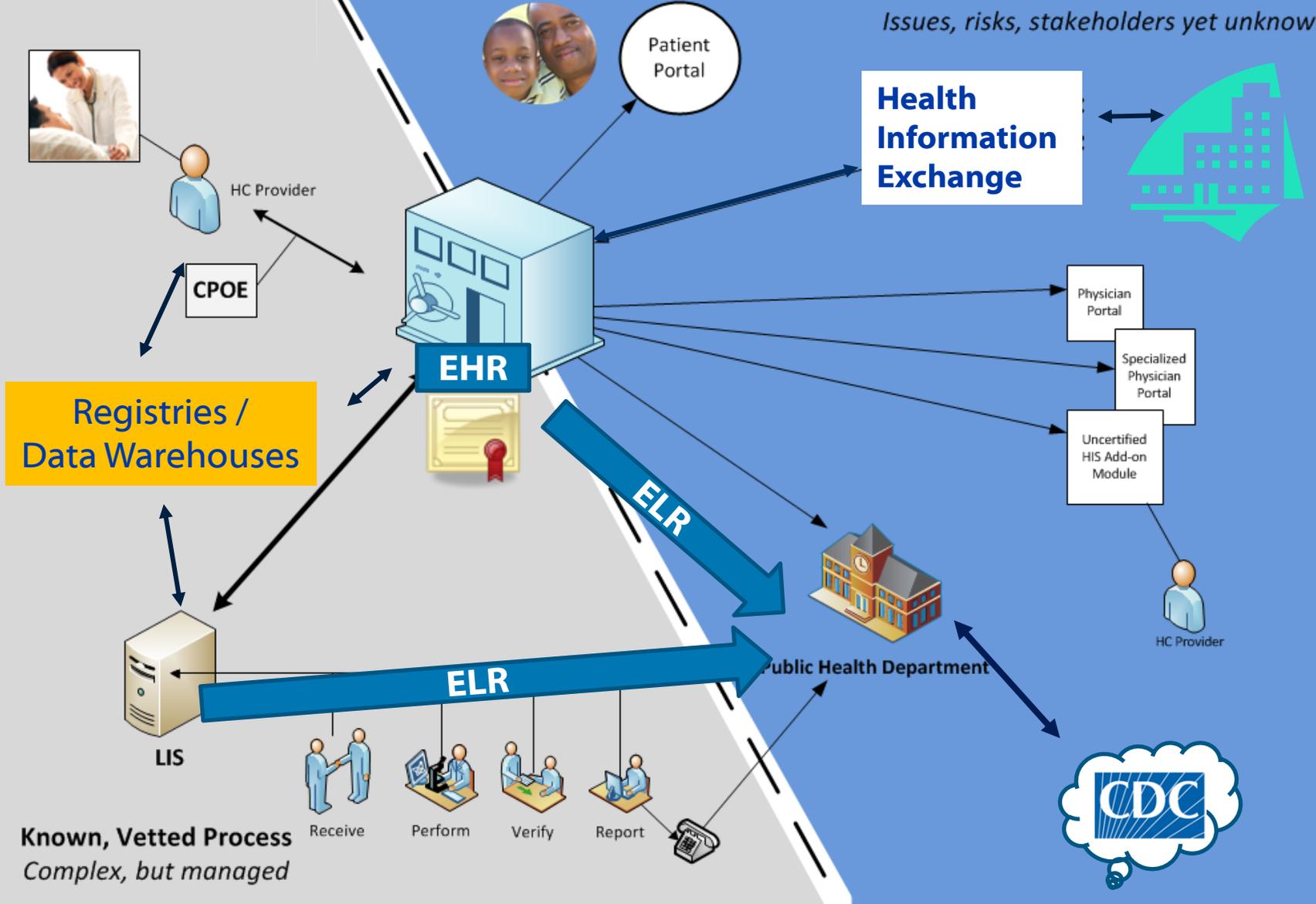
CDC DLPSS Health Information Team

The DLSS LabHIT Team was formed to facilitate the safe and effective integration of laboratory information in the EHR and health information technology for the benefit of patients and healthcare providers.



Laboratory Information Flow

Unknown Healthcare IT Frontier
Issues, risks, stakeholders yet unknown



Known, Vetted Process
Complex, but managed

The Influence of Clinical Genomics

- Diagnose rare diseases
- Directing cancer therapy selection
- Other clinical applications in development

Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease

Elizabeth A. Worthey, PhD^{1,2}, Alan N. Mayer, MD, PhD^{2,3}, Grant D. Sverson, MD²,

Daniel Helbling, BA,
Trivikram Dasu, PhD,
Ulrich Broeckel, MD,
James T. Casper, MD,
John M. Roullet, MD

Identification of a Novel *TP53* Cancer Susceptibility Mutation Through Whole-Genome Sequencing of a Patient With Therapy-Related AML

Worthey E.A. et al. *Genet Med.* 2011;13(3):255-62.

JAMA. 2011; 305(15):1568-1576.

HUMAN GENOMICS

Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing

Callum J. Bell,^{1*} Darrell L. Dinwiddie,^{1,2*} Neil A. Miller,^{1,2} Shannon L. Hateley,¹

SciTransl Med. 2011:3 (65) 65ra4

Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units

Carol Jean Saunders,^{1,2,3,4,5*} Neil Andrew Miller,^{1,2,4*} Sarah Elizabeth Soden,^{1,2,4*} Darrell Lee Dinwiddie,^{1,2,3,4,5*} Aaron Noll,¹ Noor Abu Alnadi,⁴ Nevene Andraws,³ Melanie LeAnn Patterson,^{1,3} Lisa Ann Krivohlavek,^{1,3} Joel Fellis,⁶ Sean Humphray,⁶ Peter Saffrey,⁶ Zoya Kingsbury,⁶ Jacqueline Claire Weir,⁶ Jason Betley,⁶ Russell James Grocock,⁶ Elliott Harrison Margulies,⁶ Emily Gwendolyn Farrow,¹ Michael Artman,^{2,4} Nicole Pauline Safina,^{1,4} Joshua Erin Petrikin,^{2,3} Kevin Peter Hall,⁶ Stephen Francis Kingsmore^{1,2,3,4,5†}

SciTransl Med. 2012:4, 154ra135

Genomics in Clinical and Public Health Microbiology

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Transforming clinical microbiology with bacterial genome sequencing

Xavier Didelot¹, Rory Bowden^{1,2,3}, Daniel J. Wilson^{2,4}, Tim E. A. Peto^{3,4} and Derrick W. Crook^{4,5}

Didelot X. et al., *Nature Reviews Genetics*, 2012

OPEN ACCESS Freely available online

PLoS PATHOGENS

Opinion

Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology

Claudio U. Köser^{1,2*}, Matthew J. Ellington², Edward J. P. Cartwright^{1,2}, Stephen H. Gillespie³, Nicholas M. Brown², Mark Farrington², Matthew T. G. Holden⁴, Gordon Dougan⁴, Stephen D. Bentley⁴, Julian Parkhill⁴, Sharon J. Peacock^{1,2,4,5}

Koser C.U. et al., *PLOS* Vol 8, Issue 8, 2012

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak

Jennifer L. Gardy, Ph.D., James C. Johnston, M.D., Shannan J. Ho Sui, Ph.D., Victoria J. Cook, M.D., Lena Shah, M.Sc., Elizabeth Brodtkin, M.D., Shirley Rempel, R.N., Richard Moore, Ph.D., Yongjun Zhao, D.V.M., Robert Holt, Ph.D., Richard Varhol, M.Sc., Inanc Birol, Ph.D., Marcus Lem, M.D., Meenu K. Sharma, Ph.D., Kevin Elwood, M.D., Steven J.M. Jones, Ph.D., Fiona S.L. Brinkman, Ph.D., Robert C. Brunham, M.D., and Patrick Tang, M.D., Ph.D.

Gardy J.L. et al. *N Engl J Med* 2011; 364:730-739

- Species identification
- Culture-independent microbiology
- Drug susceptibility testing and detecting virulence determinants

A framework for human microbiome research

The Human Microbiome Project Consortium*

HMPC. *Nature* 2012; 486:215-221

Next-Generation Sequencing – Standardization of Clinical Testing (Nex-StoCT)

The screenshot shows the CDC website interface. At the top left is the CDC logo and 'Centers for Disease Control and Prevention'. A search bar is at the top right. The main navigation bar includes 'A-Z Index' and 'LSPPPPO Home'. The page title is 'Laboratory Science, Policy and Practice Program Office'. The main content area features a breadcrumb trail 'LSPPPPO Home > Genetic Testing Quality Practices', social media sharing options, and the article title 'Next Generation Sequencing: Standardization of Clinical Testing (Nex-StoCT) Working Groups'. A highlighted section titled 'Outcomes from Nex-StoCT I published' contains a citation: 'Gargis AS, et al. Assuring the Quality of Next-Generation Sequencing in Clinical Laboratory Practice. *Nature Biotechnology*. 30, 1033–1036 (2012)'. Below this is a 'Background' section with a DNA double helix image and text explaining the transition of NGS from research to clinical settings and the role of CLIA regulations. A 'Nex-StoCT I Working Group' section lists a meeting held in Atlanta in 2011. A sidebar on the left contains a menu with categories like 'ABC's to Public Health', 'Genetic Testing Quality Practices', and 'Nex-StoCT Working Groups'. A contact information box on the right provides details for the 'Office of Surveillance, Epidemiology, and Laboratory Services'.

1. Implementation in a Clinical setting
2. Design/optimization of an informatics pipeline
3. Data representation and messaging

Website: www.cdc.gov/osels/lspppo/Genetic_Testing_Quality_Practices/Nex-StoCT.html

Evolution of Sequencing Capabilities

Sanger

NGS

**Gene
Panels**

Exome

Genome

**Targeted
Gene/Variant
Analysis**

Increasing Complexity

Next Generation Sequencing: The Clinical Workflow

1. Indication for testing
2. Counseling
3. Specimen Collection, Transport, Management, and preparation
4. Sequence analysis
 - a) Library preparation
 - b) Machine sequencing
 - c) Alignment
 - d) Identify sequence variants
 - e) Variant annotation
 - f) Identify relevant variants
 - g) Confirmatory testing
 - h) Clinically relevant variants and result report
5. Communicating results / Counseling
6. Integration into clinical decision making

Physical patient sample

Digital patient sample

Interpreting Sequence Results

Sequence variants (position/type)



Variant
Annotation



- Functional studies
- Predictions (PolyPhen, SIFT, etc)
- Prevalence
- Segregation
- Database representation (e.g., dbSNP, ClinVar, etc.)



Variant Analysis/Classification
(What is clinically relevant?)



Describing a Variant

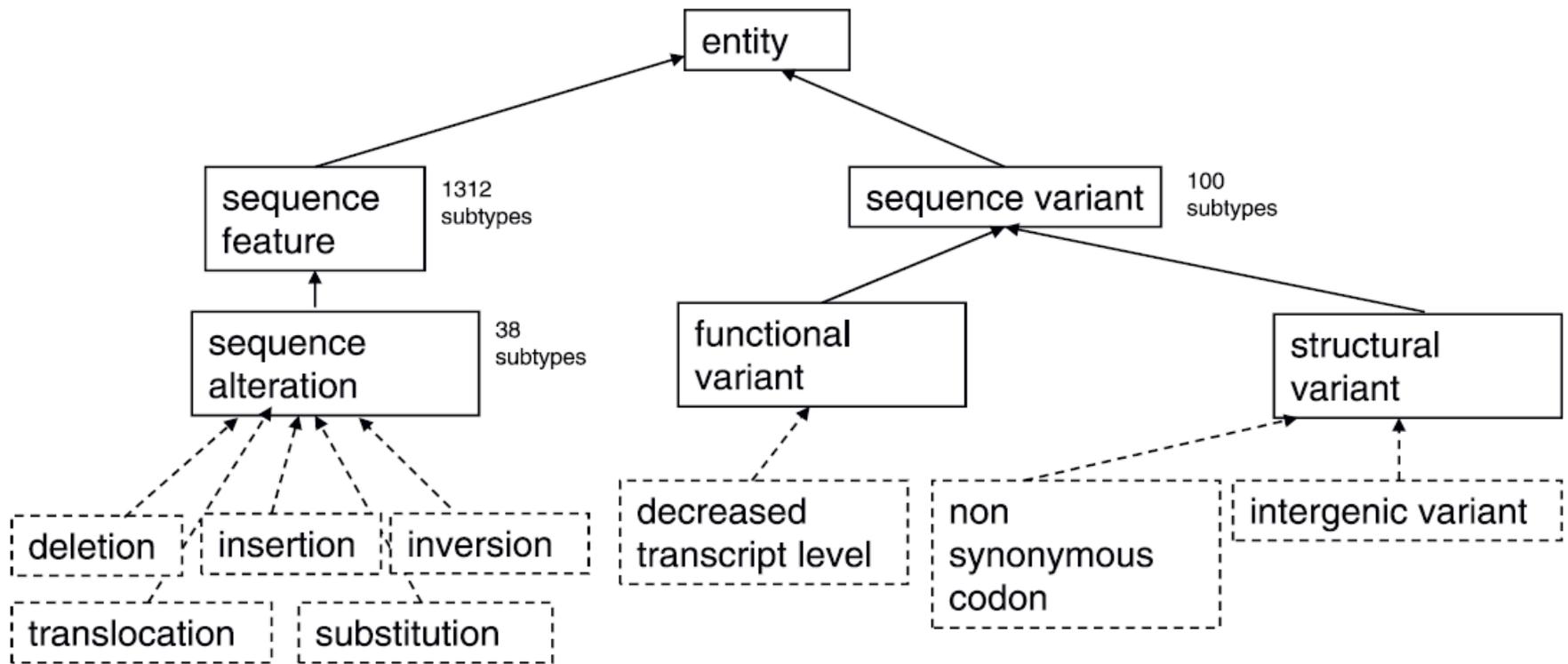


Figure 1 The top-level terms in the Sequence Ontology used in variant annotation. There are 1,792 terms in SO, most of which (1,312) are sequence features. There are 100 terms in the ontology that are kinds of sequence variant, of which the two top level terms are shown, and three sub-types, shown with dashed lines, that demonstrate the detail of these terms. The parts of SO that are used to annotate sequence variation files are sequence alteration to categorize the change (five subtypes shown with dashed lines), sequence feature to annotate the genomic features that the alteration intersects, and sequence variant to annotate the kind of sequence variant with regards to the reference sequence.

Variant Call File Format

genomic
position

Variant
alleles

variant
metadata

phasing

sample
genotypes

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001	NA00002
20	14370	rs6054257	G	A	29	0	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:HQ	0 0:48:1:51,51	1 0:48:8:51,51
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP:HQ	0 0:49:3:58,50	0 1:3:5:65,3
20	1110696	rs6040355	A	G,T	67	0	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT:GQ:DP:HQ	1 2:21:6:23,27	2 1:2:0:18,2
20	1230237	.	T	.	47	0	NS=3;DP=13;AA=T	GT:GQ:DP:HQ	0 0:54:7:56,60	0 0:48:4:51,51
20	1234567	microsat1	G	D4,IGA	50	0	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4	0/2:17:2

Designed to be used within the laboratory and not implemented for the sharing of data with others

Finding The One that is Disease Associated (Crohn disease-like illness – atypical)

Category A

Start here →

High confidence variants	16,124/1,527
Genic variants (variants within genes; i.e., excluding intergenic variants)	16,012/1,504
Insertions	222/72
Deletions	240/136
Substitutions	15,550/1,296
Protein coding variants (variants within the protein coding exons of genes)	15,272/1,407
Insertions	147/65
Deletions	239/119
Substitutions	14,886/1,223
Nonsynonymous variants (variants resulting in an amino acid change)	7,158/879
Insertions	117/51
Deletions	232/112
Substitutions	6,799/706
Substitutions—introduction of a homozygous stop	13/2

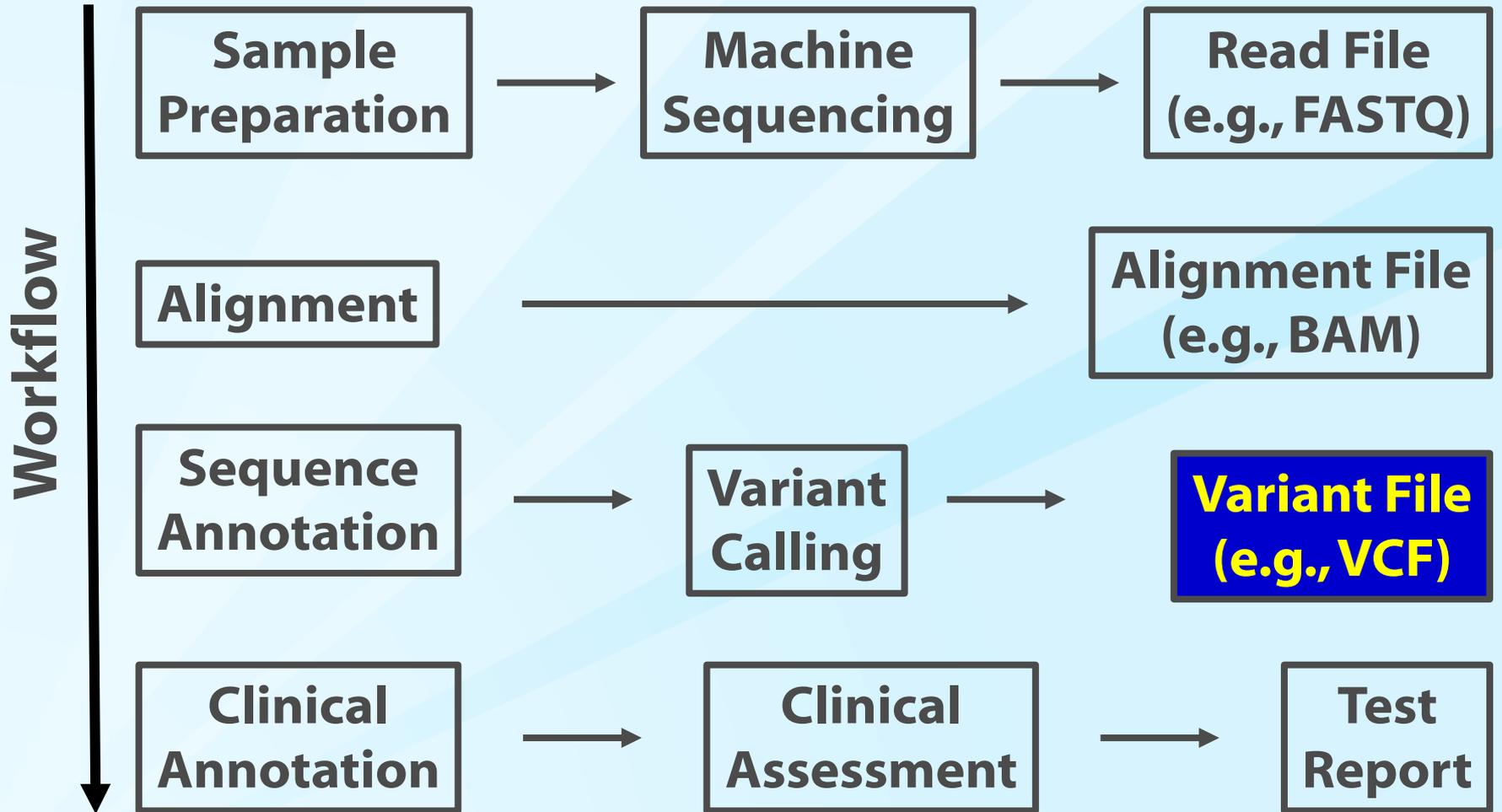
Category B

Variants in genes where two variants were predicted to be damaging	66
Altering highly conserved positions	18
Not known to frequently contain deleterious mutations	4
Novel and confirmed	0
Homozygous or hemizygous	70
Predicted to be damaging	17
Novel (against dbSNP 130)	8
Altering highly conserved positions	4
Not found in reference genome sequences	2
Not known to frequently contain deleterious mutations	1

End here →
(XIAP)

Genet Med 2011;13(3):255–262

The NGS Workflow and Associated Data Files



Clinical-Grade Variant Work Group

Established Clinical-Grade Variant Work Group



current focus

Develop principles and recommendations

- **File format / content considerations**
(e.g., use of genomic coordinates, variant representation, metadata)

Community engagement / feedback (through website)



Develop and evaluate use cases



Pilot in laboratory, clinical settings, HIEs



Share with oversight, standards, and accreditation organizations

Publications

Desired Features of a Clinical-Grade Variant File

- 1. Accurately represent (position , type, reference, quality)**
 - 1. Variant calls**
 - 2. Reference calls**
 - 3. No-calls**
- 2. Represent the full spectrum of variant calls, not just small variants (e.g. CNVs, SVs)**
- 3. Define complete allele (haplotype)**
 - 1. HLA**
 - 2. Pharmacogenomics**
- 4. Provide sufficient meta-data for variant file to be interpreted outside the laboratory**

Metadata (Proposed)

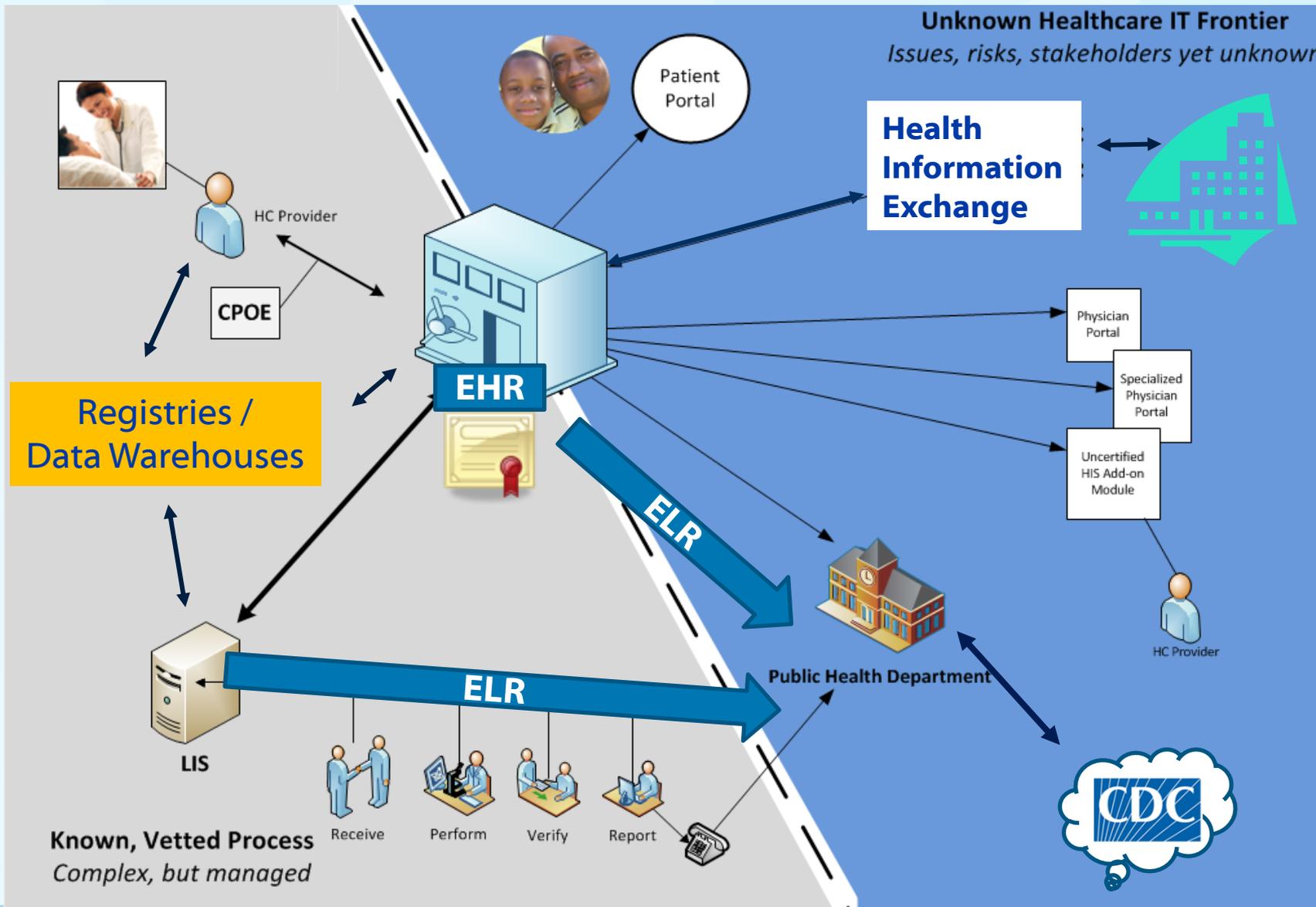
- **General information (Protected Health Information issues?)**
 - **Patient identifier**
 - **Laboratory identifier**
 - **ICD 9/10 codes**
 - **Clinical presentation features**
- **Data relevant to NGS laboratory test procedures**
 - **Instrumentation / Software**
 - **Performance specifications**
- **Data relevant to describing the genome**
 - **Reference assembly / annotation sets**
 - **Regions covered at high confidence**
- **Data relevant to describing the variants**
 - **Standard descriptors (e.g., HGVS, dbSNP, COSMIC)**

How do we develop, pilot, implement, and promote the adoption of a clinical-grade variant file?

Principles

- **Must advance as a community**
- **Must build on the existing infrastructure (to constrain and adapt existing specifications to extent possible)**
- **Must integrate into existing oversight and evolving mechanisms of oversight / professional guidance**
- **Must integrate into evolving health IT structure**
- **Must be flexible to accommodate changing technologies and practices**

Integration into the Evolving Health IT Framework



What Standards Exist

- **File specifications (VCF, gVCF, GVF, etc. – dev. for research)**
- **Reference assembly / annotation sets**
- **HGNC (gene names and symbols)**
- **HGVS Nomenclature (for variant description)**
- **dbSNP – database of short nucleotide variations**
- **RefSeq – standard non-redundant sequence representations**
- **COSMIC – catalog of somatic mutations in cancer**
- **LOINC – coded laboratory and clinical observations**
- **ClinVar (in development) – sequence database linked to health implications**
- **ICD 9/10**

Regulatory / Professional Standards and Guidance

- **Centers for Disease Control and Prevention**
Next-Generation Sequencing: Guidance for the Translation from Research to Clinical Applications
- **American College of Medical Genetics**
Standards for Next Generation Sequencing / Incidental findings
- **College of American Pathologists**
Inspection Checklist
- **HL7 Documents and Guides**
- **Clinical and Laboratory Standards Institute**
MM09 - Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine
- **Association for Molecular Pathology**
Guidance in progress (e.g., test validation)

HL7 Documents

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

C
 HL:

Initial ver:
 Next Balloted V

Chapter Chair and Principal Author:	Mollie Ullman-Dana-Farber C
Chapter Chair and Contributing Author:	Amnon Shabo IBM
Project Chair and Contributing Author:	Grant Wood Intermountain I
Contributing Author:	Kevin Hughes, Massachusetts
Contributing Author:	Daryl Thomas Life Technology
Contributing Author:	Larry Babb Partners Health
Contributing Author:	Lynn Bry, MD Brigham and V
Seeking Additional Co-Authors/Participants	

This document has been updated to re Clinical Genomics Work Group in genetics/genomics community
mullmancullere@partners.com



V3_IG_CANONPED_R1_INFORM_2013APR

HL7 Version 3 Implementation Guide:
Family History/Pedigree Interoperability,
Release 1 – US Realm
 April, 2013

HL7 Informative Document

Sponsored by:
 Clinical Genomics Work Group

Pedigree R1 Co-Editors:
 Dr. Amnon Shabo (Shabo), IBM Research Lab, Haifa; Co-chair & Modeling Facilitator
 Dr. Kevin S. Hughes, Aron Comprehensive Breast Evaluation Center, Massachusetts General Hospital

US Realm IG Co-Editors:
 Dr. Amnon Shabo (Shabo), IBM Research Lab, Haifa; Co-chair & Modeling Facilitator
 Mollie M. Ullman-Cullere, Dana-Farber Cancer Institute and Partners Healthcare
 Dr. Yan Heiss, Louisiana Consulting Group
 Namdi Ihugbo, Life Technologies
 Grant M. Wood, Intermountain Healthcare
 Dr. Kevin S. Hughes
 Dr. Brian Deiban, Comprehensive Breast Evaluation Center, Massachusetts General Hospital

Copyright © 2013 Health Level Seven International. All RIGHTS RESERVED. The reproduction of this material in any form is strictly forbidden without the written permission of the publisher. HL7 and Health Level Seven are registered trademarks of Health Level Seven International. Reg. U.S. Pat & TM Off.

V2IG.CG.LOINCENVAR.R2.INFORM.2013MAR



HL7 Version 2 Implementation Guide:
Clinical Genomics; Fully LOINC-Qualified
Genetic Variation Model, Release 2
 March 2013

HL7 Informative Document: HL7 V2IG CG LOINCENVAR R2-2013
 A Technical Report prepared by Health Level Seven International and registered with ANSII
 5/5/2013

Sponsored by:
 Clinical Genomics Work Group
 Principal Contributors:
 Mollie Ullman-Cullere
 Grant Wood
 Stan Huff
 Clement McDonald
 Amnon Shabo

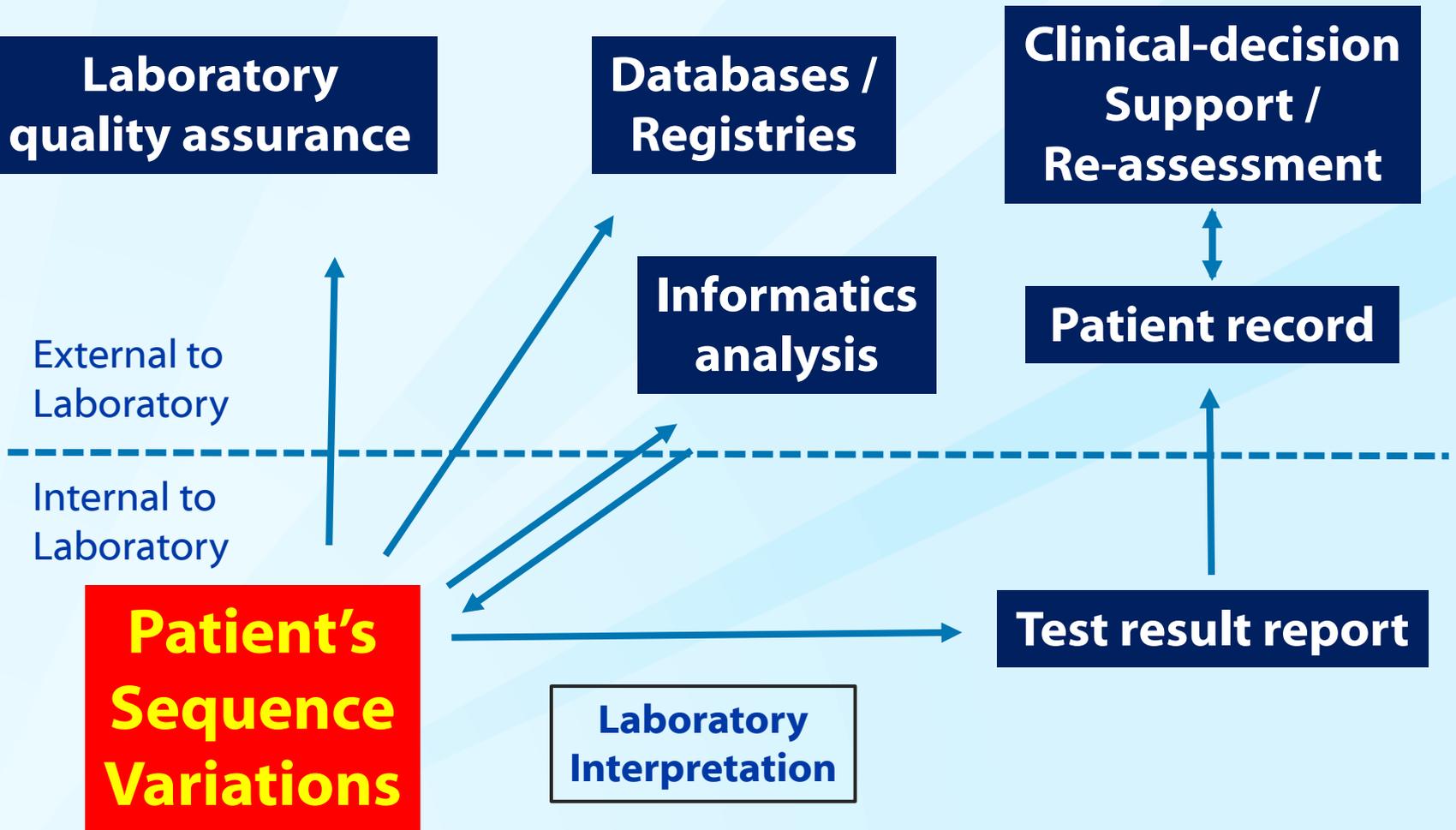
Questions or comments regarding this document should be directed to Grant Wood at grant.wood@hl7.org.

Copyright © 2013 Health Level Seven International. All RIGHTS RESERVED. The reproduction of this material in any form is strictly forbidden without the written permission of the publisher. HL7 and Health Level Seven are registered trademarks of Health Level Seven International. Reg. U.S. Pat & TM Off.

Products Sought

- **Constrained rules for describing variants in the context of the patient and his/her genome**
- **Machine and human readable file format(s) (a balance)**
- **Able to be messaged for a variety of applications**
- **Use cases and pilot studies that provide evidence for utility**

What Capability Do We Want to Develop?



Thank you!

- **Questions about presentation?**

- **Questions for HITSAC**
 1. **What aspects of the presentation are relevant to your vision for integrating genomics into Virginia's evolving healthcare IT infrastructure?**

 2. **What suggestions can you offer to integrate into the scope of our work to serve your present and future needs?**

For additional questions/discussion: Ira Lubin at ilubin@cdc.gov