Binding SNOMED CT and Genomic Data for Cancer Care: An Implementation Story

W. Scott Campbell, PhD, MBA
James R. Campbell, MD

University of Nebraska Medical Center
Nebraska Medicine

Acknowledgements

- College of American Pathologists – Mary Kennedy, Raj Dash, MD, Alexis Carter, MD, Mark Routbort ,MD PhD, Monica de Baca, MD
- Allison Cushman-Vokoun, MD PhD, Tim Griener, MD - UNMC
- Daniel Karlsson, PhD – Linköping University, Sweden
- Royal College of Pathologists and eDigital Health (NHS) – Deborah Drake, Laszlo Iglali, MBBS
- SNOMED International Observable Project – Michael Osborne (Australia), Farzaneh Ashrafi, Suzanne Santamaria, DVM
- Regenstrief Institute Inc. – Daniel Vreeman, DPT, Swapna Abhyankar, MD

Dr. W. Scott Campbell and Dr. James R. Campbell partially supported by NIH Award: 1U01HG009455-01; Patient Centered Outcomes Research Institute (PCORI) Award CDRN-1306-04631); Funding from UNMC Departments of Pathology and Microbiology and Internal Medicine

Outline

- Current environment
- Clinical Terminology Development
- Application to Molecular Pathology

Pathology - Evolving Landscape

- Biomarker data to genomic data
  - Immunohistochemistry (IHC)
  - FISH
  - Sanger Sequencing
  - Next Generation Sequencing
    - Targeted gene loci
    - Whole genome and exome

- Diagnosis to Precision Medicine
  - Histopathology – Diagnose
  - IHC – Differential Diagnoses/Prognosis
  - Molecular Pathology
    - Diagnosis?
    - Prognosis and targeted therapy

Revolution in healthcare

- Sequencing of human genome has led to flood of new observational data appearing in scientific literature and now…clinical records
- Majority of this clinical data is unstructured and not useful for decision support in the EHR or clinical research
- Understanding of genetic and molecular basis of human disease now having impact on diagnosis and therapy
- Challenge posed by precision medicine - to have sufficiently detailed molecular genetics observations and diagnostic data in treatment of cancer, infectious disease and pharmacogenetics

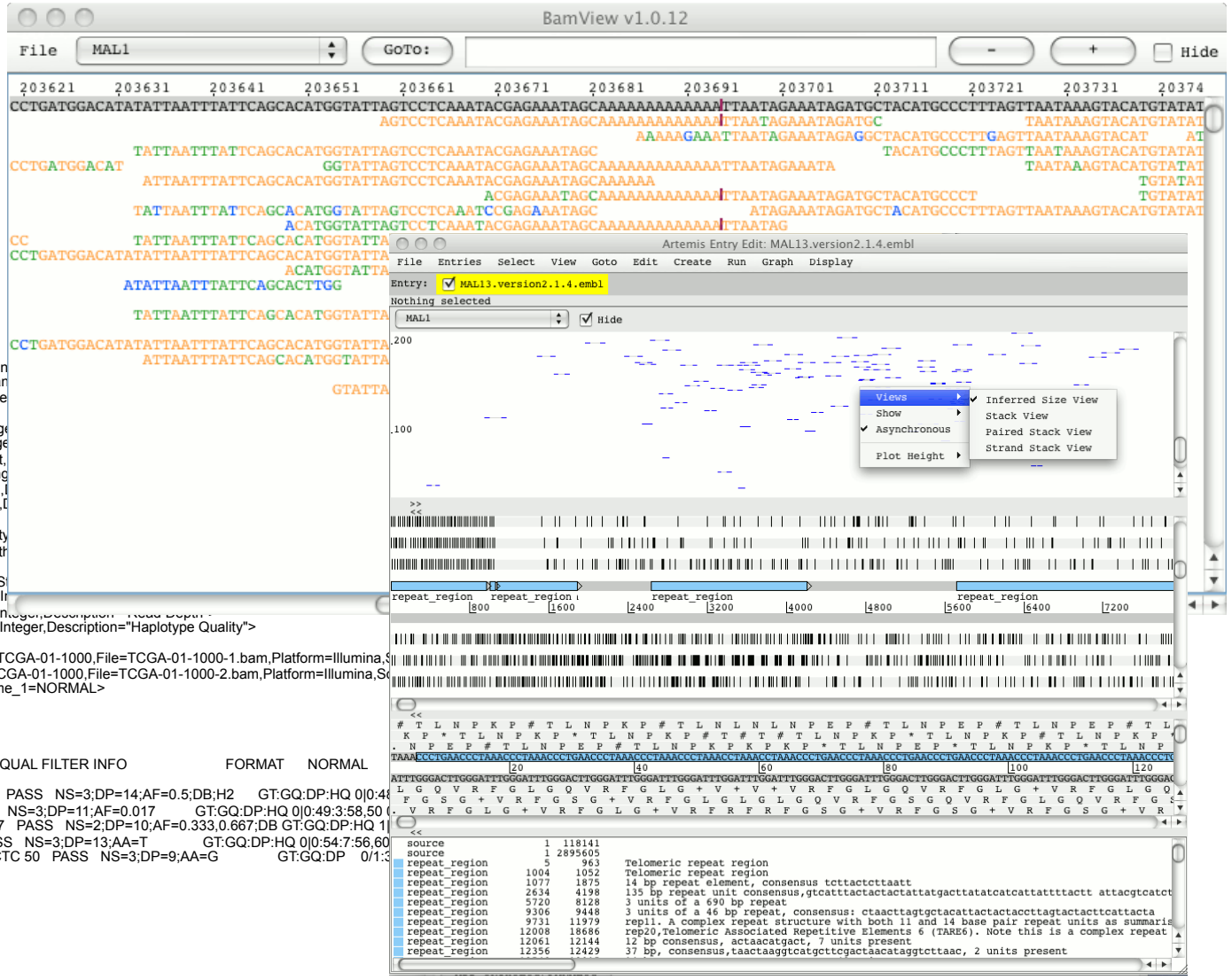
EHR Technology

- What it is:
  - Patient-centric
  - Encounter based
  - Longitudinal Medical Record
  - Transactional
  - Billing orientation

- What it can be:
  - Integrated patient-centric data
  - Basis for clinical decision support
  - Population management

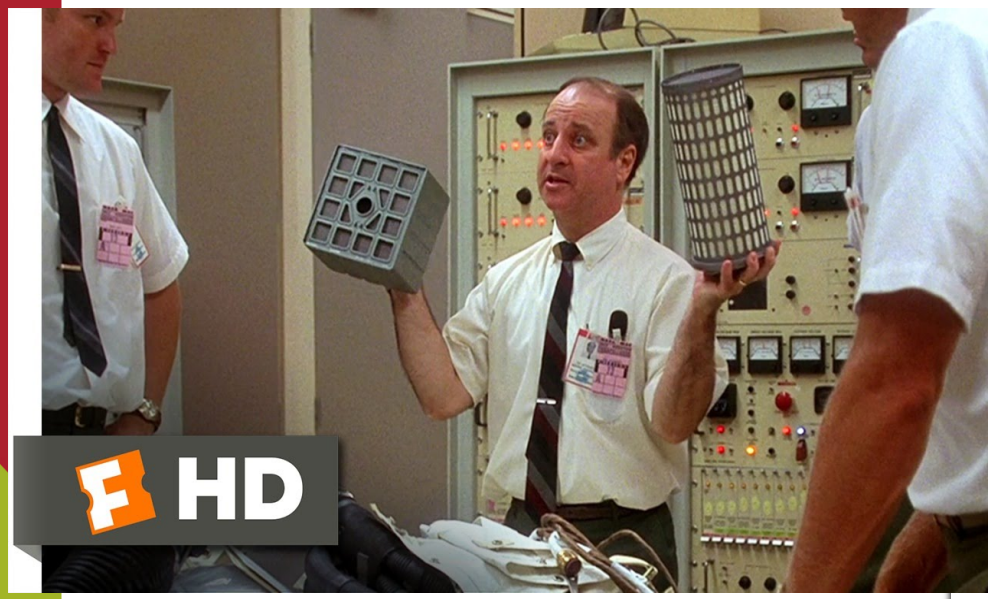
- Requirements (Just a few…)
  - Standards
  - Standard use of standards
  - System Integration/Interoperability
  - Office of the National Coordinator (ONC) directed

Who anticipated this?

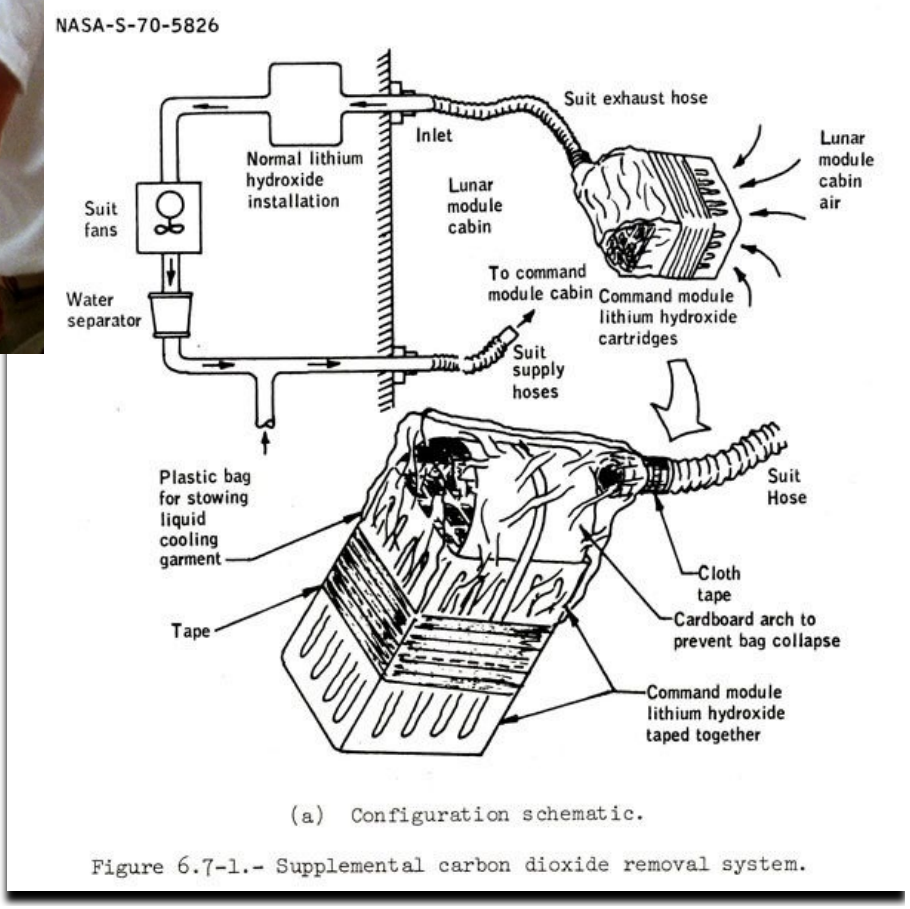



```
##fileformat=VCFv4.1
##fileDate=20140305
##vcfversion=1.2
##vcfProcessLog=<InputVCF=file1.vcf,In
##reference=ftp://ftp.ncbi.nih.gov/genbar
##contig=<ID=20,length=62435964,asse
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Intege
##INFO=<ID=DP,Number=1,Type=Intege
##INFO=<ID=AF,Number=A,Type=Float,
##INFO=<ID=AA,Number=1,Type=String
##INFO=<ID=DB,Number=0,Type=Flag,I
##INFO=<ID=H2,Number=0,Type=Flag,D

##FILTER=<ID=q10,Description="Quality
##FILTER=<ID=s50,Description="Less th

##FORMAT=<ID=GT,Number=1,Type=S
##FORMAT=<ID=GQ,Number=1,Type=In
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">

##SAMPLE=<ID=NORMAL,Individual=TCGA-01-1000,File=TCGA-01-1000-1.bam,Platform=Illumina,S
##SAMPLE=<ID=TUMOR,Individual=TCGA-01-1000,File=TCGA-01-1000-2.bam,Platform=Illumina,S
##PEDIGREE=<Name_0=TUMOR,Name_1=NORMAL>

#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NORMAL

20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;DB      GT:GQ:DP:HQ 1|
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                  GT:GQ:DP:HQ 0|0:54:7:56,60
20     1234567 microsat1 GTC    G,GTCTC 50   PASS   NS=3;DP=9;AA=G                   GT:GQ:DP    0/1:3
```

Houston: We have a problem



How to put a square peg into a round hole?



Taking Inventory of the Situation: Use Cases

Cancer treatment planning

Precision medicine research project planning

Laboratory risk and safety management

Retrieving biobank tissue for research protocols

Examples

Cancer treatment planning:

- Colon cancer with BRAF V600E mutation and KRAS mutation in codon 12. Anti-EGFR therapy contraindicated.

Research Project Planning:

- How many healthy patients do we have that are BRCA1 or BRCA2 positive
- Retrieve all breast cancer cases that tested ER-, PR- and Her2/neu

Medical-legal, compliance and safety:

- Recall of all cancer cases for treatment review with (formerly thought insignificant) somatic gene sequence variant reported
- CAP accreditation requirements for molecular laboratories.

Tissue Biobank Applications:

- Find all malignant neoplasms of any origin tested for BRAF mutation AND reported positive for lymphatic metastases.

What is needed to meet the needs of these researchers and clinicians is a domain ontology (structured, defined coding hierarchy) which we can query for detailed results

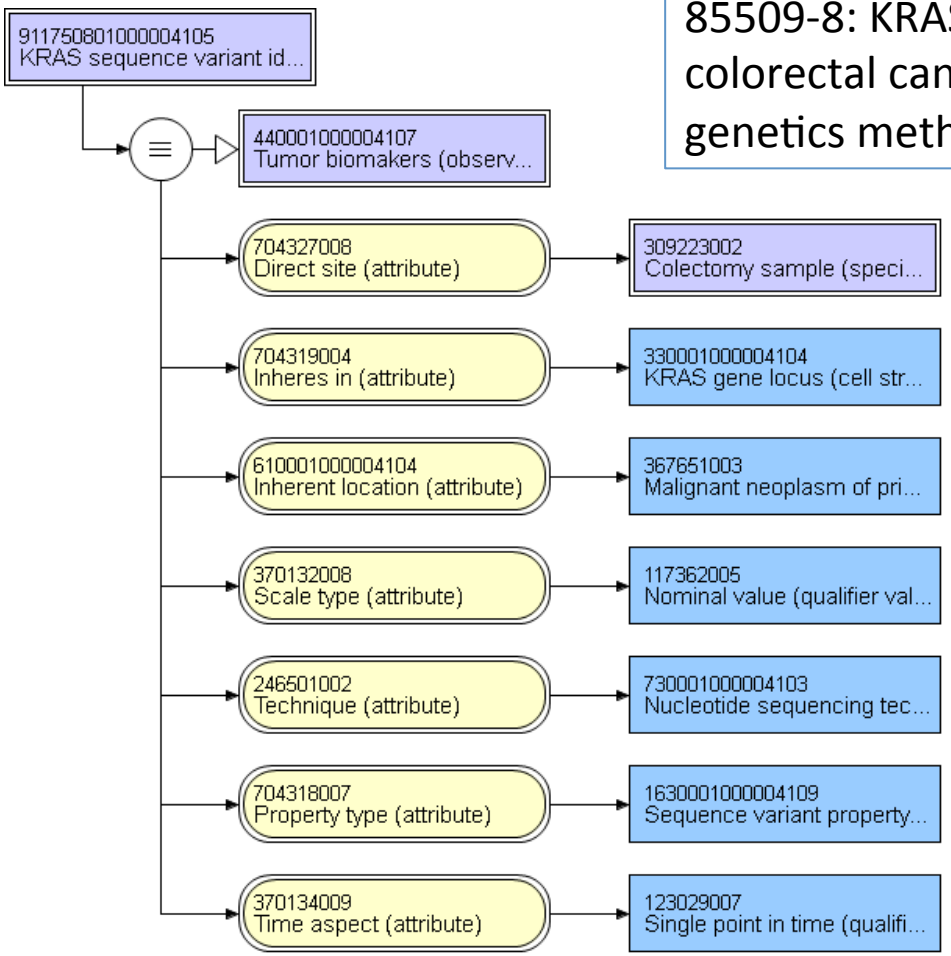
These data types are called "Observables"

Key Terminology Items

- All concepts modeled according to SNOMED-LOINC cooperative agreement
    - LOINC identifiers
    - Full SNOMED CT concept model definitions
        - Observable entity hierarchy

- New SNOMED CT content required
    - Body structures – gene loci definitions for proto-oncogenes
    - Primitive concepts with reference sets to Human Genome Nomenclature Committee (HGNC) identifiers
    - Inherent location of sample (neoplasm vs. non-neoplastic) differentiates somatic vs. germline mutation
    - New Property type –Sequence variant property
    - New Technique – Nucleotide sequencing technique

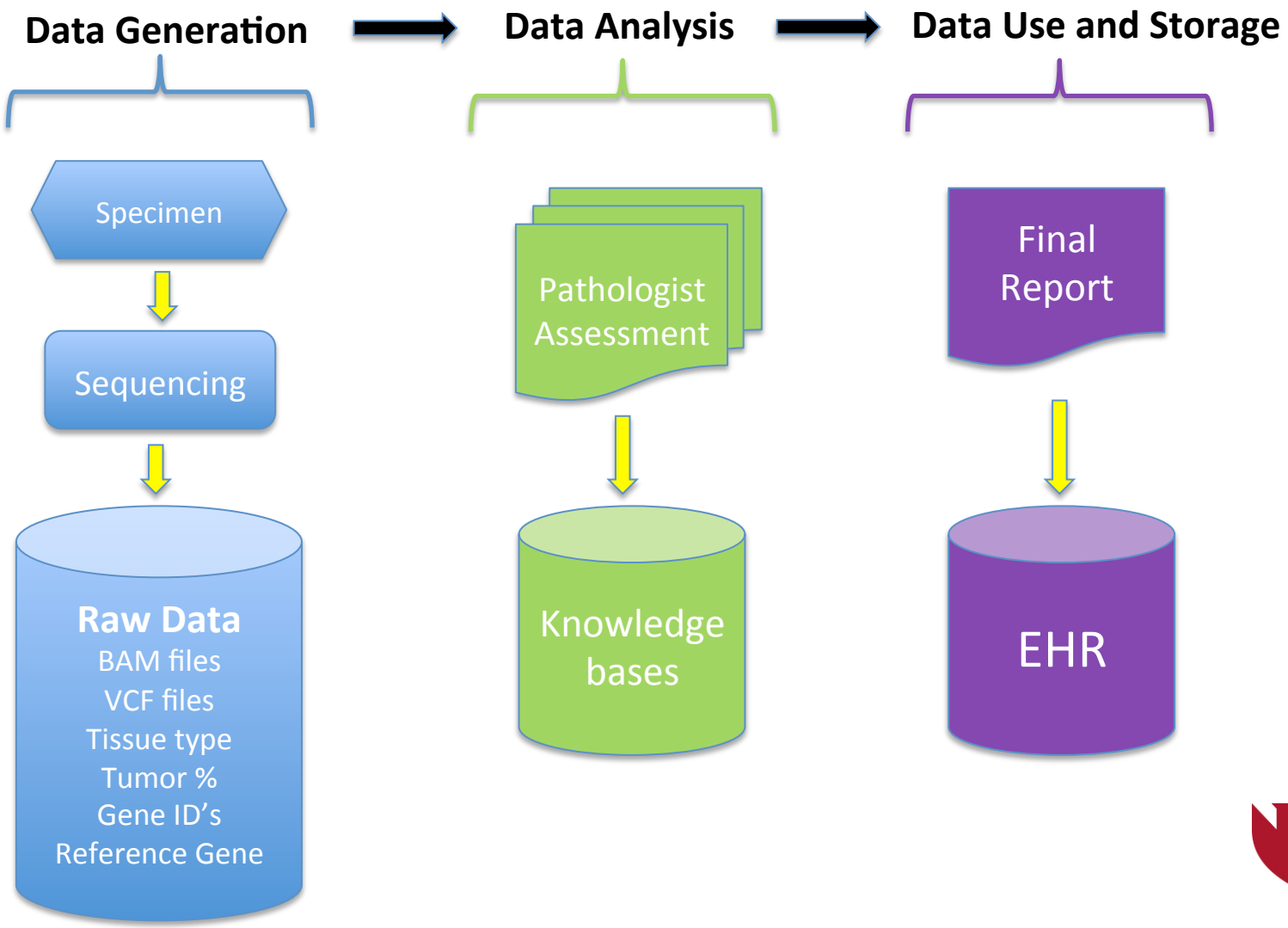
Terminology Approach
KRAS Variant Detected



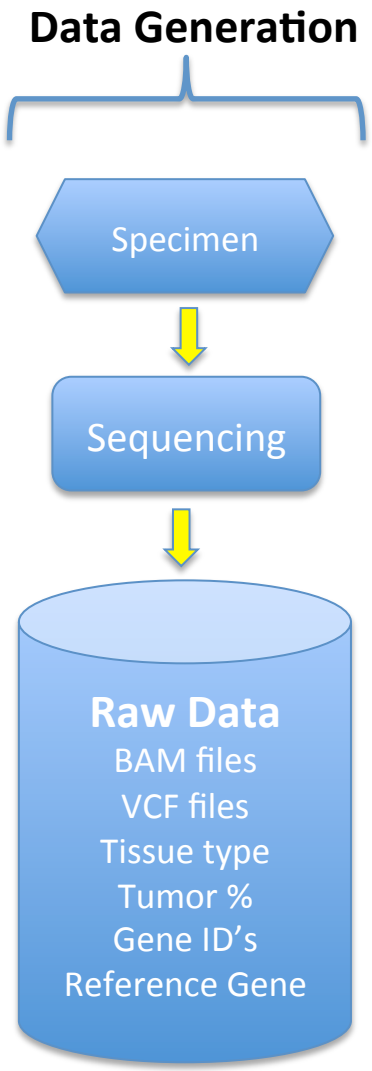
85509-8: KRAS gene mutations found in colorectal cancer specimen by molecular genetics method

How to represent the result?

Molecular Pathology Data Flow



Data Use – Data Generation



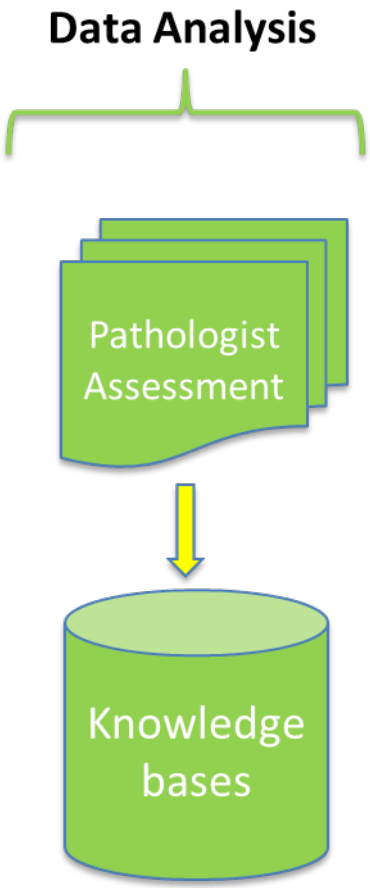
Raw Data (i.e., Primary Data)

Necessary for clinical interpretation

Valuable to Research Community

Value to consumers – clinicians, patients?

Data Use – Data Analysis



Pathologists consider outputs of Data Generation step

Access Knowledge Bases to reconcile data with current genomic understanding

Render assessment of variants detected and should be reported

Classify variants in terms of clinical significance

Data Use –Use and Storage

Final report – pdf format

Enumeration of variants detected classified by clinical significance

Additional items: Type of alteration, Allele frequency (in population), availability of potential clinical trials

Useful for individual clinical encounter but of limited value for extended clinical care, decision making and research.

What to do with all this Data?

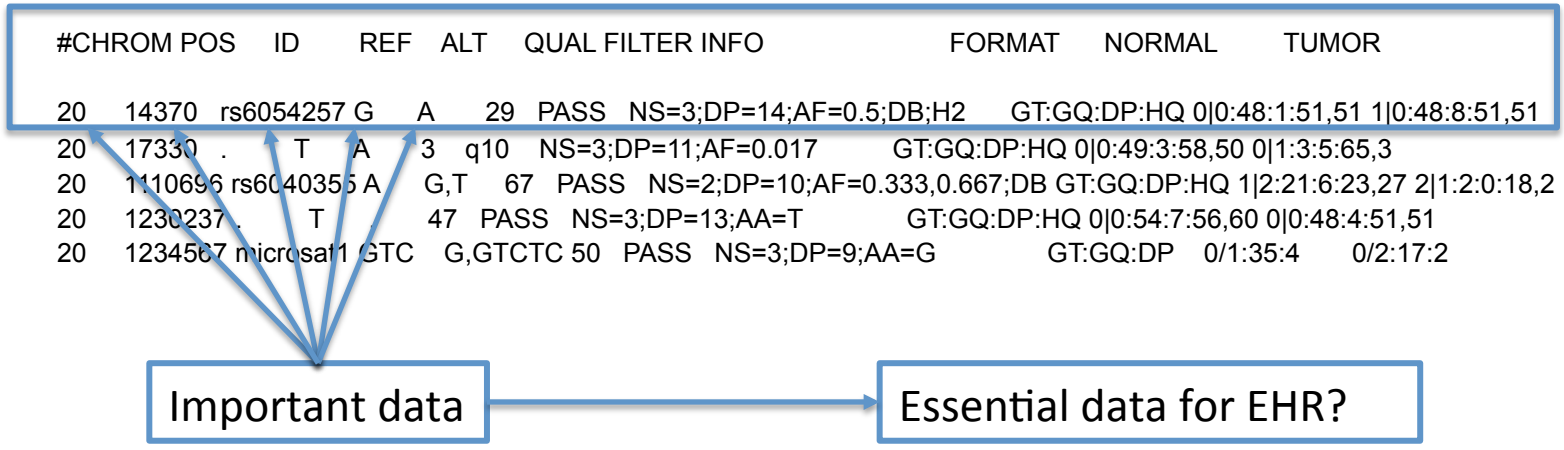
Technical Desiderata for integration of genomic data into the EHR
1. Maintain separation of primary molecular observations from the clinical interpretations of those data
2. Support lossless compression from primary molecular observations to clinically manageable subsets
3. Maintain linkage of molecular observations to the laboratory methods used to generate them
4. Support compact representation of clinically actionable subsets for optimal performance
5. Simultaneously support human viewable formats and machine readable formats in order to facilitate implementation of decision support rules
6. Anticipate fundamental changes in the understanding of human molecular variation
7. Support both individual clinical care and discovery science

1. Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, et al. Technical desiderata for the integration of genomic data into Electronic Health Records. J Biomed Inform. 2012 Jun;45(3):419-22.

More on the Final Report

- Based on the specific gene variants detected and **reported** by the pathologist.
- What elements are needed in the EHR in computable form?
- One-to-one relationship of reported variants and their representation in the variant call file (VCF)
- Can the VCF representation of the reported variants be used as discrete elements in the EHR?

Variant Call Format



Discrete and computable

Contains data that has limited value outside the context of the initial report

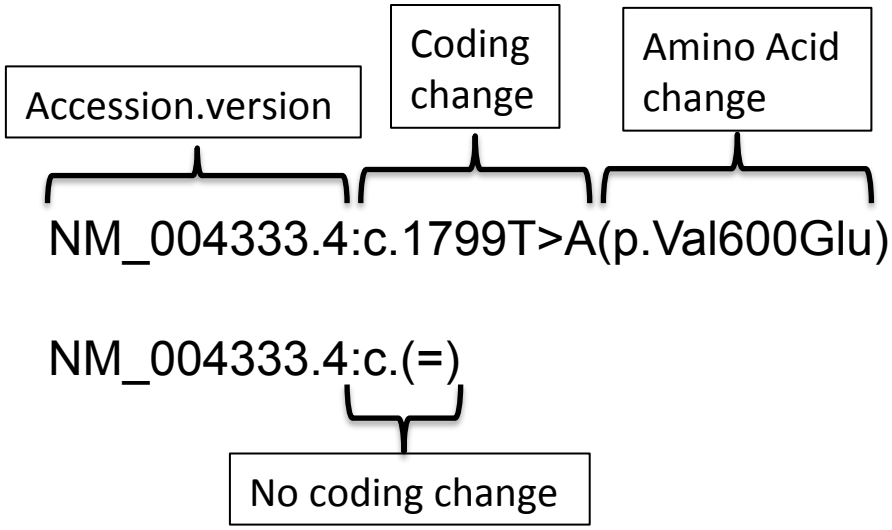
Difficult to import and manage within EHR structures

TOO COMPLICATED FOR EHR – Need something simple

Human Genome Variation Society Representation

- Human Genome Variation Society (HGVS)
  - Standardized nomenclature for variants
  - Includes nomenclature for normal sequences

Examples:



HGVS in HER and Biobank

- HGVS representation of genetic information in EHR is tractable

- Data is a structured string
  - Easily accommodated in EHR
  - Data easily queried by regular expression

- Maintain links to curated gene knowledge bases
  - Standardized representation
  - Used by reference databases, knowledge bases

- Applicable to targeted gene sequencing and whole genome sequencing

- Retains clinically relevant and actionable information

- Can be readily moved via HL7

Moving Molecular Data

HL7 Lab order and results message types, of course

Orders:
Specimen Type – SNOMED CT
Ordered Service – LOINC
Diagnosis – ICD-10 or SNOMED CT

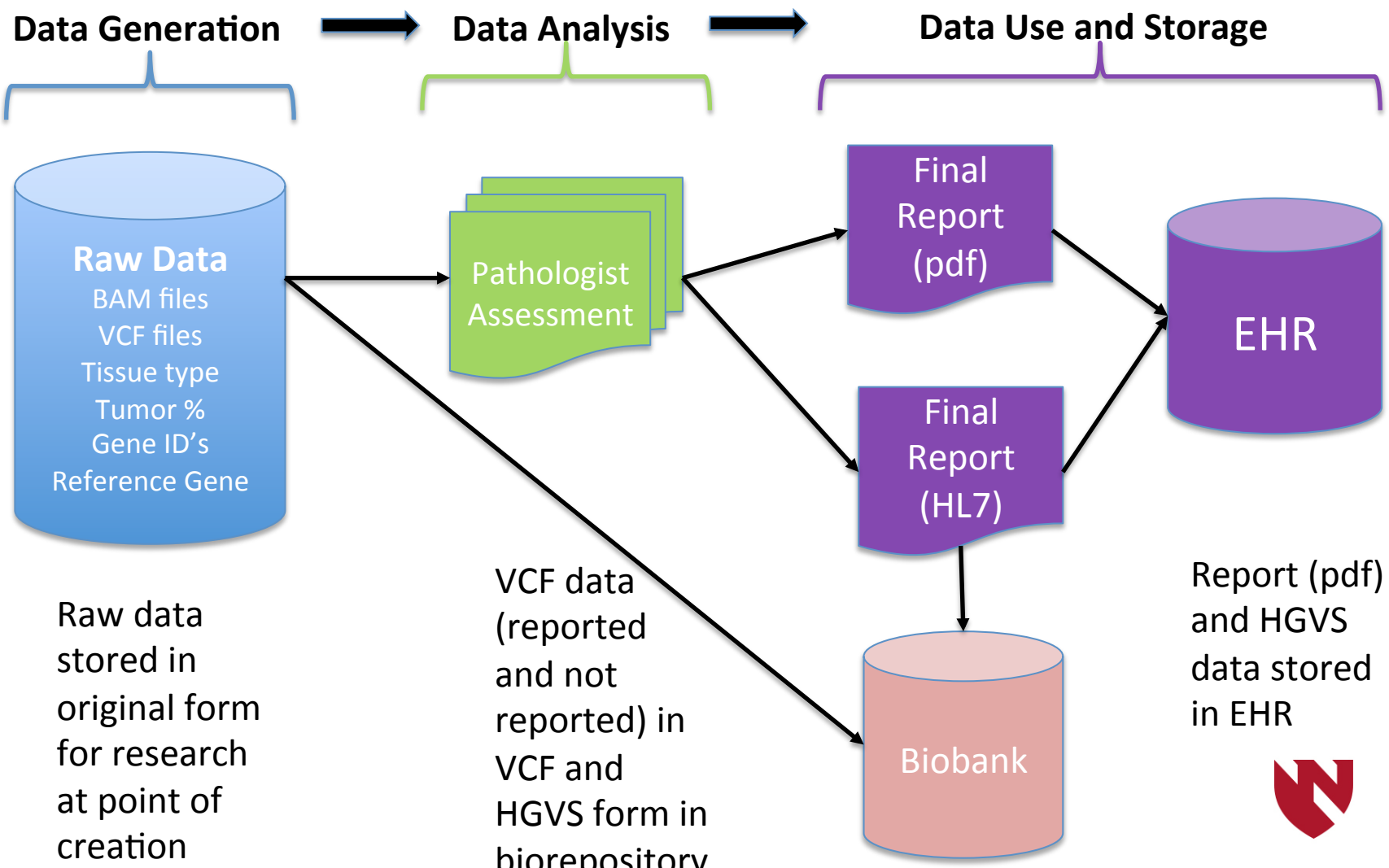
Results:
Performed tests – LOINC (OBR-4)
Results – LOINC (OBX-3) and HGVS (OBX-5)
Abnormal flag – Associate of Molecular Pathologists Tiers 1 - 4

Example:
OBR|1|CE|segment 3|51966-0^Genetic disease DNA analysis panel^LN|….

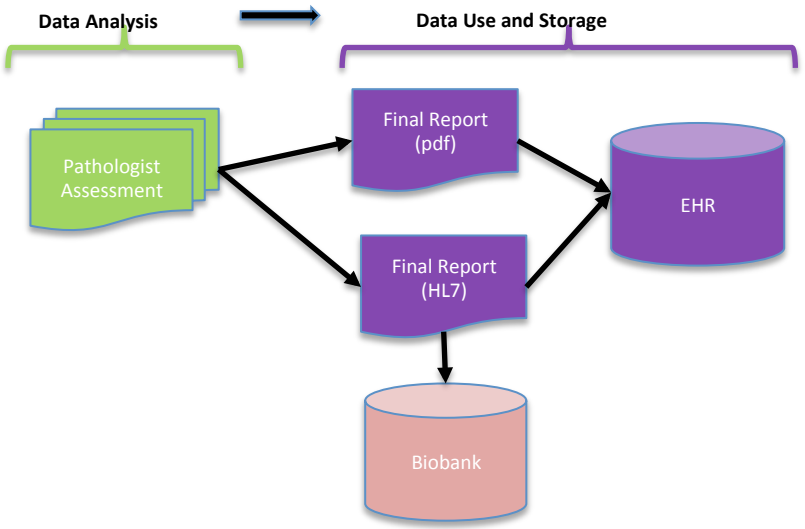
OBX|1|CE|85511-4^BRAF gene mutations found in colorectal cancer specimen by molecular genetics method^LN||*NM_004333.4(BRAF):c.1799T>A (p.Val600Glu)*|…|Pathogenic|

OBX|2|CE|85509-8^KRAS gene mutations found in colorectal cancer specimen by molecular genetics method^LN||*NM_004985.4(KRAS):c.35G>A (p.Gly600Asp)*|…|Pathogenic|

Genomic Data Flow and Store



Genomic Data Flow Example



1. Pathologist sign-out is the trigger event
2. PDF report sent per usual practice
3. HL7 version 2.5.1 message sent to biobank and EHR, simultaneously

HL7 Message Sent

MSH|^~\&|GenomOncology Workbench|UNMC|Mirth|UNMC|||ORU^R01^ORU_R01|77801|P|2.5.1|
PID|1||12345||Doe^Jane^||19850206|F
ORC|1||G17-xxx||CM||^^^^
OBR|1||G17-xxx|55232-3^Genetic analysis summary panel^LN|||
OBX|1|FT|51969-4^Genetic analysis summary report^LN||<p>For this specific specimen there was 200X coverage for the following regions,
therefore low frequency variants in these regions may not be identified: three amplicons of CEBPA exon1, CUX1 exons 1, 19, and 23, and STAG2 exon 7.
Only clinical trials that pertain to genes with identified somatic mutations are reported.
OBR|2||G17-xxx|55207-5^Genetic analysis discrete result panel^LN|||||||||||^Bruce Willis, MD
OBX|1|CWE|911752541000004109^TP53 sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|TP53 NP_000537.3:R175H NM_000546.5:c.524G>A^TP53 R175H|||Pathogenic|||F
OBX|2|CWE|911752871000004102^ASXL1 sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|ASXL1 NP_056153.2:N986S NM_015338.5:c.2957A>G^ASXL1 N986S|||Likely Benign|||F
OBX|3|CWE|911752061000004102^ABL1 sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|ABL1 NM_005157.4:c.(=)|||Normal|||F
OBX|4|CWE|911752881000004104^ATRX sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|ATRX NM_000489.3:c.(=)|||Normal|||F
OBX|5|CWE|911752891000004101^BCOR sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|BCOR NM_001123385.1:c.(=)|||Normal|||F
OBX|6|CWE|911752901000004102^BCORL1 sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|BCORL1 NM_021946.4:c.(=)|||Normal|||F
OBX|7|CWE|911752111000004101^BRAF sequence variant identified in excised malignant neoplasm (observable entity)^SCT|1|BRAF NM_004333.4:c.(=)|||Normal|||F

Patient Care Use

Find all patients with colorectal cancer AND a mutation in the PIK3CA locus and/or the ERBB2 gene locus.

Aspirin therapy shown to be beneficial

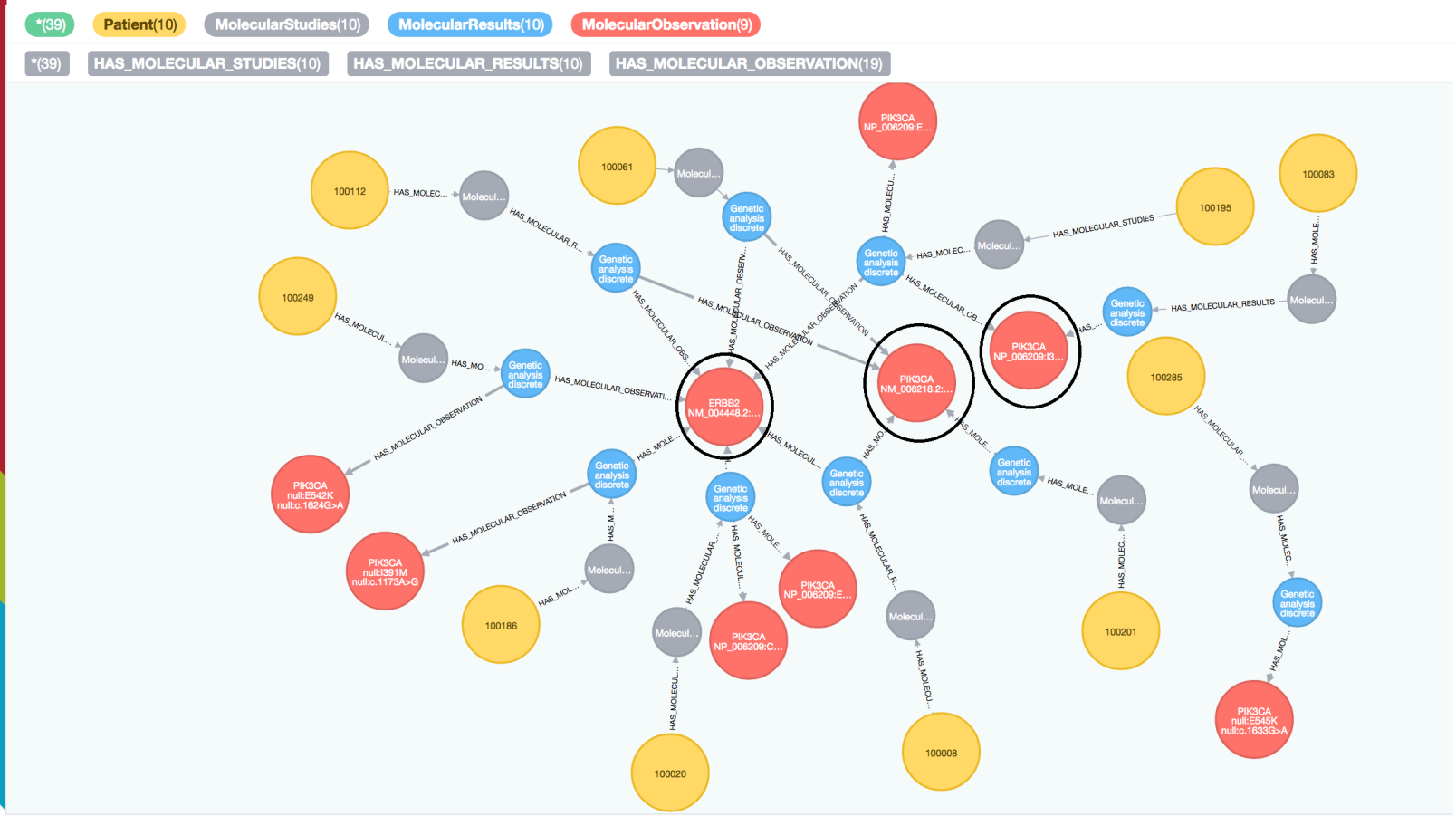
DeidPatId	Observable	HGVS	Pathogenicity
"100020"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NP_006209:C420R NM_006218:c.1258T>C"	"Uncertain Significance"
"100020"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NP_006209:E453del NM_006218:c.1359_1361delAGA"	"Likely Pathogenic"
"100020"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100008"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NM_006218.2:c.(=)"	"Normal"
"100008"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100285"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA null:E545K null:c.1633G>A"	"Pathogenic"
"100195"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NP_006209:I391M NM_006218:c.1173A>G"	"Likely Benign"
"100195"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NP_006209:E109_I112delinsD NM_006218:c.327_335delAGAAAAGAT"	"Likely Pathogenic"
"100195"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100061"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NM_006218.2:c.(=)"	"Normal"
"100061"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100083"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NP_006209:I391M NM_006218:c.1173A>G"	"Likely Benign"
"100112"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100112"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA NM_006218.2:c.(=)"	"Normal"
"100249"	"ERBB2 sequence variant identified in excised malignant neoplasm (observable entity)"	"ERBB2 NM_004448.2:c.(=)"	"Normal"
"100249"	"PIK3CA sequence variant identified in excised malignant neoplasm (observable entity)"	"PIK3CA null:E542K null:c.1624G>A"	"Likely Pathogenic"

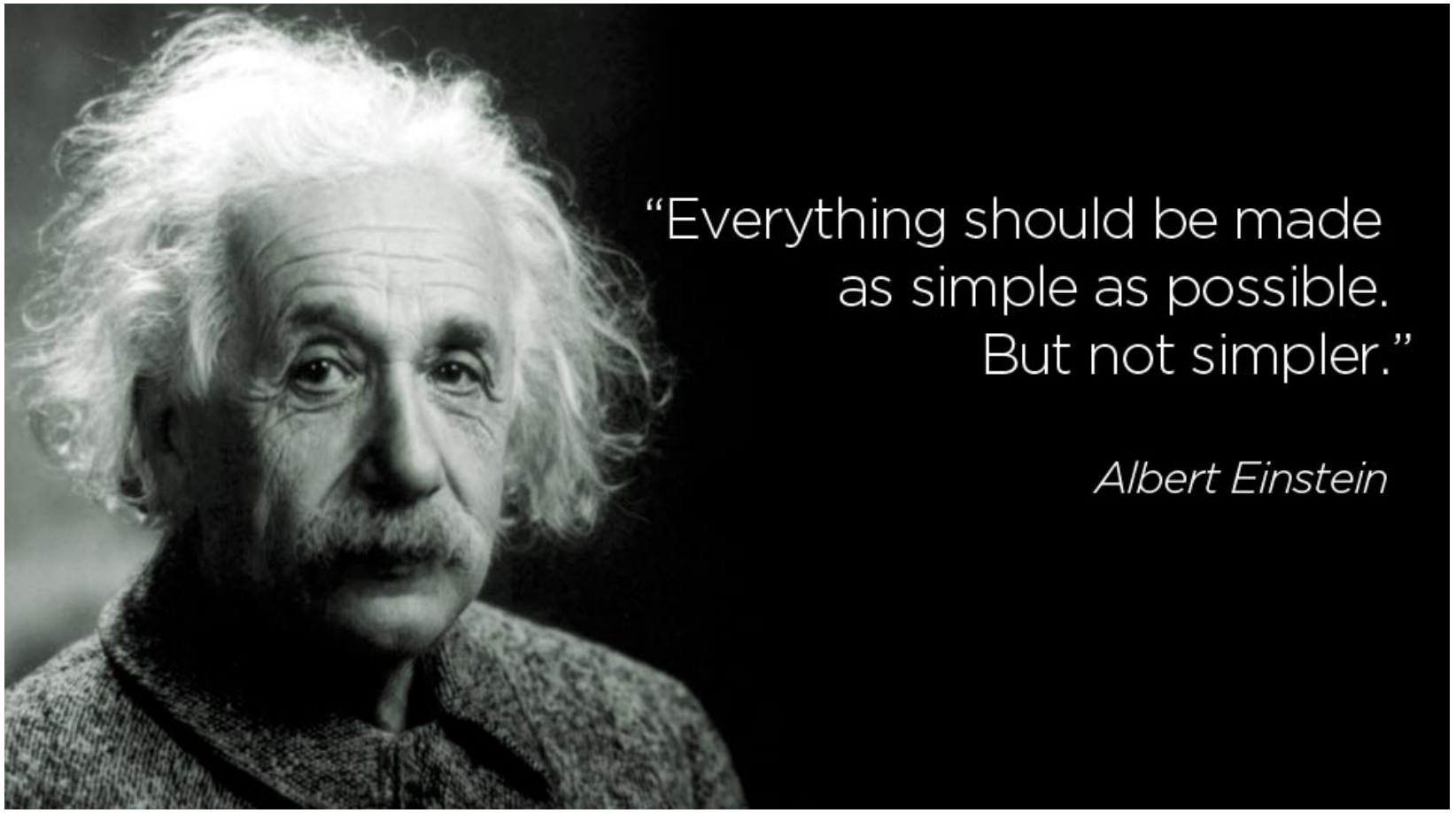
Quality

Frequency of BRAF mutations by disorder

Frequency	HGVS	Observable	Pathogenicity	Diagnosis
99	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Other"
95	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Colorectal Cancer"
59	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Non-Small Cell Lung Cancer"
15	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Tumor of Unknown Origin"
15	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Melanoma"
12	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"GIST"
12	"BRAF null:V600E null:c.1799T>A"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Pathogenic"	"Colorectal Cancer"
8	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	"Breast Cancer"
6	"BRAF null:V600E null:c.1799T>A"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Pathogenic"	"Melanoma"
4	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Normal"	(empty)
3	"BRAF NP_004324:V600E NM_004333:c.1799T>A"	"BRAF sequence variant identified in excised malignant neoplasm (observable entity)"	"Pathogenic"	"Melanoma"
3	"BRAF NM_004333.4:c.(=)"	"BRAF sequence variant identified in excised malignant neoplasm (observable	"Normal"	"Glioma"

Research – Common Variant in Colorectal Cancer Patients




“Everything should be made
as simple as possible.
But not simpler.”
Albert Einstein

Questions

University of Nebraska Medical Center

Nebraska Medicine