Validation of Genomic CDM Extension with Real World Clinical Genomics Data

Kyu-pyo Kim MD, PhD

kkp1122@amc.seoul.kr

Asan Institute for Life Sciences

Dept of Oncology, Asan Medical Center, University of Ulsan College of Medicine, South Korea

Disclosures

- I have no conflict of interests to declare related to the presentation.

Precision Medicine in Population Health Outcomes

- As genomics expands from research to clinical practice, identifying outcomes at the population and each patient level becomes more important.
  - Subpopulation stratification
  - Targeted treatment protocols
  - Predictive analytics

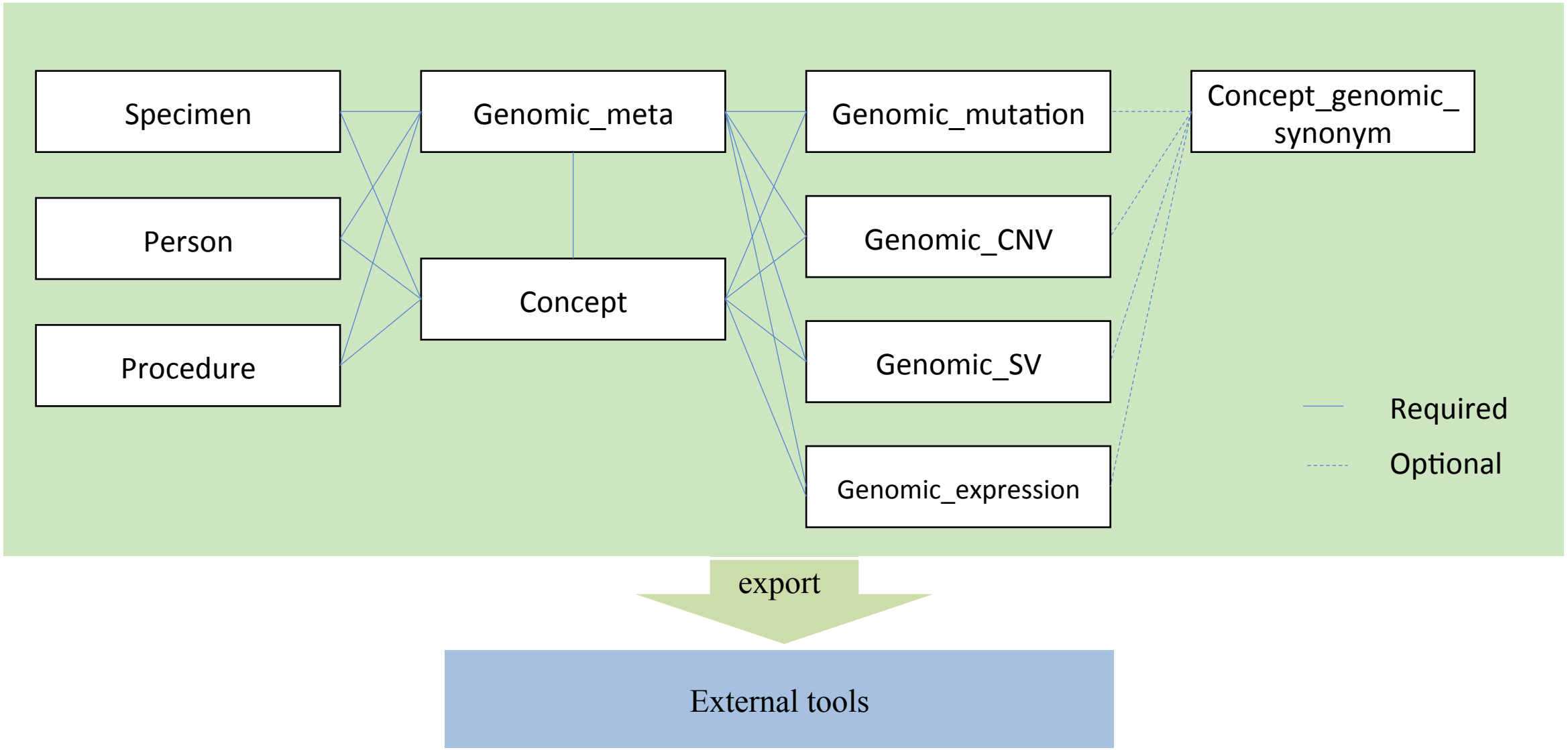
Precision Medicine Enters Real World Data

- Since March 2017, Next Generation Sequencing (NGS) has become nationally reimbursed for cancer and rare diseases in Korea.
- Asan Medical Center treats around 10% of the total Korean cancer patients; over 50 cancer patients are sequenced weekly in Asan Medical Center alone.
- Learning the differences from NGS real world data and research data
  - Heterogeneity in NGS report items: tumor cellularity, panels
  - Trade-off between flexibility and operationality
  - Rich amount of prior Sanger sequence based genetic results

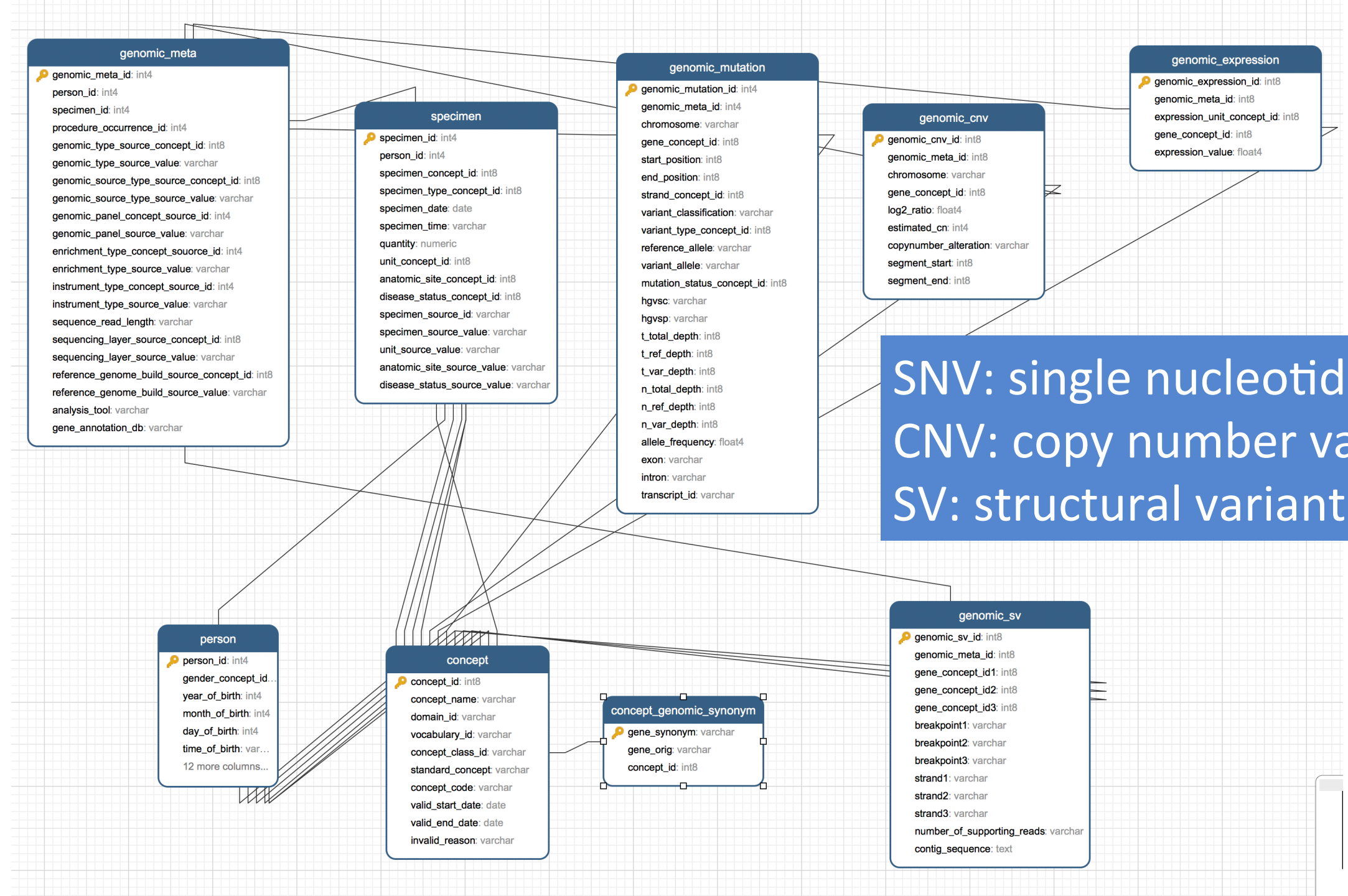
Various Scenarios of NGS

- Source: Somatic vs Germline, Tissue vs Blood, Fresh Frozen vs Formalin Fixed Paraffin Embedded (FFPE)
- Timing: Baseline, Resistance
- Consent: De-identified, Honest broker, Biobank
- Various assays: RNA-seq, micro-array, DNA methylation, protein

→ Everything is centered around the specimen. Tissue is the issue.



* Mutation table includes basic fields such as chr, star, end, ref, alt, gene symbol, depth, HGVSc, HGVSp, strand, exon, intron, transcript id.
** Structural variations, gene fusion and virus integration use same table (Genomic_SV).


genomic_meta
genomic_meta_id: int4
person_id: int4
specimen_id: int4
procedure_occurrence_id: int4
genomic_type_source_concept_id: int8
genomic_type_source_value: varchar
genomic_source_type_source_concept_id: int8
genomic_source_type_source_value: varchar
genomic_panel_concept_source_id: int4
genomic_panel_source_value: varchar
enrichment_type_concept_souorce_id: int4
enrichment_type_source_value: varchar
instrument_type_concept_source_id: int4
instrument_type_source_value: varchar
sequence_read_length: varchar
sequencing_layer_source_concept_id: int8
sequencing_layer_source_value: varchar
reference_genome_build_source_concept_id: int8
reference_genome_build_source_value: varchar
analysis_tool: varchar
gene_annotation_db: varchar
specimen
specimen_id: int4
person_id: int4
specimen_concept_id: int8
specimen_type_concept_id: int8
specimen_date: date
specimen_time: varchar
quantity: numeric
unit_concept_id: int8
anatomic_site_concept_id: int8
disease_status_concept_id: int8
specimen_source_id: varchar
specimen_source_value: varchar
unit_source_value: varchar
anatomic_site_source_value: varchar
disease_status_source_value: varchar
genomic_mutation
genomic_mutation_id: int4
genomic_meta_id: int4
chromosome: varchar
gene_concept_id: int8
start_position: int8
end_position: int8
strand_concept_id: int8
variant_classification: varchar
variant_type_concept_id: int8
reference_allele: varchar
variant_allele: varchar
mutation_status_concept_id: int8
hgvsc: varchar
hgvsp: varchar
t_total_depth: int8
t_ref_depth: int8
t_var_depth: int8
n_total_depth: int8
n_ref_depth: int8
n_var_depth: int8
allele_frequency: float4
exon: varchar
intron: varchar
transcript_id: varchar
genomic_cnv
genomic_cnv_id: int8
genomic_meta_id: int8
chromosome: varchar
gene_concept_id: int8
log2_ratio: float4
estimated_cn: int4
copynumber_alteration: varchar
segment_start: int8
segment_end: int8
genomic_expression
genomic_expression_id: int8
genomic_meta_id: int8
expression_unit_concept_id: int8
gene_concept_id: int8
expression_value: float4
SNV: single nucleotide variant
CNV: copy number variant
SV: structural variant
person
person_id: int4
gender_concept_id...
year_of_birth: int4
month_of_birth: int4
day_of_birth: int4
time_of_birth: var...
12 more columns...
concept
concept_id: int8
concept_name: varchar
domain_id: varchar
vocabulary_id: varchar
concept_class_id: varchar
standard_concept: varchar
concept_code: varchar
valid_start_date: date
valid_end_date: date
invalid_reason: varchar
concept_genomic_synonym
gene_synonym: varchar
gene_orig: varchar
concept_id: int8
genomic_sv
genomic_sv_id: int8
genomic_meta_id: int8
gene_concept_id1: int8
gene_concept_id2: int8
gene_concept_id3: int8
breakpoint1: varchar
breakpoint2: varchar
breakpoint3: varchar
strand1: varchar
strand2: varchar
strand3: varchar
number_of_supporting_reads: varchar
contig_sequence: text

omics_meta (table for meta information)

Column Name	Type		Explanation	Optional (O)
omics_meta_id	int	PK	Each meta_id PK should be identical within a single institute, “institute number” AMC_1, AMC_2	
specimen_id	int	FK (specimen)	Matched with specimen table	
source_type	varchar(32)		The specimen type evaluated. Common values: DNA, RNA, protein	O
assay_type	varchar(64)		Common : WGS, WXS, RNA-seq, Targeted-Seq Allowed : Sanger sequencing, IHC, qPCR, FISH, CISH..	O
panel_name	varchar(64)		(optional, if Targeted-Seq) OP_AMCv3 (panel information on targeted sequencing)	O
enrichment_methods	varchar(64)		(optional, capture, amplification,…) Which method was used for enrichment e.g. WXS, targeted sequencing	O
instrument	varchar(64)		Sequencing instrument (e.g. Illumina, Ion Torrent)	O
center	varchar(64)		Organizations	O
sequence_read_length	varchar(16)		Length of read for seqeuncing reagent	O
sequencing_layout	varchar(16)		single-end, paired-end, mate-pair	O
reference_genome_build	varchar(64)		Reference genome	O
analysis_tool	varchar(64)		Mandatory for NGS-based test) (tool name):(version)	
gene_annotation_db	varchar(64)		Gene db used for gene annotation	
tumor_cellularity	int		Tumor cellularity of specimen	O

mutation information for SNV & INDEL (single nucleotide variant & insertion or deletion)

Column Name	Type		Explanation	Optional (O)
mutation_id	int8	PK		
omics_meta_id	int8	FK (genomic_omics_meta)		
chromosome	varchar(5)		Chromosome number	O
gene_concept_id	int8	FK (concept)	Genomic	O
start_position	int8		Start position of mutations	O
end_position	int8		End position of mutations	O
strand	varchar(2)		Strand	
variant_classification	varchar(50)		Mutation classification	O
variant_type	varchar(50)		Mutation type	O
reference_allele	varchar(50)		Reference allele	O
mutation_status	varchar(10)		Mutation sequence	O
hgvsc	varchar(50)		cDNA change	
hgvsp	varchar(50)		Amino acid change	O
t_total_depth	int8		Total read depth in tumor	
t_ref_depth	int8		Reference allele read depth in tumor	
t_var_depth	int8		Variant allele read depth in tumor	
n_total_depth	int8		Total read depth in normal	
n_ref_depth	int8		Reference allele read depth in normal	
n_var_depth	int8		Variant allele read depth in normal	
allele_frequency	float4		Allele frequency (variant allele/total)	O
exon	varchar(10)		Exon number	O
intron	int8		Intron number	
transcript_id	int4		Transcript id	

genomic_CNV (for copy number variations)

Column Name	Type		Explanation	Optional (O)
cnv_id	int8	PK		
omics_meta_id	int8	FK genomic_omics_meta		
chromosome	varchar(10)		Chromosome number	O
gene_concept_id	int8	FK concept	Gene	O
log2_ratio	float4		log2_ratio	
estimated_cn	int4		copy number	
segment_start	int8		segment start	
segment_end	int8		segment end	

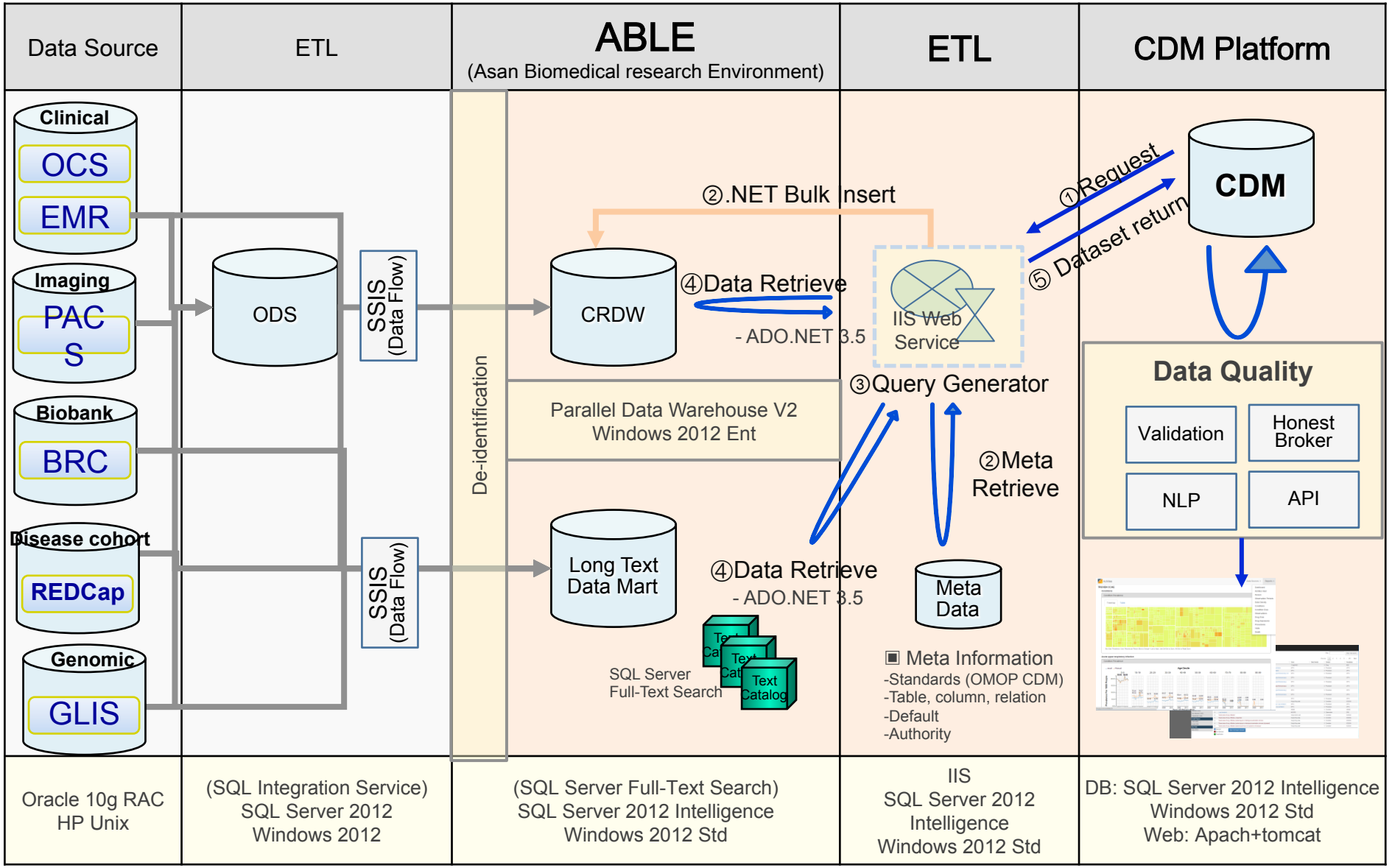
genomic_SV (structural variations, fusions for RNA-seq, virus integration)

Column Name	Type		Explanation	Optional (O)
sv_id	int8	PK		
omics_meta_id	int8	FK genomic_omics_meta		
gene1_concept_id	int8	FK concept	gene1	O
gene2_concept_id	int8	FK concept	gene2	O
gene3_concept_id			Gene3 (not mandatory)	O
breakpoint1	varchar(64)		Breakpoint for gene1	
breakpoint2	varchar(64)		Breakpoint for gene2	
breakpoint3	varchar(64)		Breakpoint for gene3	
strands1	varchar(10)		Strand for gene1	
strands2			Strand for gene2	
strands3			Strand for gene3	
supporting_reads1	varchar(10)		The number of supporting reads for gene1	O
supporting_reads2	varchar(10)		The number of supporting reads for gene2	O
supporting_reads3	varchar(10)		The number of supporting reads for gene3	O
contig_sequence	text		Contig sequence	

expression

Column Name	Type		Explanation	Optional (O)
expression_id	int8	pk		
omics_meta_id	int8	FK genomic_omics_meta		
expression_unit	varchar(20)		Unit for the value	O
gene_concept_id	int4	FK concept	Gene	O
expression_value	float4		Expression value	

Asan Medical Center: Data Flow



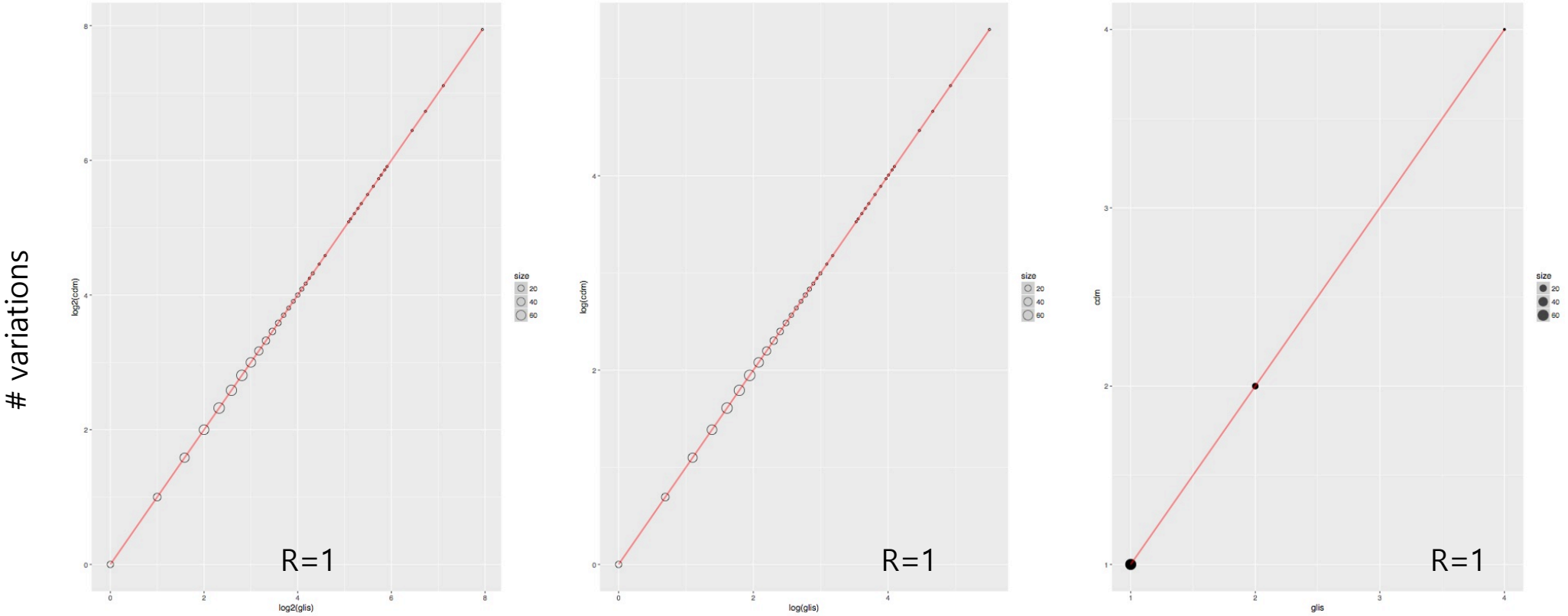
Asan Cancer CDM

- Extracts CDM data for cancer patients
- Linked to cancer biobank (BRC; BioResource Center)
- Extension tables: genomics
- API (application programming interface)
- Precision Medicine Apps

- Further plans: imaging, cancer registries, validating tables for cancer diagnosis and treatment

Results

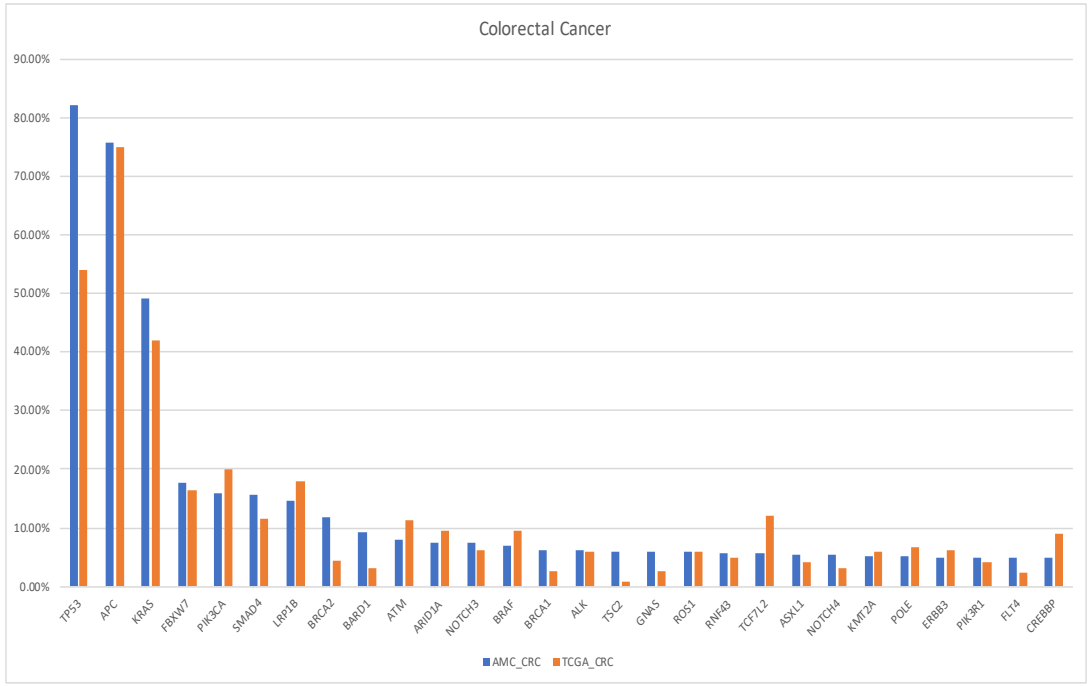
- Internal validation (Clinical Genome DB-GLIS vs CDM)



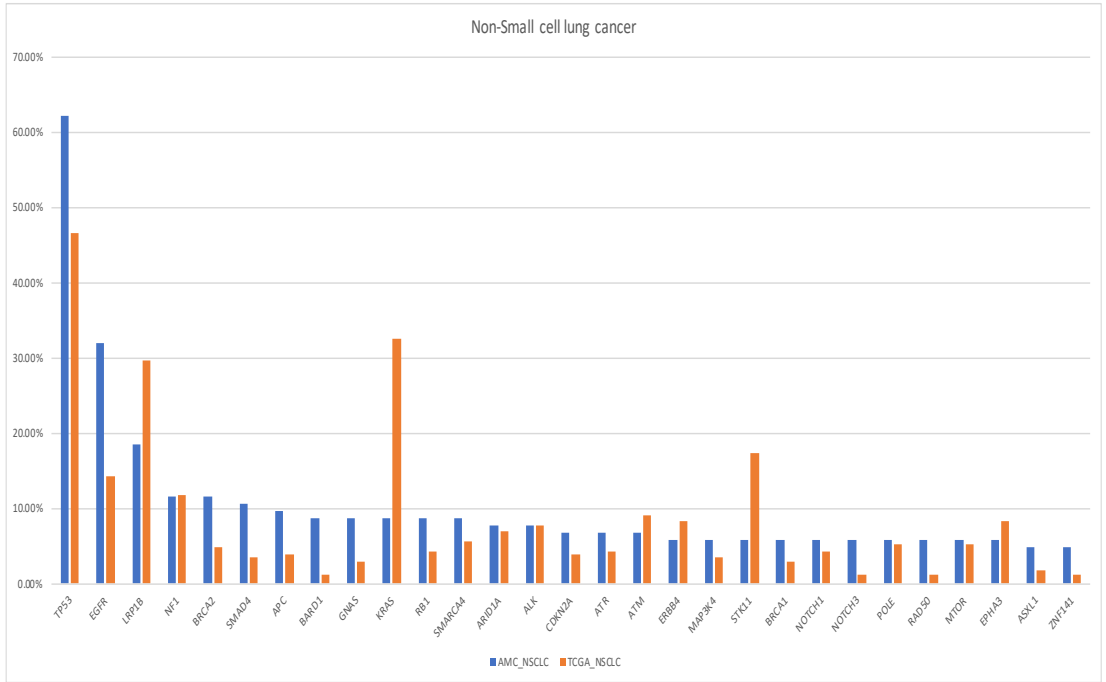
SNV CNV SV

Results

- External validation (Asan vs TCGA; Asan vs Bundang Seoul Nat. Univ)

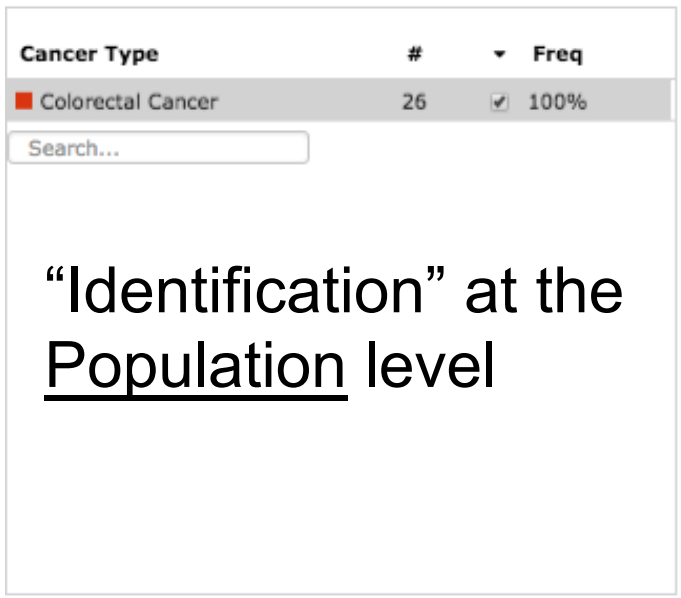


vs TCGA: East Asian NSCLC profile confirmed
vs Bundang: SQL differences and concept ID

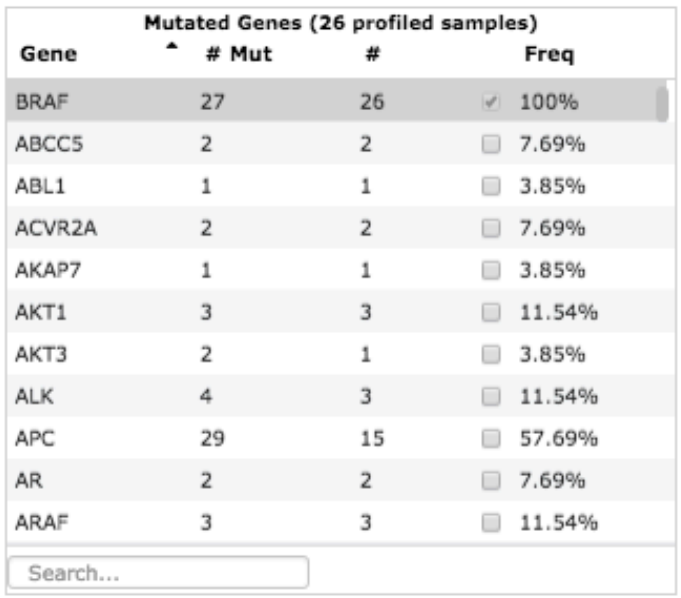


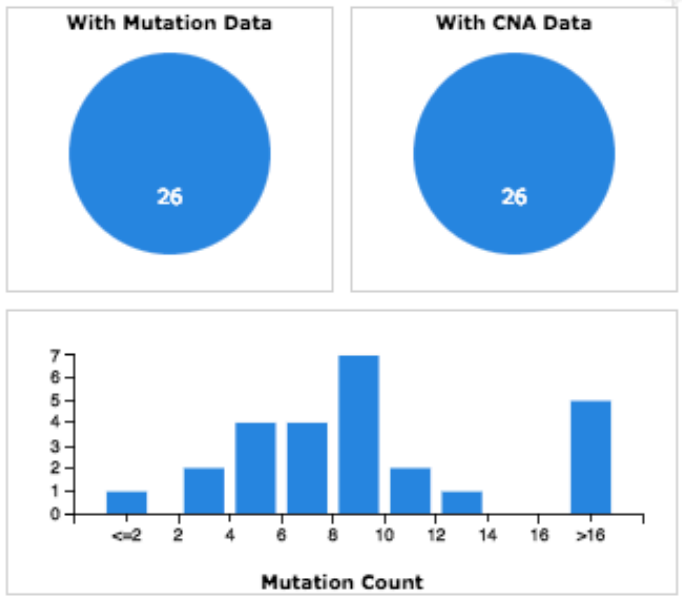
https://tcga-data.nci.nih.gov/docs/publications/luad_2014/
http://tcga-data.nci.nih.gov/docs/publications/coadread_2012/

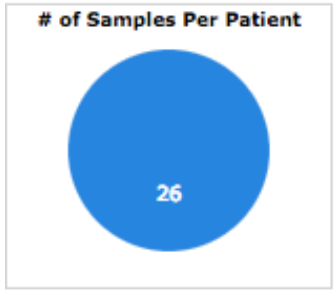
ACCESS: Asan Precision Medicine App

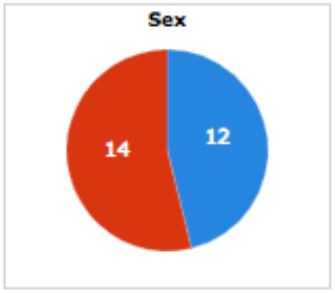


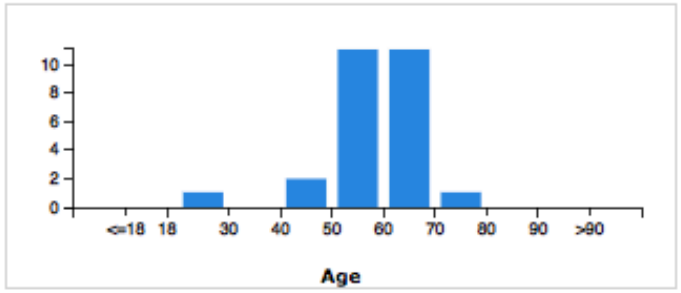
"Identification" at the Population level



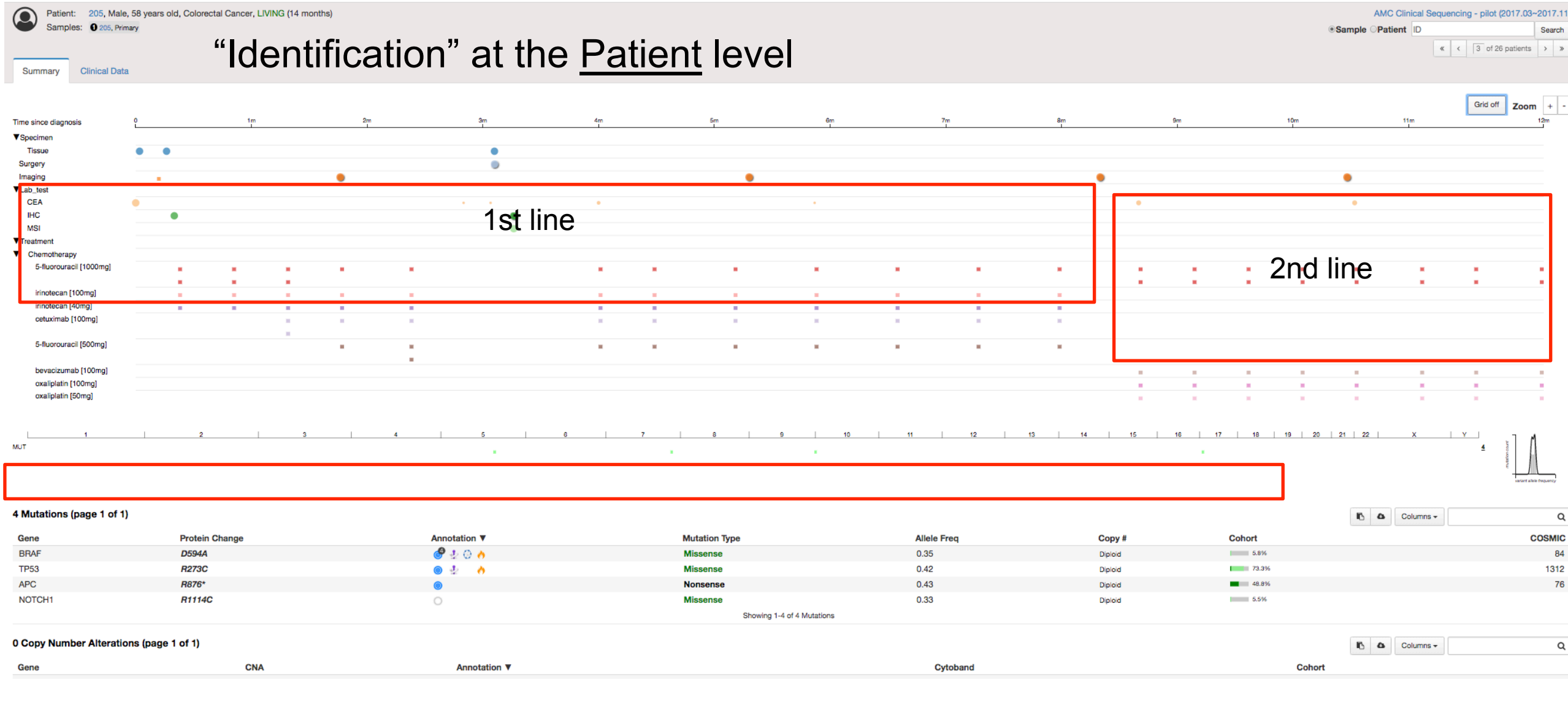




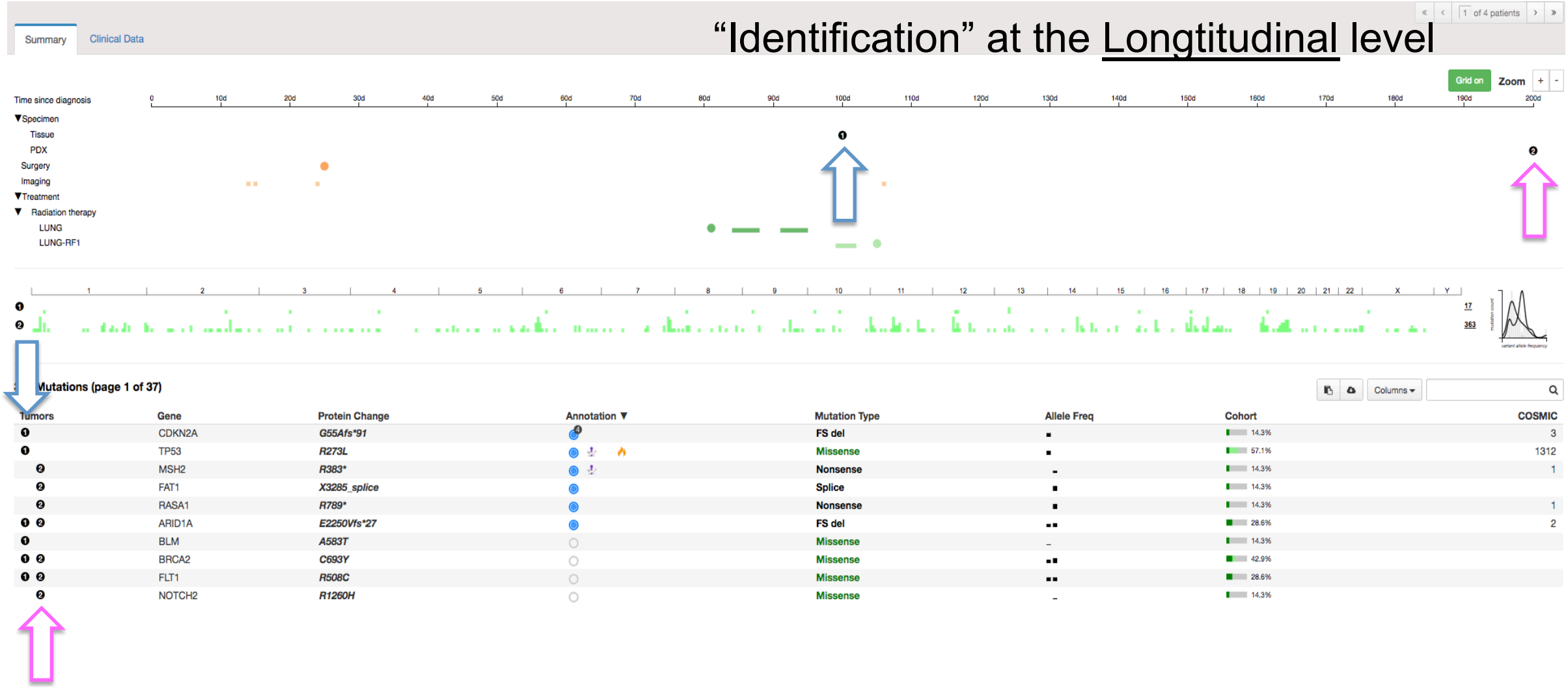




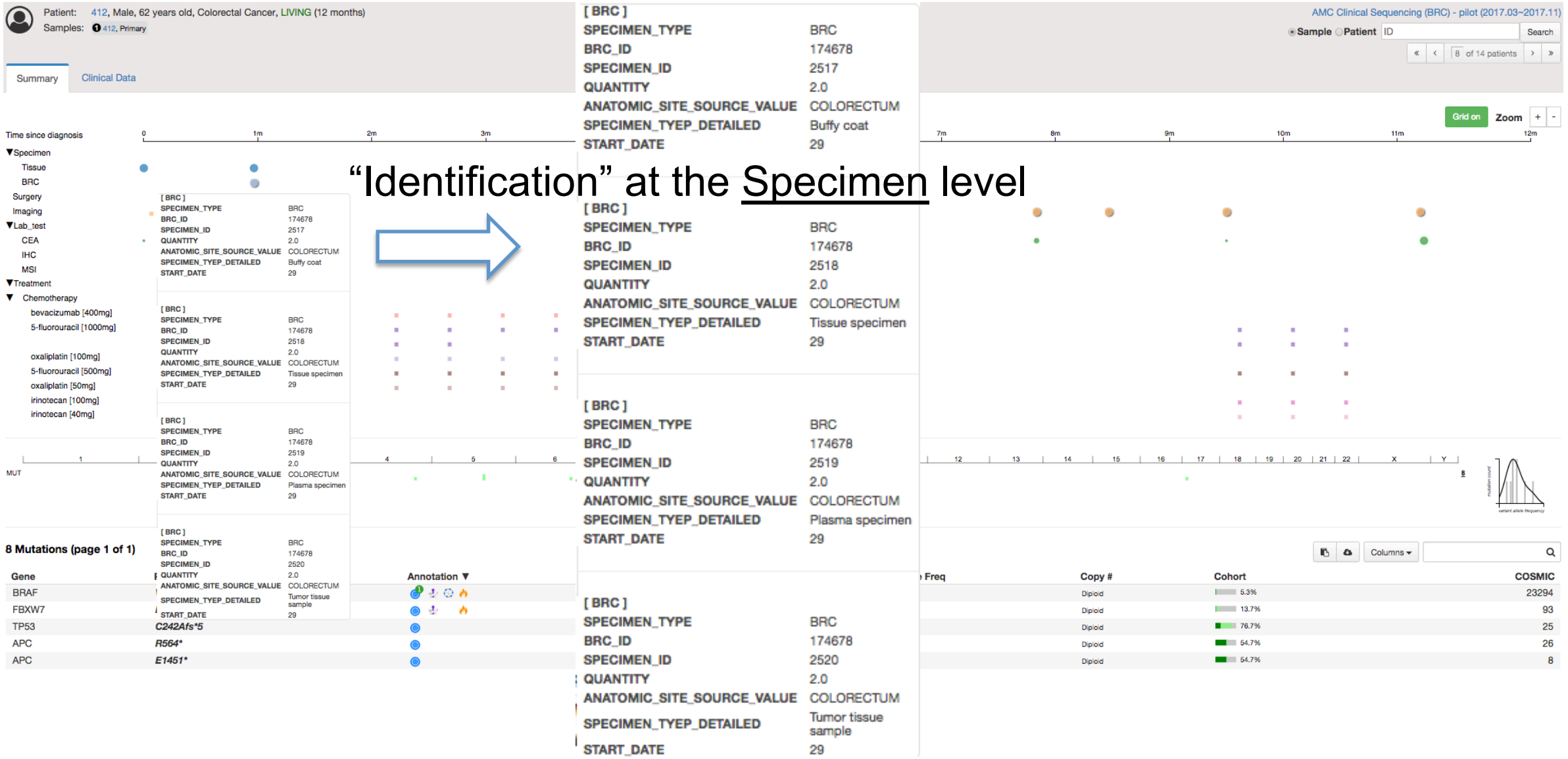
ACCESS: Asan Precision Medicine App



ACCESS: Asan Precision Medicine App



ACCESS: Asan Precision Medicine App



"Identification" at the Specimen level

Discussion

- Translational platforms can allow integration of in-house and central lab results/databases within OMOP CDM.
- Data sharing policy for genomic clinical data have to be explored (e.g. Genomic Data Sharing policies).
- Most of the present NGS data is being generated by cancer, which have issues *per se* in diagnosis and treatment within OMOP CDM.

Acknowledgements

- Asan Medical Center
  - Deokhoon Kim
  - Yoonjung Choi
  - Seulgi Lee
  - Jungeun Hong
  - Ji Seon Oh
  - Jeongsik Oh
  - Eunae Kang
  - Jinsoon Park
  - Jeongeun Lee
  - Tae Won Kim
  - Sejin Jang
  - Young Hak Kim
  - Eun Kyung Choi
  - Chang-Sik Yu
  - Chong Jai Kim
  - Jay Lee
- Seoul National University Bundang Hospital
  - Hee Hwang
  - Sooyoung Yoo
  - Sejoon Lee
- Yonsei University
  - Yu Rang Park