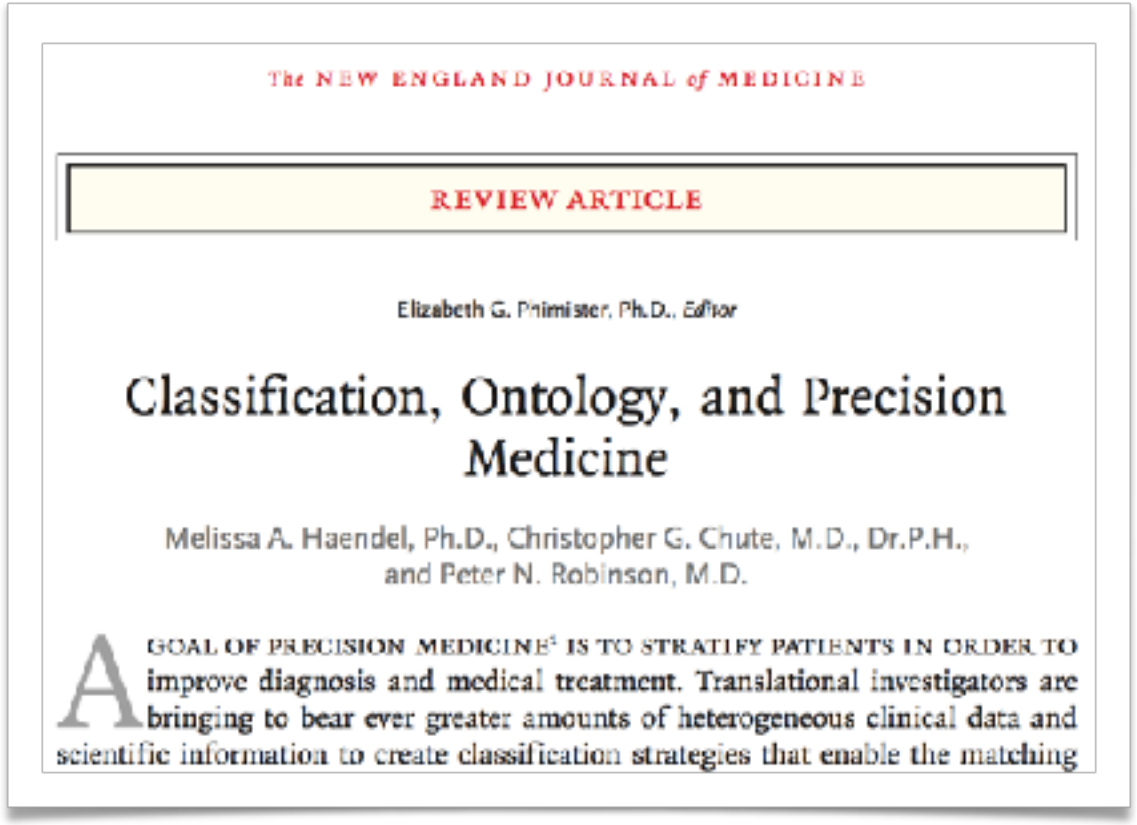
2018 OHDSI Symposium

Genomic Working Group Meeting

# Genomic CDM Extension

Seojeong Shin, Seng Chan You

Ajou University



The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

Elizabeth G. Phimister, Ph.D., Editor

# Classification, Ontology, and Precision Medicine

Melissa A. Haendel, Ph.D., Christopher G. Chute, M.D., Dr.P.H., and Peter N. Robinson, M.D.

A GOAL OF PRECISION MEDICINE[1] IS TO STRATIFY PATIENTS IN ORDER TO improve diagnosis and medical treatment. Translational investigators are bringing to bear ever greater amounts of heterogeneous clinical data and scientific information to create classification strategies that enable the matching

***Data standards*** *can ultimately be reduced to two components:*

*Structure* & *Semantics*.

**N Engl J Med 2018;379:1452-62.**

# Contents

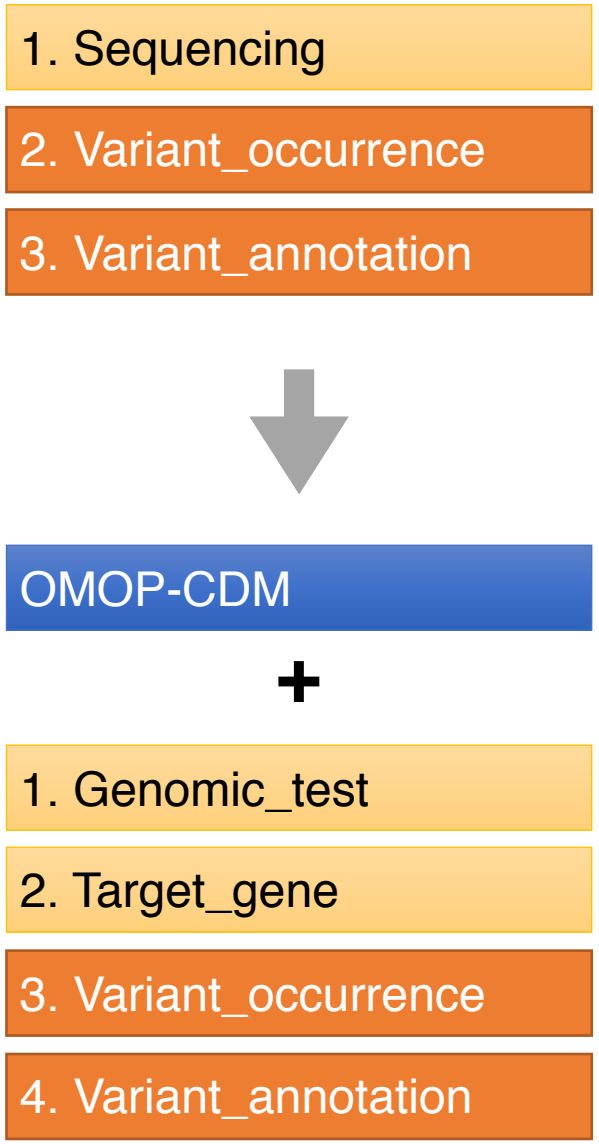

- Structure of Genomic Extension Model
  - Genomic-CDM (G-CDM)
- Vocabulary used in G-CDM
  - Concept ID request

# Contents

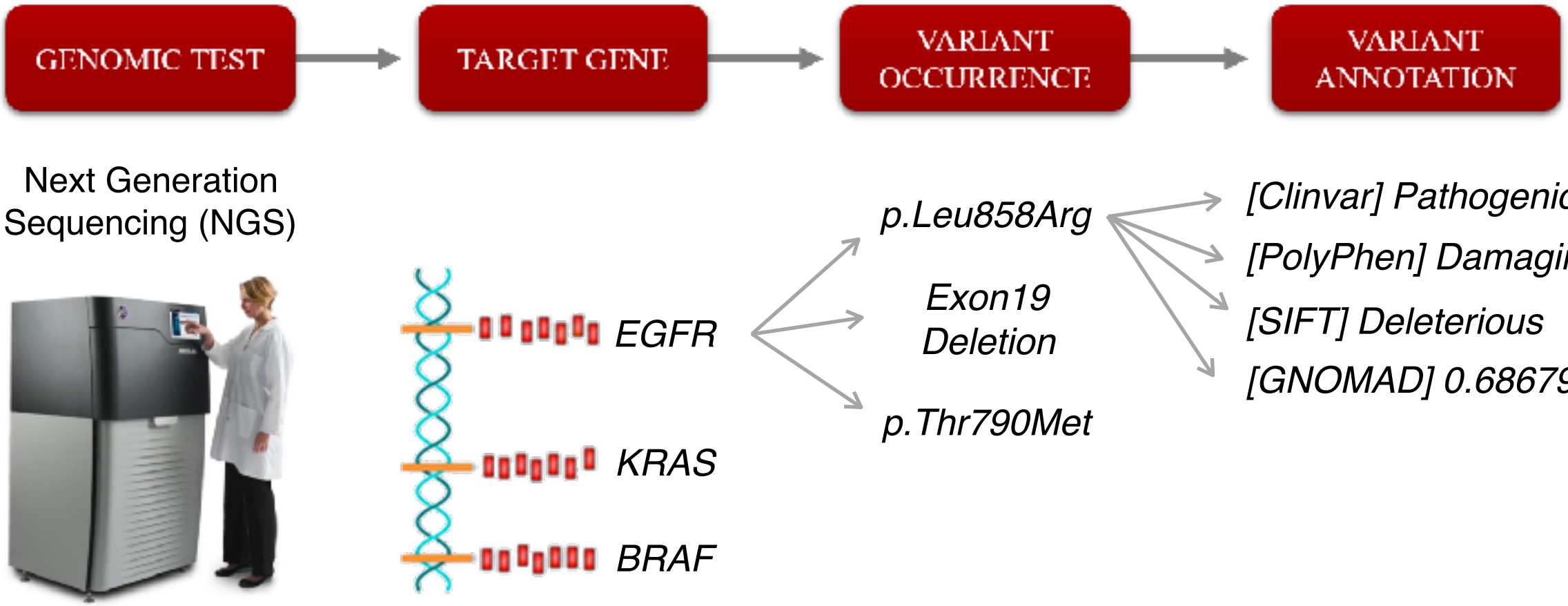

- **Structure of Genomic Extension Model**
  - **Genomic-CDM (G-CDM)**
- Vocabulary used in G-CDM
  - Concept ID request

# G-CDM Structure

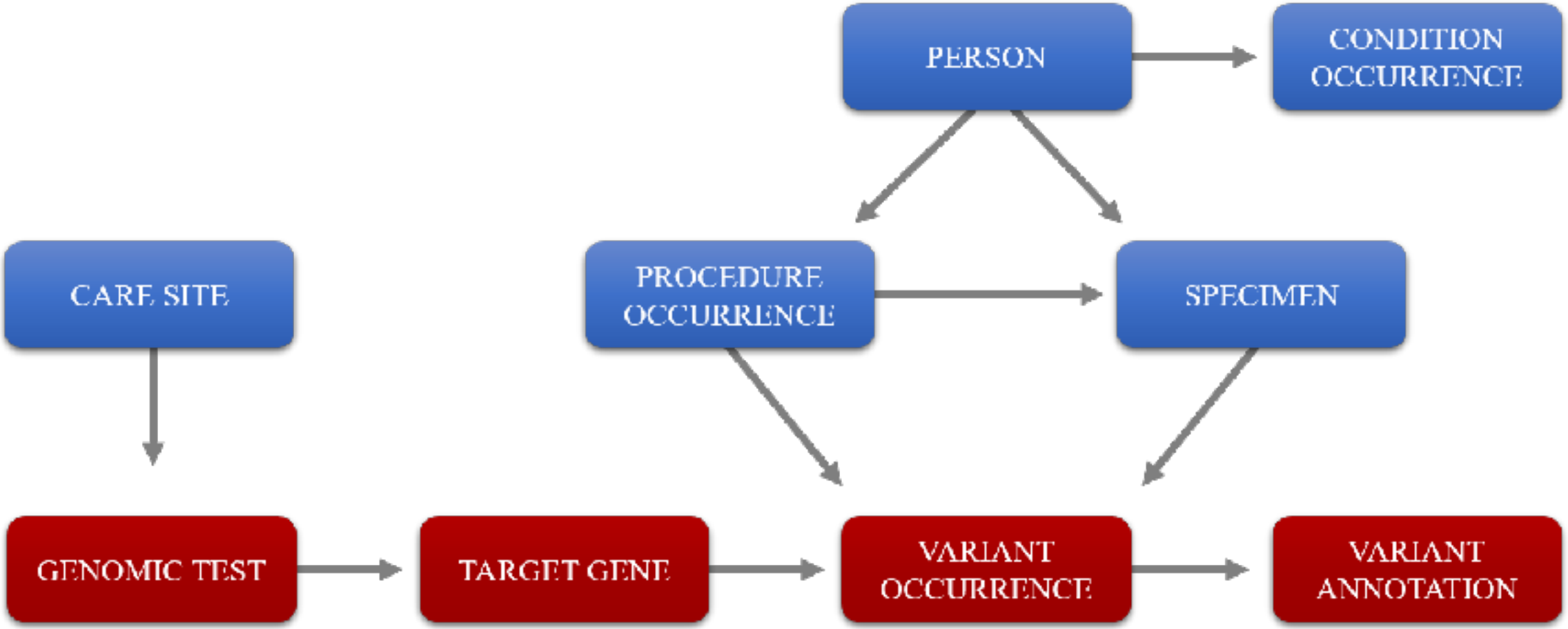
- Beginning version
  - In the OHDSI Symposium in May, 2018
- Upgrade version
  - Take full utilize of the existing OMOP-CDM tables
  - Adapt a standard vocabulary system



# Genomic Test Process



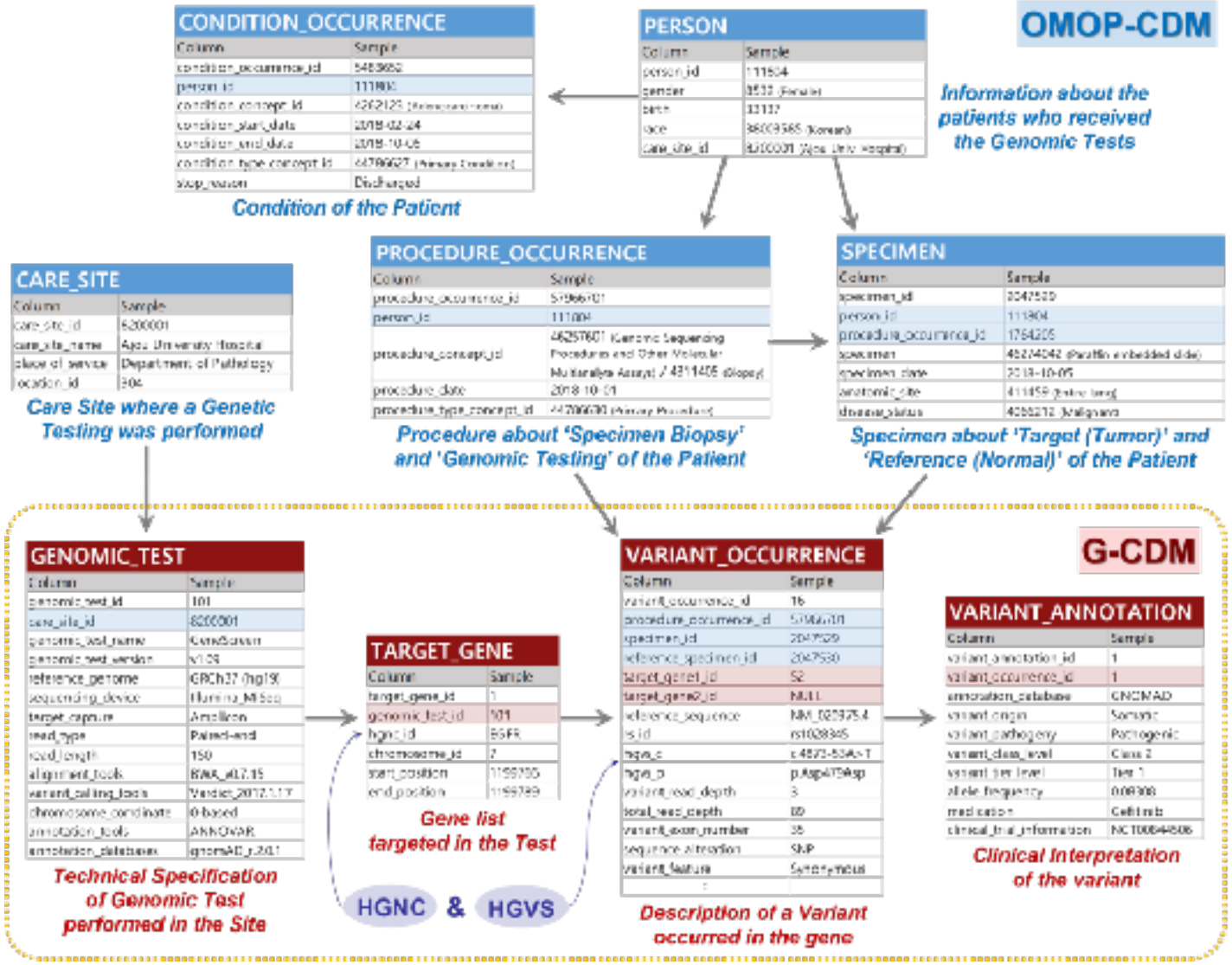
# G-CDM Structure




Schematic diagram of the relationship between the tables that make up the GCDM.

# G-CDM Structure

Entity-relationship diagram (ERD) of G-CDM as an extension to the OMOP-CDM.



# OMOP-CDM

## CONDITION_OCCURRENCE

| Column | Sample |
|---|---|
| condition_occurrence_id | 5483652 |
| person_id | 111804 |
| condition_concept_id | 4262123 (Adenocarcinoma) |
| condition_start_date | 2018-02-24 |
| condition_end_date | 2018-10-05 |
| condition_type_concept_id | 44786627 (Primary Condition) |
| stop_reason | Discharged |

***Condition of the Patient***

## PERSON

| Column | Sample |
|---|---|
| person_id | 111804 |
| gender | 8532 (Female) |
| birth | 33137 |
| race | 38003585 (Korean) |
| care_site_id | 8200001 (Ajou Univ. Hospital) |

***Information about the patients who received the Genomic Tests***

## CARE_SITE

| Column | Sample |
|---|---|
| care_site_id | 8200001 |
| care_site_name | Ajou University Hospital |
| place_of_service | Department of Pathology |
| location_id | 304 |

***Care Site where a Genetic Testing was performed***

## PROCEDURE_OCCURRENCE

| Column | Sample |
|---|---|
| procedure_occurrence_id | 57966701 |
| person_id | 111804 |
| procedure_concept_id | 46257501 (Genomic Sequencing Procedures and Other Molecular Multianalyte Assays) / 4311405 (Biopsy) |
| procedure_date | 2018-10-01 |
| procedure_type_concept_id | 44786630 (Primary Procedure) |

***Procedure about 'Specimen Biopsy' and 'Genomic Testing' of the Patient***

## SPECIMEN

| Column | Sample |
|---|---|
| specimen_id | 2047529 |
| person_id | 111804 |
| procedure_occurrence_id | 1764205 |
| specimen | 46274042 (Paraffin embedded slide) |
| specimen_date | 2018-10-05 |
| anatomic_site | 4111459 (Entire lung) |
| disease_status | 4066212 (Malignant) |

***Specimen about 'Target (Tumor)' and 'Reference (Normal)' of the Patient***

# G-CDM

## GENOMIC_TEST

| Column | Sample |
|---|---|
| genomic_test_id | 101 |
| care_site_id | 8200001 |
| genomic_test_name | GeneScreen |
| genomic_test_version | v1.09 |
| reference_genome | GRCh37 (hg19) |

## TARGET_GENE

| Column | Sample |
|---|---|

## VARIANT_OCCURRENCE

| Column | Sample |
|---|---|
| variant_occurrence_id | 16 |
| procedure_occurrence_id | 57966701 |
| specimen_id | 2047529 |
| reference_specimen_id | 2047530 |
| target_gene1_id | 52 |

## VARIANT_ANNOTATION

| Column | Sample |
|---|---|
| variant_annotation_id | 1 |
| variant_occurrence_id | 1 |

## CARE_SITE

| Column | Sample |
|---|---|
| care_site_id | 8200001 |
| care_site_name | Ajou University Hospitel |
| place_of_service | Department of Pathology |
| location_id | 304 |

***Care Site where a Genetic Testing was performed***

| Column | Sample |
|---|---|
| procedure_occurrence_id | 57966701 |
| person_id | 111804 |
| procedure_concept_id | 46257501 (Genomic Sequencing Procedures and Other Molecular Multianalyte Assays) / 4311405 (Biopsy) |
| procedure_date | 2018-10-01 |
| procedure_type_concept_id | 44786630 (Primary Procedure) |

***Procedure about 'Specimen Biopsy' and 'Genomic Testing' of the Patient***

| Column | Sample |
|---|---|
| specimen_id | 2047529 |
| person_id | 111804 |
| procedure_occurrence_id | 1764205 |
| specimen | 46274042 (Paraffin embedded slide) |
| specimen_date | 2018-10-05 |
| anatomic_site | 411459 (Entire lung) |
| disease_status | 4066212 (Malignant) |

***Specimen about 'Target (Tumor)' and 'Reference (Normal)' of the Patient***

**G-CDM**

## GENOMIC_TEST

| Column | Sample |
|---|---|
| genomic_test_id | 101 |
| care_site_id | 8200001 |
| genomic_test_name | GeneScreen |
| genomic_test_version | v1.09 |
| reference_genome | GRCh37 (hg19) |
| sequencing_device | Illumina_MiSeq |
| target_capture | Amplicon |
| read_type | Paired-end |
| read_length | 150 |
| alignment_tools | BWA_v0.7.15 |
| variant_calling_tools | Vardict_2017.1.17 |
| chromosome_corrdinate | 0-based |
| annotation_tools | ANNOVAR |
| annotation_databases | gnomAD_r.2.0.1 |

***Technical Specification of Genomic Test performed in the Site***

## TARGET_GENE

| Column | Sample |
|---|---|
| target_gene_id | 1 |
| genomic_test_id | 101 |
| hgnc_id | EGFR |
| chromosome_id | 7 |
| start_position | 1199766 |
| end_position | 1199789 |

***Gene list targeted in the Test***

**HGNC & HGVS**

## VARIANT_OCCURRENCE

| Column | Sample |
|---|---|
| variant_occurrence_id | 16 |
| procedure_occurrence_id | 57966701 |
| specimen_id | 2047529 |
| reference_specimen_id | 2047530 |
| target_gene1_id | 52 |
| target_gene2_id | NULL |
| reference_sequence | NM_020975.4 |
| rs_id | rs1028345 |
| hgvs_c | c.4873-53A>T |
| hgvs_p | p.Asp479Asp |
| variant_read_depth | 3 |
| total_read_depth | 89 |
| variant_exon_number | 35 |
| sequence_alteration | SNP |
| variant_feature | Synonymous |
| : | |

***Description of a Variant occurred in the gene***

## VARIANT_ANNOTATION

| Column | Sample |
|---|---|
| variant_annotation_id | 1 |
| variant_occurrence_id | 1 |
| annotation_database | GNOMAD |
| variant_origin | Somatic |
| variant_pathogeny | Pathogenic |
| variant_class_level | Class 2 |
| variant_tier_level | Tier 1 |
| allele_frequency | 0.08308 |
| medication | Gefitinib |
| clinical_trial_information | NCT00844506 |

***Clinical Interpretation of the variant***

# Contents


- Structure of Genomic Extension Model
  - Genomic-CDM (G-CDM)
- **Vocabulary used in G-CDM**
  - **Concept ID request**

*Ontologies are systematic representations of knowledge*

*that can be used to integrate and analyze large amounts of heterogeneous data,*

*allowing precise classification of a patient.*

**N Engl J Med 2018;379:1452-62.**

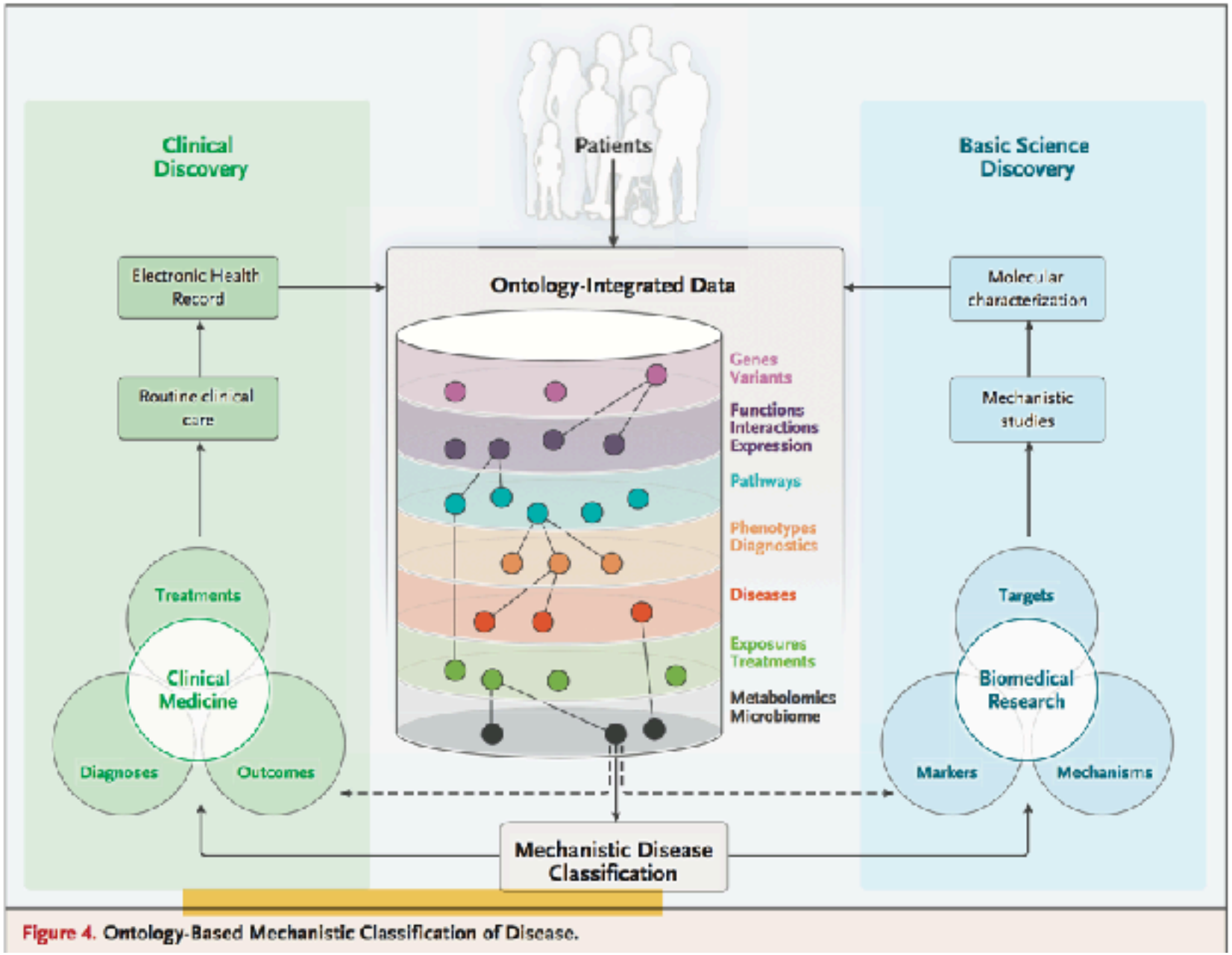



Figure 4. Ontology-Based Mechanistic Classification of Disease.

N Engl J Med 2018;379:1452-62.

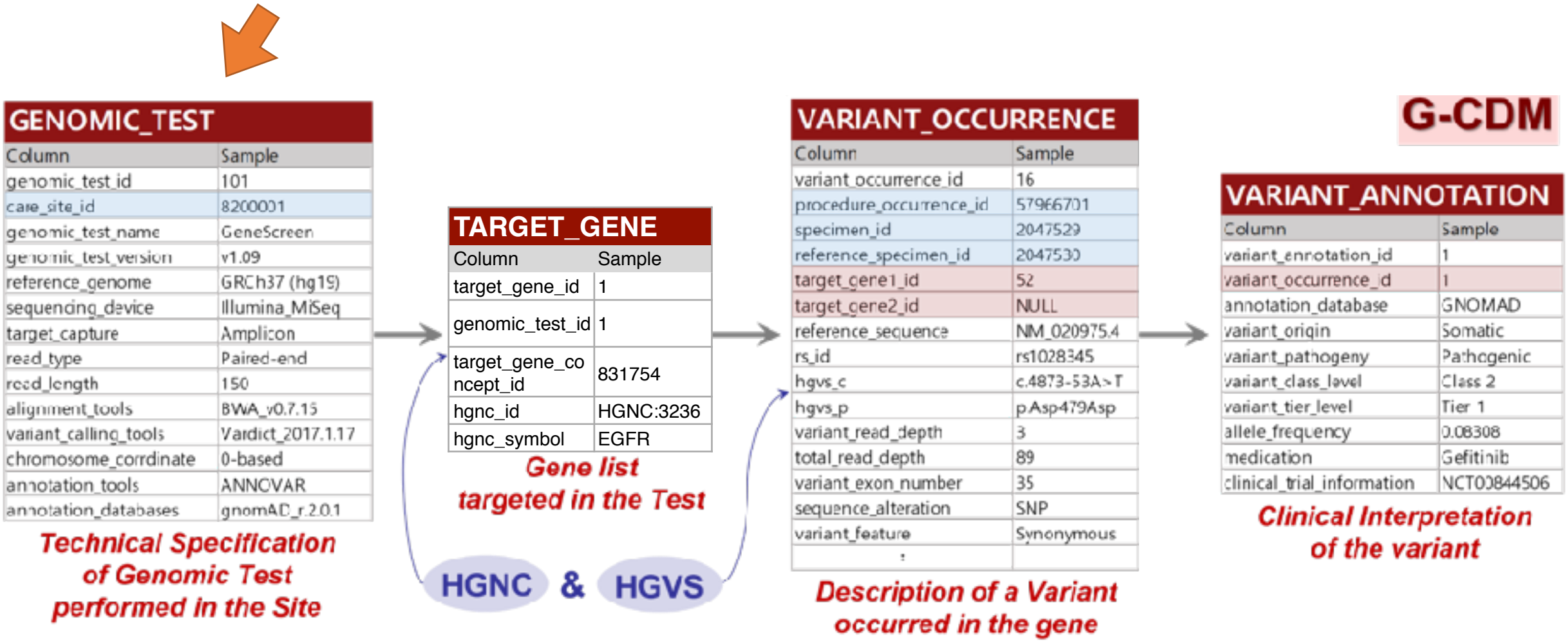
# Concept ID Request

| Table | Column | Concept_name | Number of Concept_id |
|---|---|---|---|
| Target_gene | target_gene_concept_id | Approved Gene Symbols of HGNC Database | 41503 |
| Variant_occurrence | sequence_alteration | MNP | 1 |
| | variant_feature | Upstream, Downstream, Stop-loss, Inframe, 5_prime_UTR, 3_prime_UTR, Intron, Splice_donor, Splice_acceptor | 9 |
| Variant_annotation | annotation_database | Clinvar, PolyPhen, SIFT, SnpEff, 1000G, GNOMAD, ExAC | 7 |
| | variant_pathogeny | Low, Modifier, Moderate, High | 4 |
| | | Benign, Benign/Likely benign, Likely benign, Unknown significance, Likely pathogenic, Likely pathogenic/ Pathogenic, Conflict pathogenic, Pathogenic, Drug response | 9 |
| | | (Benign), Possibly damaging, Probably damaging | 2 |
| | | Tolerated, Tolerated (low confidence), Deleterious, Deleterious (low confidence) | 4 |
| | variant_class_level | Class 1~5 | 5 |
| | variant_tier_level | Tier 1~4 | 4 |
| | | | 41548 |

45

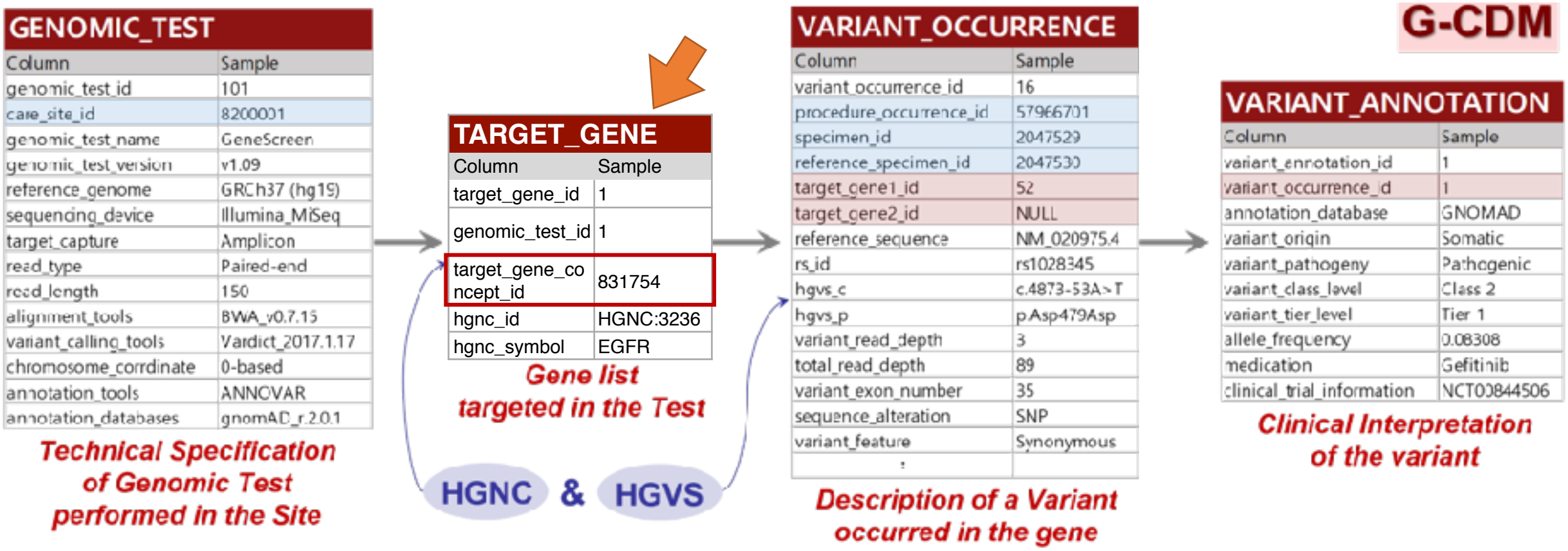
# Concept ID Request

**No** concept ID is needed for '**Genomic_test**' table.



# Concept ID Request

**HGNC** concept ID is needed for '**Target_gene**' table.



ATHENA | SEARCH | DOWNLOAD | LOGIN

Genes

DETAILS

| | |
|---|---|
| Domain ID | Measurement |
| Concept Class ID | LOINC Hierarchy |
| Vocabulary ID | LOINC |
| Concept ID | 40798129 |
| Concept code | LP32747-6 |
| Invalid reason | Valid |
| Standard concept | Classification |
| Synonyms | Genes |
| Valid start | 01/01/1970 |
| Valid end | 12/31/2099 |

TERM CONNECTIONS (626)

| RELATIONSHIP | RELATES TO | CONCEPT ID | VOCABULARY |
|---|---|---|---|
| Is a | Molecular Pathology | 40772980 | LOINC |
| Subsumes | AAHS2 gene | 21497771 | LOINC |
| | ABCA3 gene | 40778070 | LOINC |
| | ABCB1 gene | 40782878 | LOINC |

| | | |
|---|---|---|
| ABCB4 gene | 40782834 | LOINC |
| ABCC8 gene | 40782876 | LOINC |
| AXIN2 gene \| Blood or Tissue | 21497770 | LOINC |
| BBS1 gene | 40782863 | LOINC |
| BBS10 gene | 21497826 | LOINC |
| BBS2 gene | 40794582 | LOINC |
| BCHE gene | 40778611 | LOINC |
| BCKDHB gene | 21196138 | LOINC |
| BCL2 gene | 21496875 | LOINC |
| BCL6 gene | 21496340 | LOINC |
| BCS1L gene | 40777889 | LOINC |
| BLM gene | 40772926 | LOINC |
| BMPR1A gene | 36306253 | LOINC |
| BOR syndrome gene | 40794526 | LOINC |
| BRAF gene | 40782877 | LOINC |

| | | |
|---|---|---|
| TYR gene | 40779618 | LOINC |
| TYRO3P gene | 21496654 | LOINC |
| UBE3A gene | 40782871 | LOINC |
| UGT1A1 gene | 40782850 | LOINC |
| UGT2B15 gene | 40787429 | LOINC |
| UMOD gene/s | 40788810 | LOINC |
| UNC13D gene | 40782877 | LOINC |
| USH1C gene | 21496003 | LOINC |
| USH2A gene | 40784887 | LOINC |
| VAPB gene | 21496004 | LOINC |
| VHL gene | 40779805 | LOINC |
| VIM gene methylation \| Stool | 40778812 | LOINC |
| VKORC1 gene | 40774819 | LOINC |
| WRS1B gene | 40787878 | LOINC |
| WWF gene | 40779575 | LOINC |
| WAS gene | 40780384 | LOINC |
| WFS1 gene | 40782858 | LOINC |
| WS2A gene | 40775117 | LOINC |
| YY1 gene | 40774887 | LOINC |

Not Enough to cover all genes targeted by genomic test

# **HGNC** 41,503 approved genes (ID and Symbol) —> Concept ID

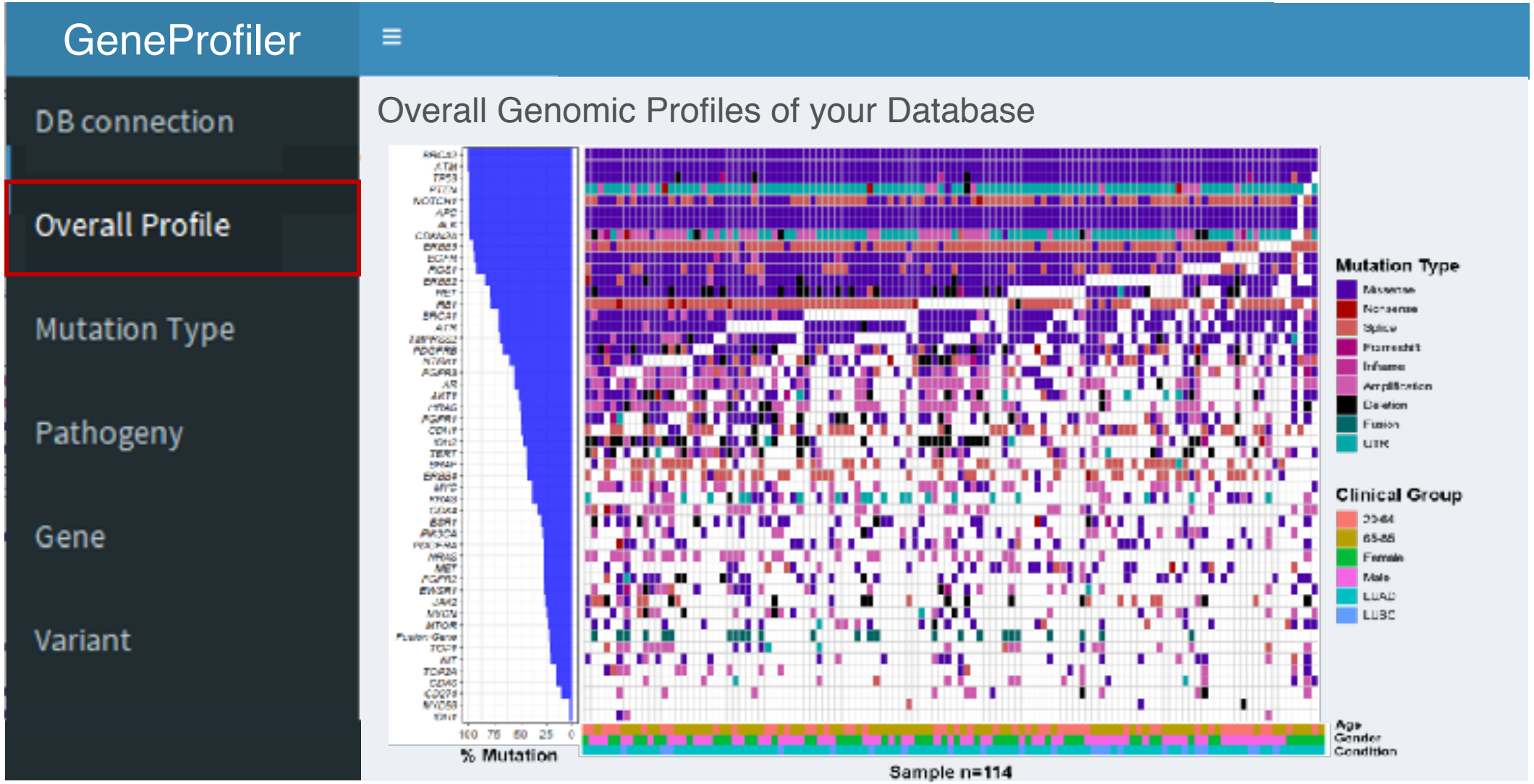

| Locus Group | Total by Locus Group |
|---|---|
| protein-coding gene | 19197 |
| non-coding RNA | 7373 |
| phenotype | 571 |
| pseudogene | 13188 |
| other | 1174 |
| **Total Approved Symbols** | 41503 |

| hgnc_id | symbol | name | locus_group | locus_type | status | location | location_so | alias_symbol | alias_name | prev_symbol | prev_name |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HGNC:5 | A1BG | alpha-1-B glycoprotein | protein-coding gene | gene with protein produ | Approved | 19q13.43 | 19q13.43 | | | | |
| HGNC:37133 | A1BG-AS1 | A1BG antisense RNA 1 | non-coding RNA | RNA, long non-coding | Approved | 19q13.43 | 19q13.43 | FLJ23569 | | NCRNA00181\|A1 | non-protein co |
| HGNC:24086 | A1CF | APOBEC1 complementatio | protein-coding gene | gene with protein produ | Approved | 10q11.23 | 10q11.23 | ACF\|ASP\|ACF64\|ACF65\|A | | APOBEC1CF | |
| HGNC:7 | A2M | alpha-2-macroglobulin | protein-coding gene | gene with protein produ | Approved | 12p13.31 | 12p13.31 | FWP007\|S863-7\|CPAMD5 | | | |
| HGNC:27057 | A2M-AS1 | A2M antisense RNA 1 | non-coding RNA | RNA, long non-coding | Approved | 12p13.31 | 12p13.31 | | | | A2M antisense |
| HGNC:23336 | A2ML1 | alpha-2-macroglobulin like | protein-coding gene | gene with protein produ | Approved | 12p13.31 | 12p13.31 | FLJ25179\|p170 | | CPAMD9 | C3 and PZP-l |
| HGNC:41022 | A2ML1-AS1 | A2ML1 antisense RNA 1 | non-coding RNA | RNA, long non-coding | Approved | 12p13.31 | 12p13.31 | | | | A2ML1 antisen |
| HGNC:41523 | A2ML1-AS2 | A2ML1 antisense RNA 2 | non-coding RNA | RNA, long non-coding | Approved | 12p13.31 | 12p13.31 | | | | A2ML1 antisen |
| HGNC:8 | A2MP1 | alpha-2-macroglobulin pse | pseudogene | pseudogene | Approved | 12p13.31 | 12p13.31 | | | A2MP | alpha-2-macro |
| HGNC:30005 | A3GALT2 | alpha 1,3-galactosyltransfe | protein-coding gene | gene with protein produ | Approved | 1p35.1 | 01p35.1 | IGBS3S\|IGB3 | iGb3 synthas | A3GALT2P | alpha 1,3-gala |
| HGNC:18149 | A4GALT | alpha 1,4-galactosyltransfe | protein-coding gene | gene with protein produ | Approved | 22q13.2 | 22q13.2 | A14GALT\|Gb | Gb3 synthas | P1 | alpha 1,4-gala |
| HGNC:17968 | A4GNT | alpha-1,4-N-acetylglucosan | protein-coding gene | gene with protein produ | Approved | 3q22.3 | 03q22.3 | alpha4GnT | | | |
| HGNC:13666 | AAAS | aladin WD repeat nucleopo | protein-coding gene | gene with protein produ | Approved | 12q13.13 | 12q13.13 | | aladin\|Allgrove, triple-A\|adraca | | achalasia, adi |
| HGNC:21298 | AACS | acetoacetyl-CoA synthetase | protein-coding gene | gene with protein produ | Approved | 12q24.31 | 12q24.31 | FLJ12389\|SU | acyl-CoA synthetase family member 1 | | |
| HGNC:18226 | AACSP1 | acetoacetyl-CoA synthetase | pseudogene | pseudogene | Approved | 5q35.3 | 05q35.3 | | | AACSL | acetoacetyl-Co |
| HGNC:17 | AADAC | arylacetamide deacetylase | protein-coding gene | gene with protein produ | Approved | 3q25.1 | 03q25.1 | DAC\|CES5A1 | | | arylacetamide |
| HGNC:24427 | AADACL2 | arylacetamide deacetylase | protein-coding gene | gene with protein produ | Approved | 3q25.1 | 03q25.1 | MGC72001 | | | |
| HGNC:50301 | AADACL2-AS1 | AADACL2 antisense RNA 1 | non-coding RNA | RNA, long non-coding | Approved | 3q25.1 | 03q25.1 | | | | |
| HGNC:32037 | AADACL3 | arylacetamide deacetylase | protein-coding gene | gene with protein produ | Approved | 1p36.21 | 01p36.21 | OTTHUMG00000001887 | | | |
| HGNC:32038 | AADACL4 | arylacetamide deacetylase | protein-coding gene | gene with protein produ | Approved | 1p36.21 | 01p36.21 | OTTHUMG00000001889 | | | |
| HGNC:50305 | AADACP1 | arylacetamide deacetylase | pseudogene | pseudogene | Approved | 3q25.1 | 03q25.1 | | | | |
| HGNC:17929 | AADAT | aminoadipate aminotransfe | protein-coding gene | gene with protein produ | Approved | 4q33 | 04q33 | KAT\|\|KAT2\|K | kynurenine aminotransferase \|\|\|L | kynurenine/ | |
| HGNC:25662 | AAGAB | alpha and gamma adaptin b | protein-coding gene | gene with protein produ | Approved | 15q23 | 15q23 | FLJ11506\|p34 | | | |
| HGNC:19679 | AAK1 | AP2 associated kinase 1 | protein-coding gene | gene with protein produ | Approved | 2p13.3 | 02p13.3 | KIAA1048\|DKFZp686K16132 | | | |
| HGNC:30205 | AAMDC | adipogenesis associated M | protein-coding gene | gene with protein produ | Approved | 11q14.1 | 11q14.1 | PTD015\|FLJ21035\|CK067 | | C11orf67 | chromosome |
| HGNC:18 | AAMP | angio associated migratory | protein-coding gene | gene with protein produ | Approved | 2q35 | 02q35 | | | | |
| HGNC:19 | AANAT | aralkylamine N-acetyltransf | protein-coding gene | gene with protein produ | Approved | 17q25.1 | 17q25.1 | SNAT | serotonin N-acetyltransferase | | arylalkylamine |
| HGNC:15886 | AAR2 | AAR2 splicing factor homol | protein-coding gene | gene with protein produ | Approved | 20q11.23 | 20q11.23 | bA234K24.2 | | C20orf4 | chromosome 2 |
| HGNC:33842 | AARD | alanine and arginine rich do | protein-coding gene | gene with protein produ | Approved | 8q24.11 | 08q24.11 | LOC441376 | Alanine and | C8orf85 | chromosome 8 |
| HGNC:20 | AARS | alanyl-tRNA synthetase | protein-coding gene | gene with protein produ | Approved | 16q22.1 | 16q22.1 | CMT2N\|AlaR | alanine tRNA ligase 1, cytoplasmic | | |
| HGNC:21022 | AARS2 | alanyl-tRNA synthetase 2, r | protein-coding gene | gene with protein produ | Approved | 6p21.1 | 06p21.1 | KIAA1270\|bA | alanine tRNA | AARSL | alanyl-tRNA s |
| HGNC:28417 | AARSD1 | alanyl-tRNA synthetase dor | protein-coding gene | gene with protein produ | Approved | 17q21.31 | 17q21.31 | MGC2744 | | | |

# Use-case of the Concept ID

http://github.com/OHDSI/StudyProtocolSandbox/tree/master/GeneProfiler

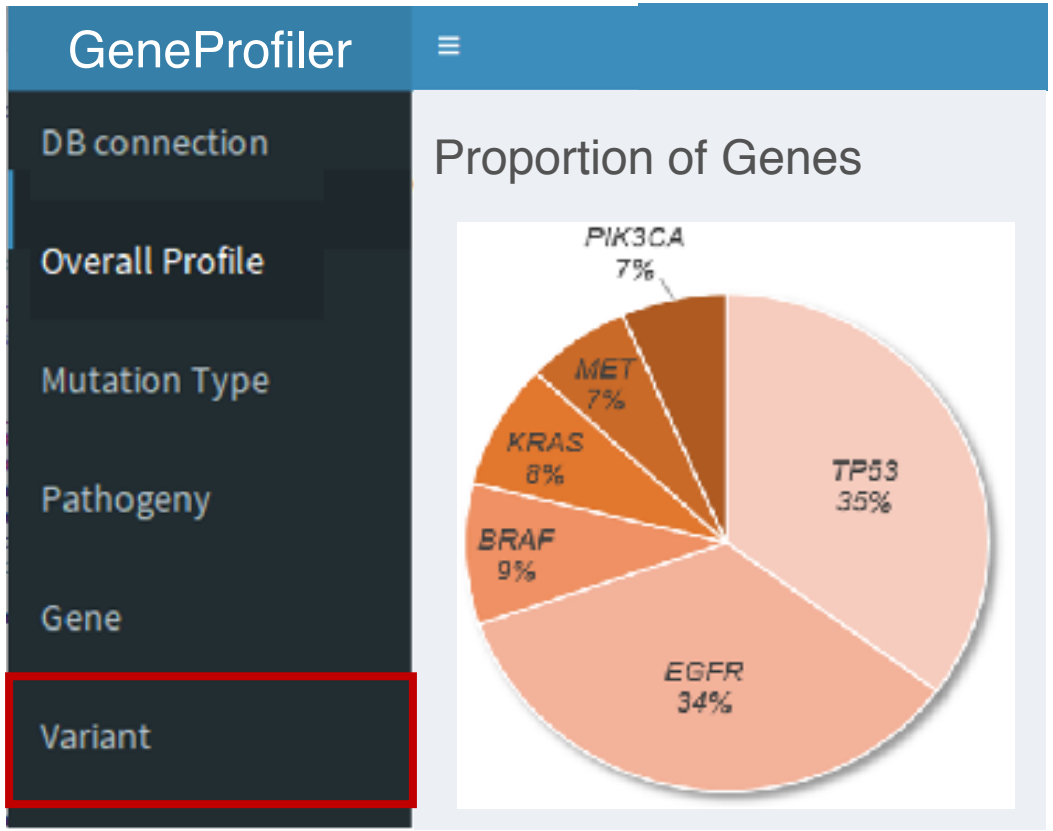
Data Profiling Tool for Genomic Data



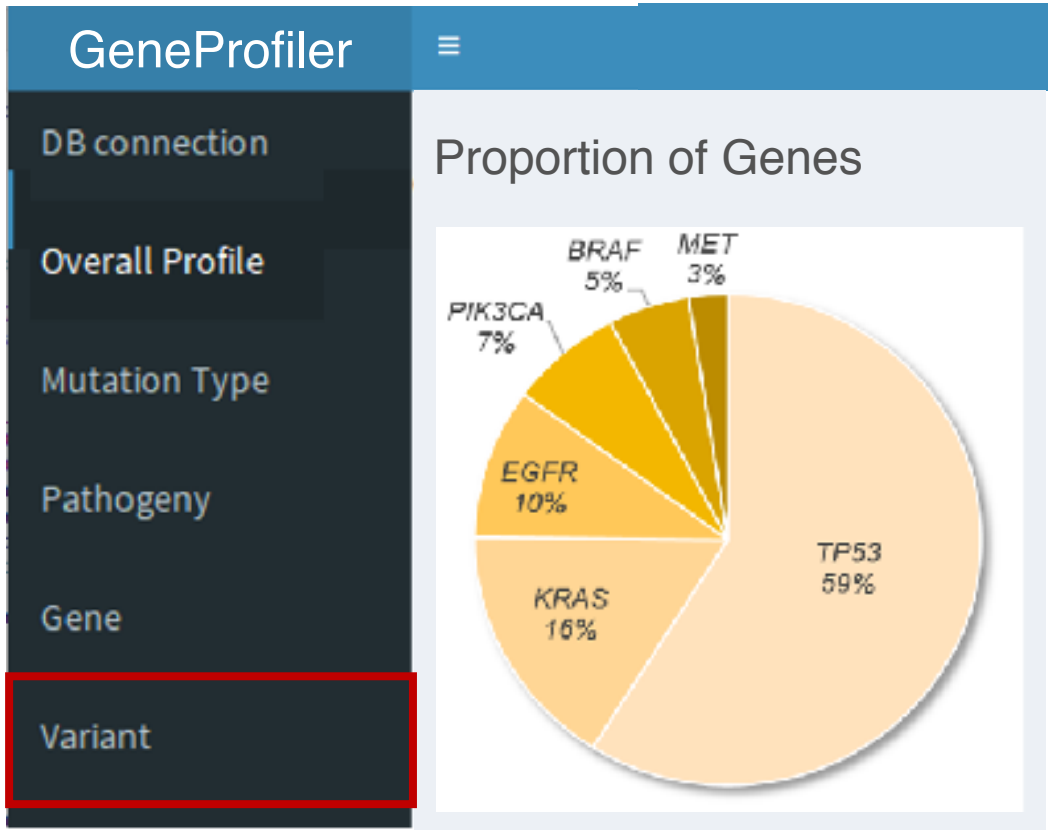
# Use-case of the Concept ID

http://github.com/OHDSI/StudyProtocolSandbox/tree/master/GeneProfiler

Data Profiling Tool for Genomic Data



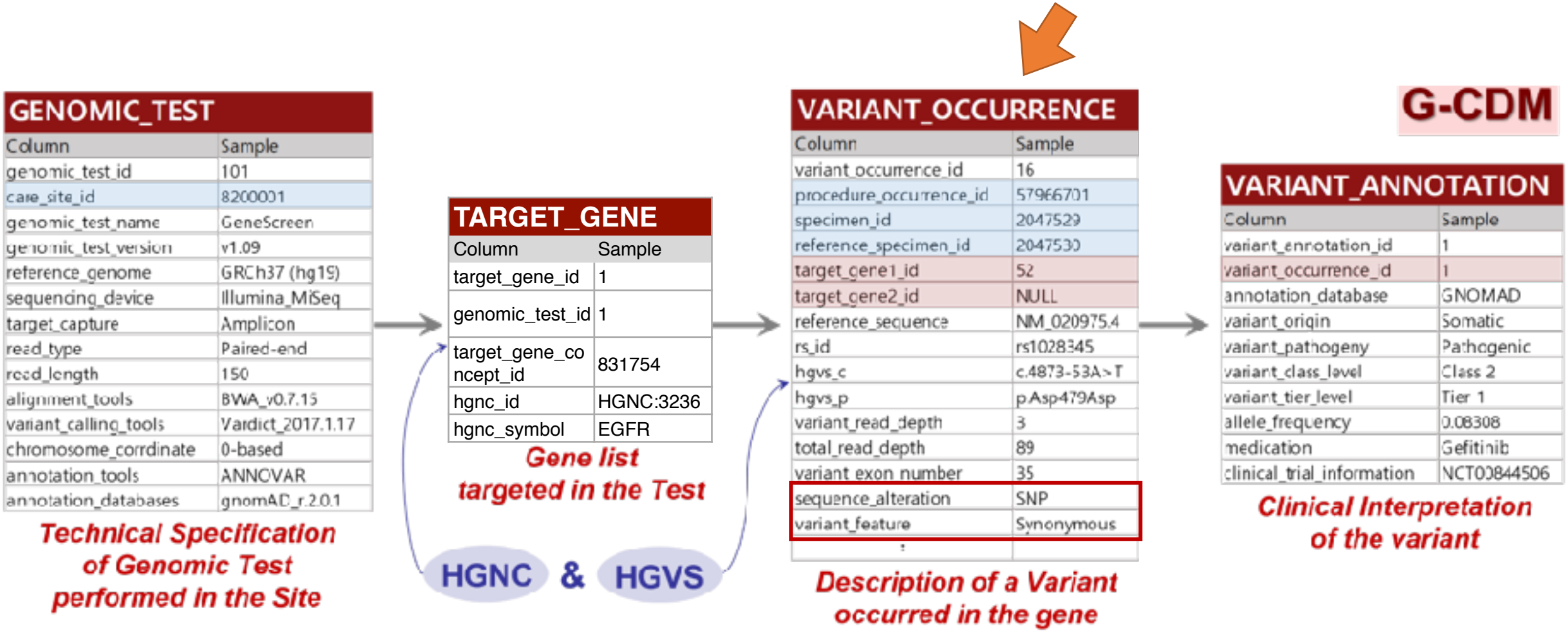

Cohort A




Cohort B

# Concept ID Request

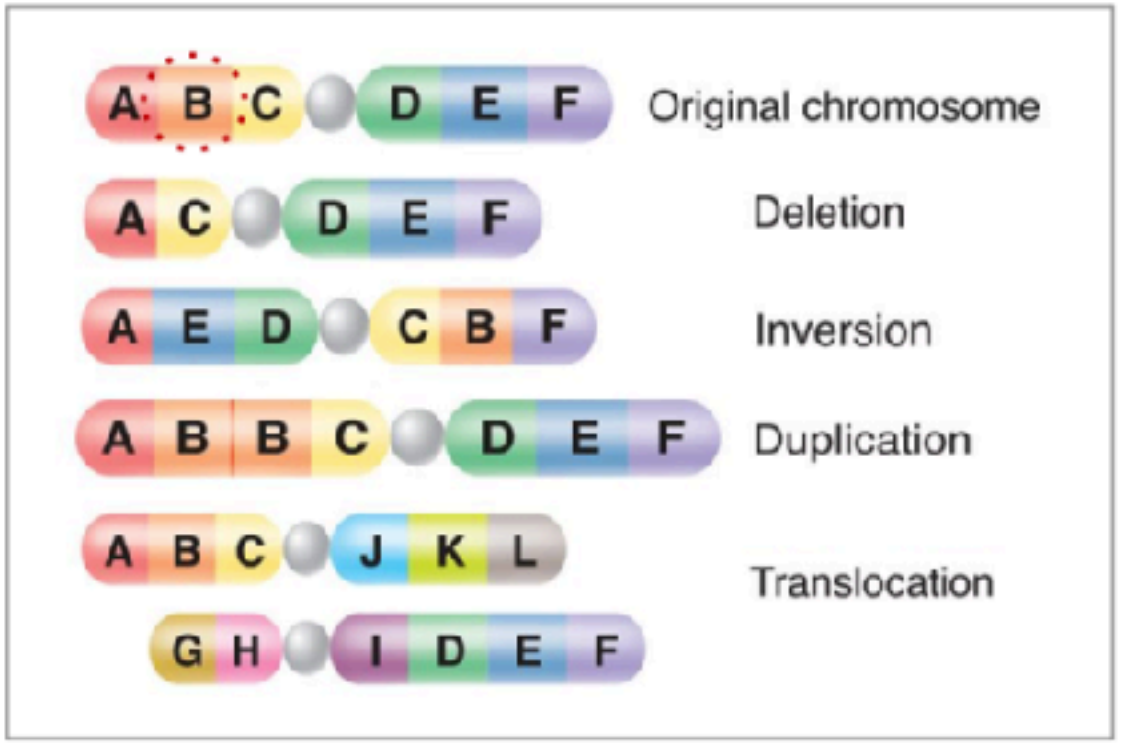
Concept IDs are needed for '**Variant_occurrence**' table.



# Concept ID Request

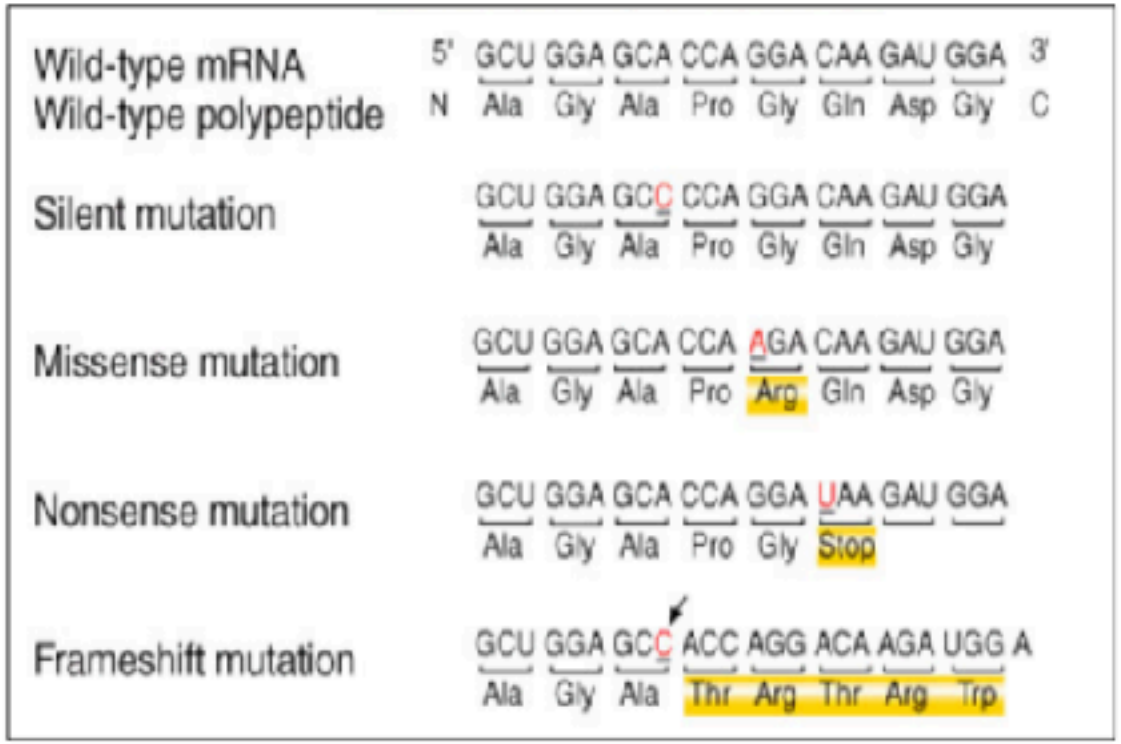
Concept IDs are needed for '**Variant_occurrence**' table.

## 1. Sequence Alteration




DNA Level **Structural** Variant Types

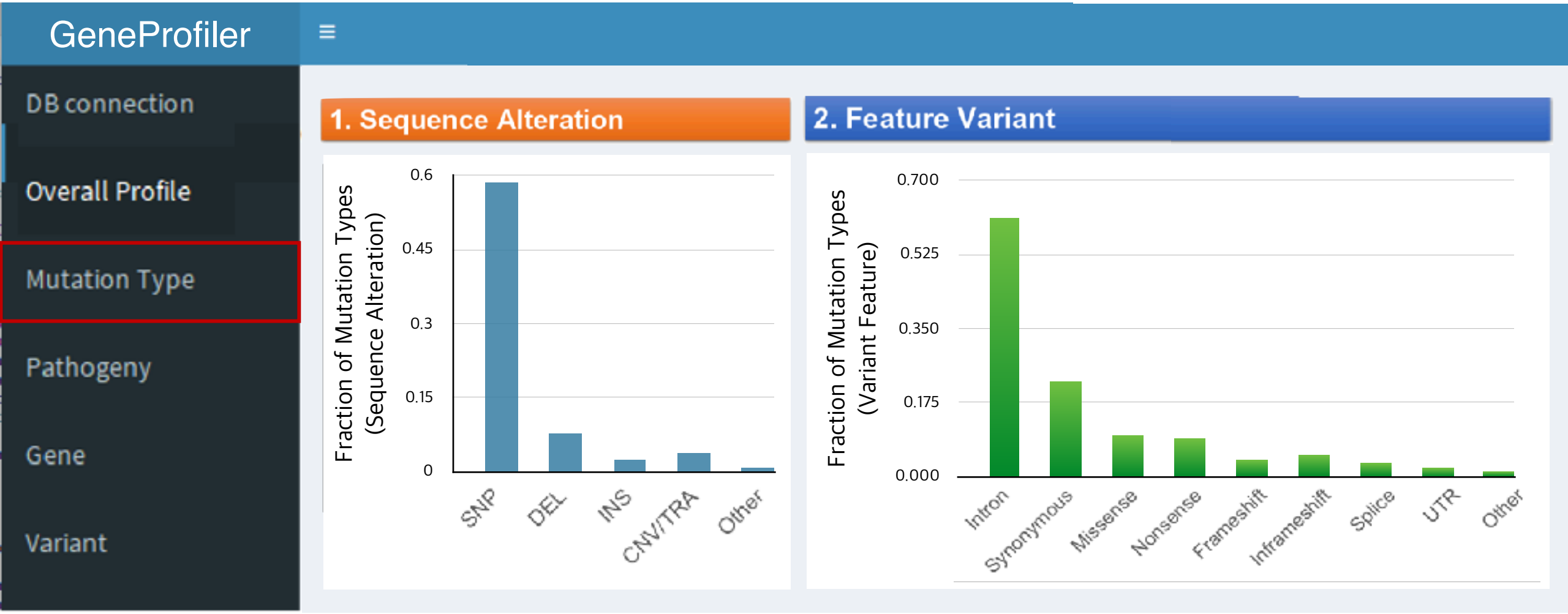
## 2. Feature Variant




Protein Level **Functional** Variant Types

# Use-case of the Concept ID

Data Profiling Tool for Genomic Data

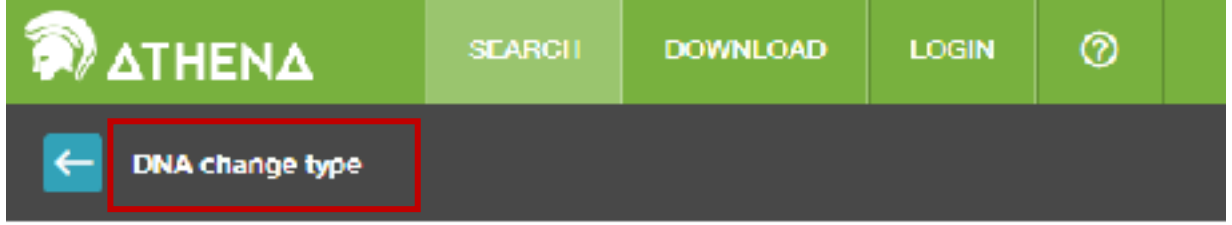



**DETAILS**

| | |
|---|---|
| Domain ID | Measurement |
| Concept Class ID | Lab Test |
| Vocabulary ID | LOINC |
| Concept ID | 3051414 |
| Concept code | 48006-1 |
| Invalid reason | Valid |
| Standard concept | Standard |
| Synonyms | Acid; Acids; Blood; Exchange; Genetic; Genomic; HL7.GENETICS; Molecular genetics; MOLPATH.GENERAL; Nominal; PCR; Point in time; Random; Replace; Tissue; Tissue, unspecified; Typ; WB; Whole blood; Whole blood or Tissue<br>Amino acid change type<br>Amino acid change [Type] |
| Valid start | |
| Valid end | 12/31/2099 |

| | | | |
|---|---|---|---|
| Has Answer (LOINC) | Deletion | 45879448 | LOINC |
| | Duplication | 45879451 | LOINC |
| | Frameshift | 45878252 | LOINC |
| | Initiating Methionine | 45884100 | LOINC |
| | Insertion | 45878601 | LOINC |
| | Insertion and Deletion | 45878251 | LOINC |
| | Missense | 45881183 | LOINC |
| | Nonsense | 45879449 | LOINC |
| | Silent | 45879450 | LOINC |
| | Stop Codon Mutation | 45884101 | LOINC |
| | | | LOINC |

Variant types were assigned by several categories and **not enough.**




**DETAILS**

| | |
|---|---|
| **Domain ID** | Measurement |
| **Concept Class ID** | Lab Test |
| **Vocabulary ID** | LOINC |
| **Concept ID** | 3051367 |
| **Concept code** | 48019-4 |
| **Invalid reason** | Valid |
| **Standard concept** | Standard |
| **Synonyms** | Blood; Deoxyribonucleic acid; Exchange; Genetic; Genomic; HLA GENETICS; Molecular genetics; MOLPATH.GENERAL; Nominal; PCR; Point in time; Random; Replace; Tissue; Tissue, unspecified; Typ; WB; Whole blood; Whole blood or Tissue DNA change type DNA Change Type |
| **Valid start** | |
| **Valid end** | |

| | | | |
|---|---|---|---|
| Has Answer (LOINC) | Complex | 21498573 | LOINC |
| | Copy number gain | 45879160 | LOINC |
| | Copy number loss | 45880003 | LOINC |
| | Deletion | 45879440 | LOINC |
| | Duplication | 45879451 | LOINC |
| | Insertion | 45879001 | LOINC |
| | Insertion/Deletion | 45879002 | LOINC |
| | Interchromosomal breakpoint | 21499301 | LOINC |
| | Intrachromosomal breakpoint | 21499300 | LOINC |
| | Inversion | 45881102 | LOINC |
| | Mobile element insertion | 21498441 | LOINC |
| | Novel sequence insertion | 21498571 | LOINC |
| | Sequence alteration | 21499039 | LOINC |
| | Substitution | 45880312 | LOINC |
| | Tandem duplication | 21499572 | LOINC |
| | | | LOINC |
| | Wild type | | LOINC |

Variant types were assigned by several categories and **not enough.**

ATHENA | SEARCH | DOWNLOAD | LOGIN

Structural variant [Type]

**DETAILS**

| | |
|---|---|
| Domain ID | Measurement |
| Concept Class ID | Lab Test |
| Vocabulary ID | LOINC |
| Concept ID | 21406047 |
| Concept code | 81288-1 |
| Invalid reason | Valid |
| Standard concept | Standard |
| Synonyms | Molecular pathology; MOLPATH; Nominal; Point in time; Random; Struct var; Typ<br>Structural variant [Type]<br>Struct var type |
| Valid start | 08/21/2018 |
| Valid end | |

| RELATIONSHIP | RELATES TO | CONCEPT ID | VOCABULARY |
|---|---|---|---|
| Has Answer (LOINC) | Complex | 21498573 | LOINC |
| | Copy number gain | 45878168 | LOINC |
| | Copy number loss | 45880603 | LOINC |
| | Deletion | 45879449 | LOINC |
| | Duplication | 45879451 | LOINC |
| | Insertion | 45878601 | LOINC |
| | Interchromosomal breakpoint | 21400301 | LOINC |
| | Intrachromosomal breakpoint | 21400300 | LOINC |
| | Inversion | 45884102 | LOINC |
| | Mobile element insertion | 21408441 | LOINC |
| | Novel sequence insertion | 21498571 | LOINC |
| | Sequence alteration | 21498830 | LOINC |
| | Tandem duplication | 21498572 | LOINC |
| | | | LOINC |

Variant types were assigned by several categories and **not enough.**

# Concept ID Request

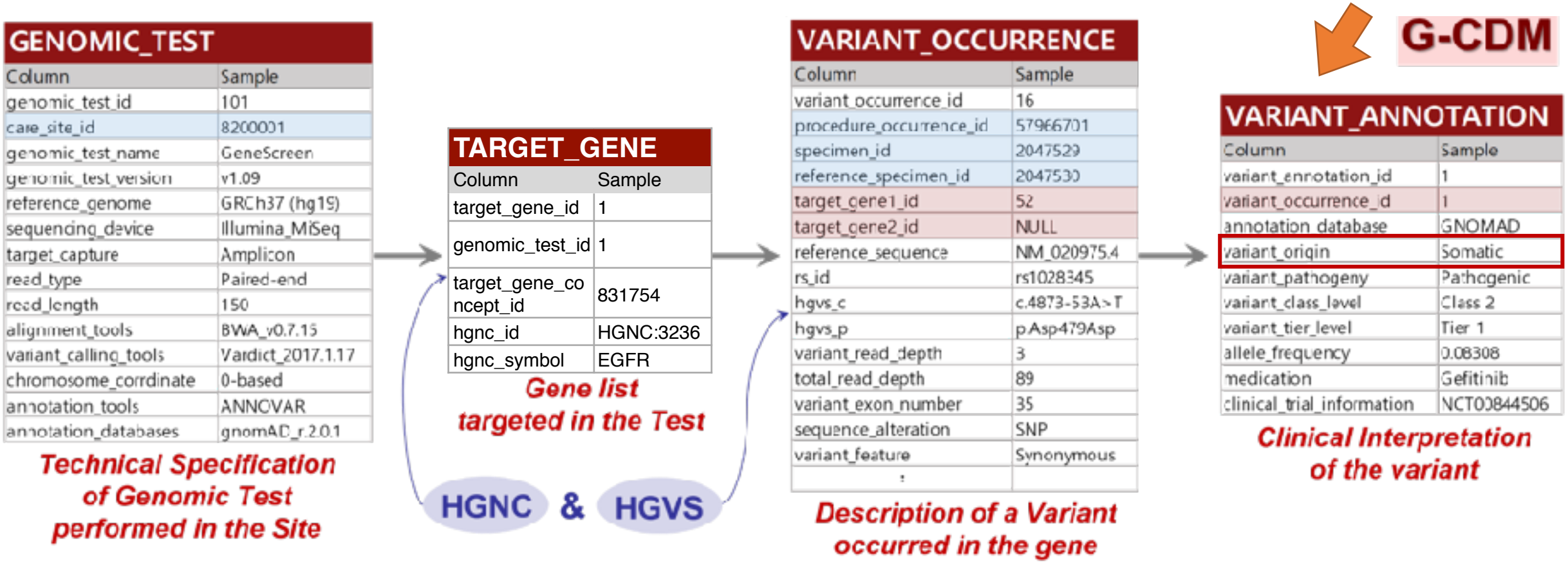
**10** more concept_id are needed to create for variant types.

| # | Sequence_alteration (Structural Types) | | | Concept ID |
|---|---|---|---|---|
| 1 | SNP<br>(Single Nucleotide Polymorphism) | SNV<br>(Single Nucleotide Variation) | Substitution | **45880312 (Substitution)** |
| 2 | DIP<br>(Deletion Insertion Polymorphism) | INS | Insertion | **45878601 (Insertion)** |
| 3 | | DEL | Deletion | **45879448 (Deletion)** |
| 4 | MNP<br>(Multiple nucleotide polymorphism) | MNP | | **MNP** |
| 5 | CNV<br>(Copy Number Variation) | Copy number gain | Amplification | **45878168 (Copy number gain)** |
| 6 | | Copy number loss | Deletion | **45880603 (Copy number loss)** |
| 7 | Translocation | Translocation | Fusion | **21499257 (Translocation)** |
| 8 | MIXED | MIXED | Mixed / Complex | **21498573 (MIXED)** |

| # | Variant_feature (Functional Types) | | | Concept ID |
|---|---|---|---|---|
| 1 | Locus Region | Upstream | | **Upstream** |
| 2 | | Downstream | | **Downstream** |
| 3 | Coding Region | Synonymous | Silent | **45879450 (Silent)** |
| 4 | | Missense | | **45881183 (Missense)** |
| 5 | | Nonsense | Stop-gained /<br>Stop-codon-mutation | **45879449 (Nonsense),**<br>**45884101 (Stop Codon Mutation)** |
| 6 | | Stop-loss | Stop-lost | **Stop-loss** |
| 7 | | Frameshift | | **45878252 (Frameshift)** |
| 8 | | Inframe | | **Inframe** |
| 9 | Untranslated Region | UTR | 5_prime_UTR | **5_prime_UTR** |
| 10 | | | 3_prime_UTR | **3_prime_UTR** |
| 11 | Intron Region | Intron | | **Intron** |
| 12 | Spliece Site | Splice | Splice_donor_variant | **Splice_donor** |
| 13 | | | Splice_acceptor_variant | **Splice_acceptor** |

# Concept ID Request

Concept IDs are needed for '**Variant_annotation**' table.



ATHENA | SEARCH | DOWNLOAD | LOGIN

Genomic source class [Type]

**DETAILS**

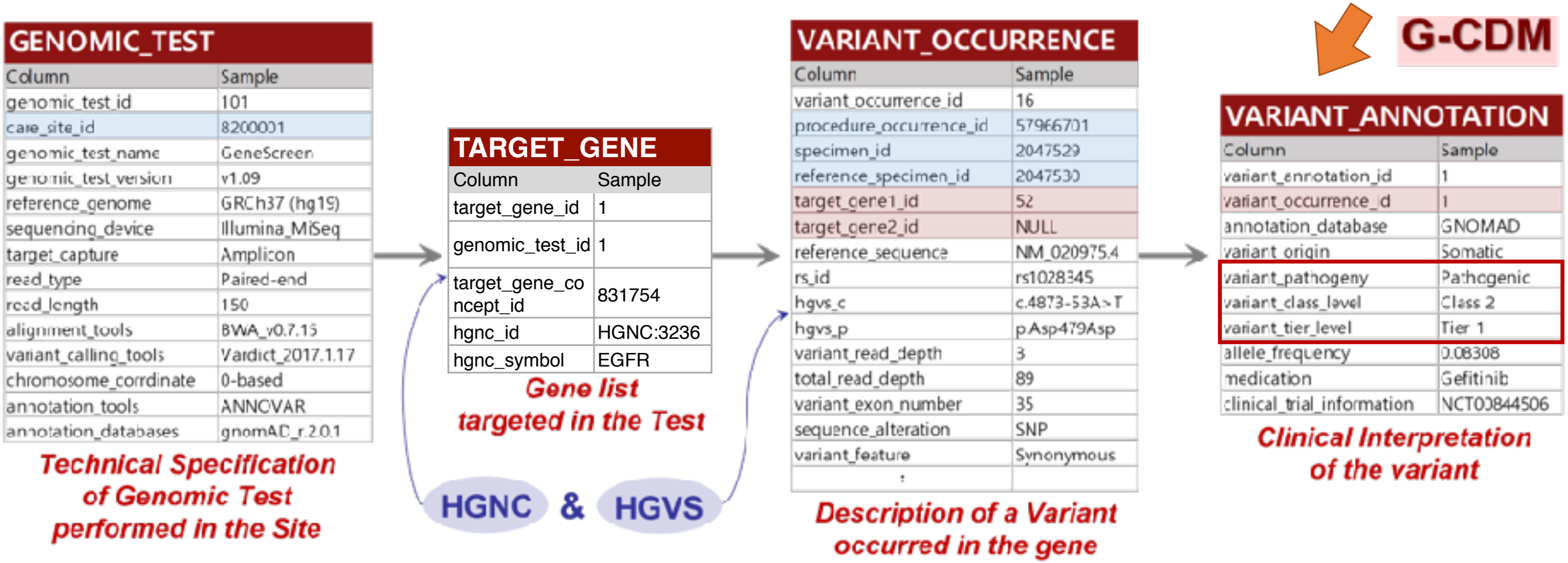
| | |
|---|---|
| **Domain ID** | Measurement |
| **Concept Class ID** | Lab Test |
| **Vocabulary ID** | LOINC |
| **Concept ID** | 3018520 |
| **Concept code** | 48002-0 |
| **Invalid reason** | Valid |
| **Standard concept** | Standard |
| **Synonyms** | Blood; Genetic; Genetic variant source; Genetics; Genomic; HL7.GENETICS; Molecular genetics; MOLPATH.GENERAL; Nominal; PCR; Point in time; Random; Src; Tissue; Tissue, unspecified; Typ; WB; Whole blood; Whole blood or Tissue Genomic source class Genomic source class [Type] |
| **Valid start** | |
| **Valid end** | |

| | | |
|---|---|---|
| De novo | 42531103 | LOINC |
| Fetal | 45881810 | LOINC |
| Germline | 45880308 | LOINC |
| Likely fetal | 45881181 | LOINC |
| Likely germline | 45878600 | LOINC |
| Likely somatic | 45881180 | LOINC |
| Somatic | 45880309 | LOINC |
| Unknown genomic origin | 45881182 | LOINC |

It’s enough to express an origin of the variant.

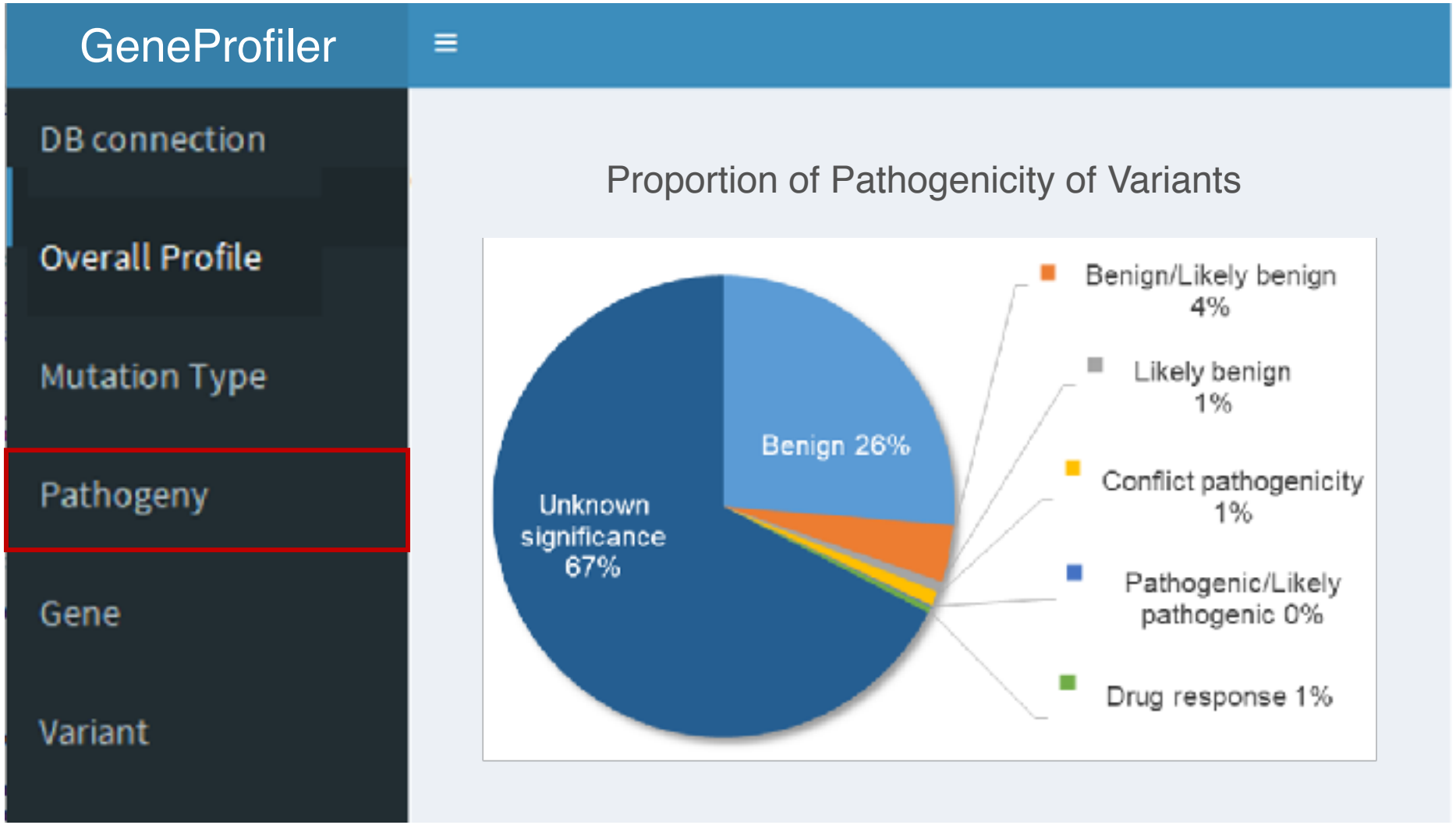
# Concept ID Request

Concept IDs are needed for '**Variant_annotation**' table.



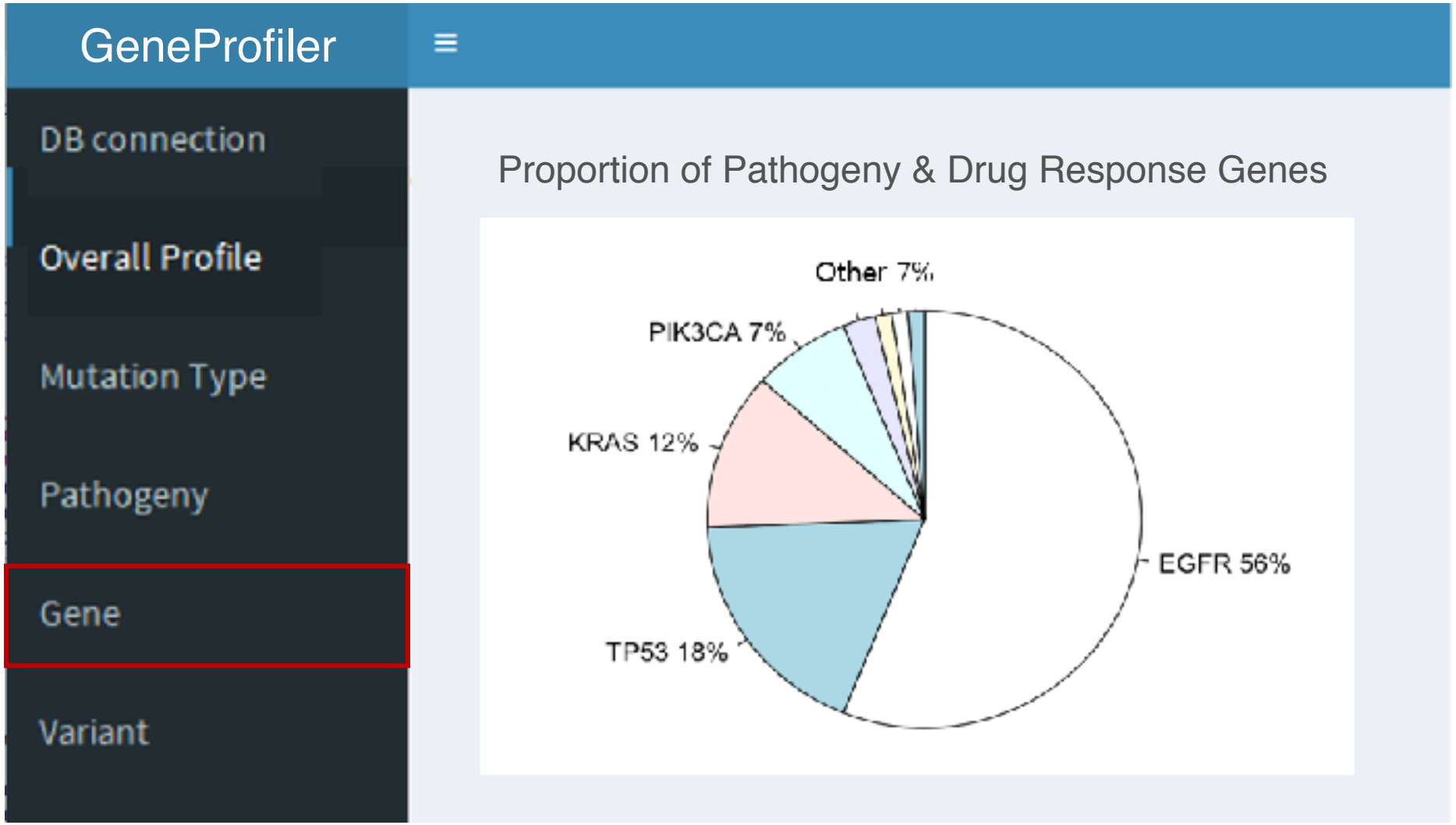
# Use-case of the Concept ID

Data Profiling Tool for Genomic Data



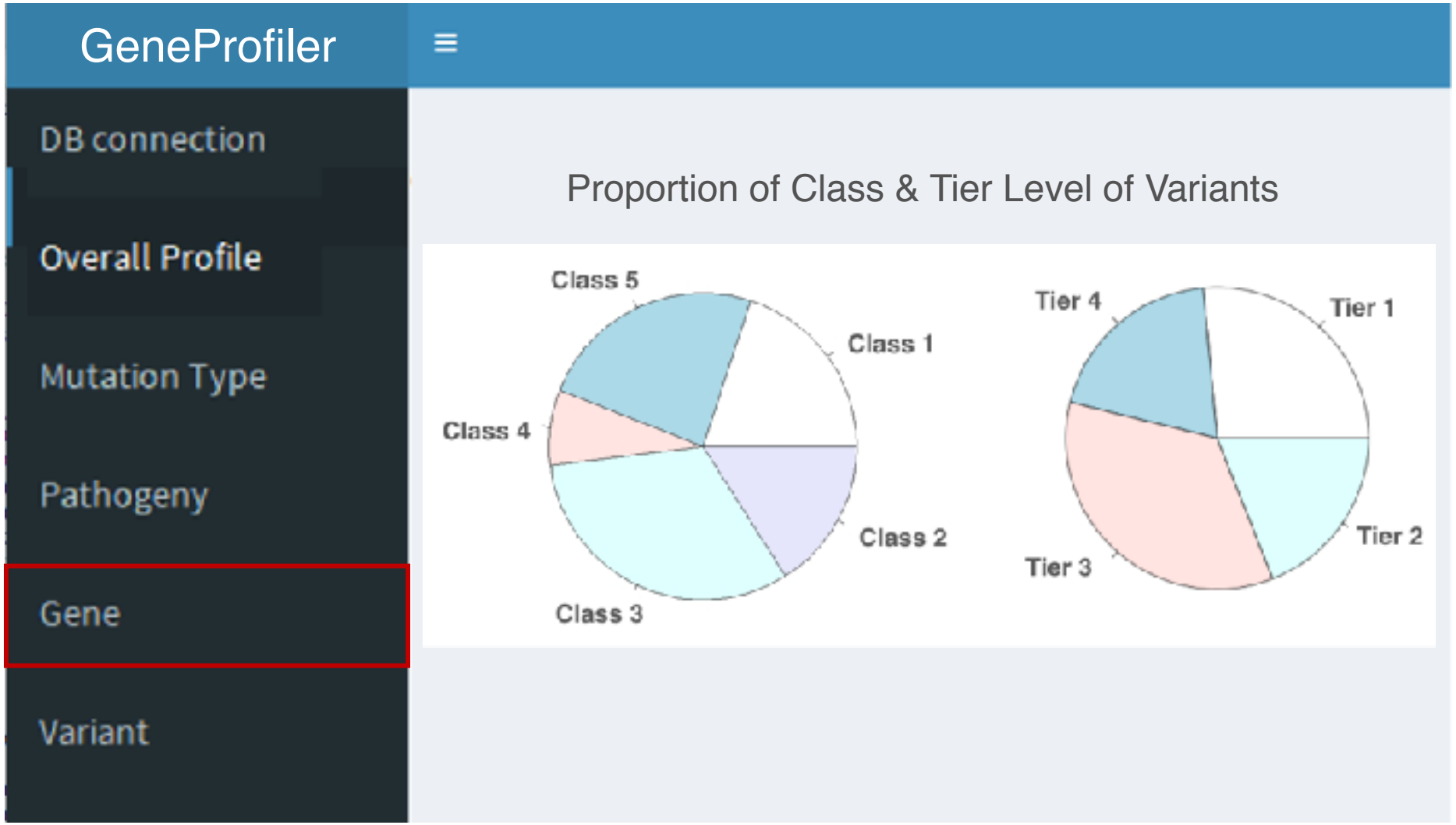
# Use-case of the Concept ID

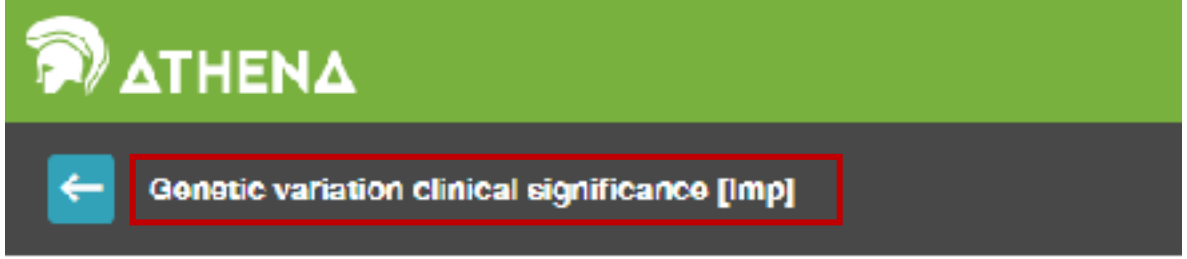
Data Profiling Tool for Genomic Data



# Use-case of the Concept ID

Data Profiling Tool for Genomic Data






| | | |
|---|---|---|
| Benign | 45880925 | LOINC |
| Likely benign | 21498359 | LOINC |
| Likely pathogenic | 21498358 | LOINC |
| Pathogenic | 45884094 | LOINC |
| Uncertain significance | 21498841 | LOINC |

—> Fit for only a '**Clinvar**' annotation database.

| | |
|---|---|
| annotation database | Clinvar, PolyPhen, SIFT, SnpEff, 1000G, GNOMAD, ExAC |

Can't express all annotation results driven from **diverse databases.**

# Concept ID Request

Concept IDs are needed for '**Variant_annotation**' table.

Class Level

**Table 2**

**Classification methods for somatic cancer variants**

| Classification[a] | SVC method [21••] | PHIAL method [22••] | BWH/DFCI method |
|---|---|---|---|
| Class 1 | Clinically actionable for therapeutic, prognostic, or diagnostic purposes for same tumor type | Validated therapeutic, prognostic, or diagnostic implications for same tumor type | Validated therapeutic, prognostic, or diagnostic implications for same tumor type |
| Class 2 | Clinically actionable for therapeutic, prognostic, or diagnostic purposes for a different tumor type | Limited evidence of therapeutic, prognostic, or diagnostic implications for same tumor type | Validated therapeutic implications for a different tumor type, or limited evidence of prognostic or diagnostic implications for same tumor type |
| Class 3 | Other variants in this gene in this primary tumor are established as actionable for same tumor type | Clinical evidence of therapeutic response from another tumor type | Preclinical or inferential therapeutic, prognostic, or diagnostic implications |
| Class 4 | Other variants in this gene in this primary tumor are established as actionable for a different tumor type | Preclinical association to therapeutic response | Novel or unstudied in cancer |
| Class 5 | (A) Gene is not actionable for any tumor type (B) Established as benign | Inferential association to therapeutic response | Established as benign |

BWH/DFCI – Brigham Women's Institute/Dana Farber Cancer Institute.
[a] PHIAL method classifications referred to as Levels A–E, while other methods refer to as Tiers.

# Concept ID Request

Concept IDs are needed for '**Variant_annotation**' table.

Tier Level

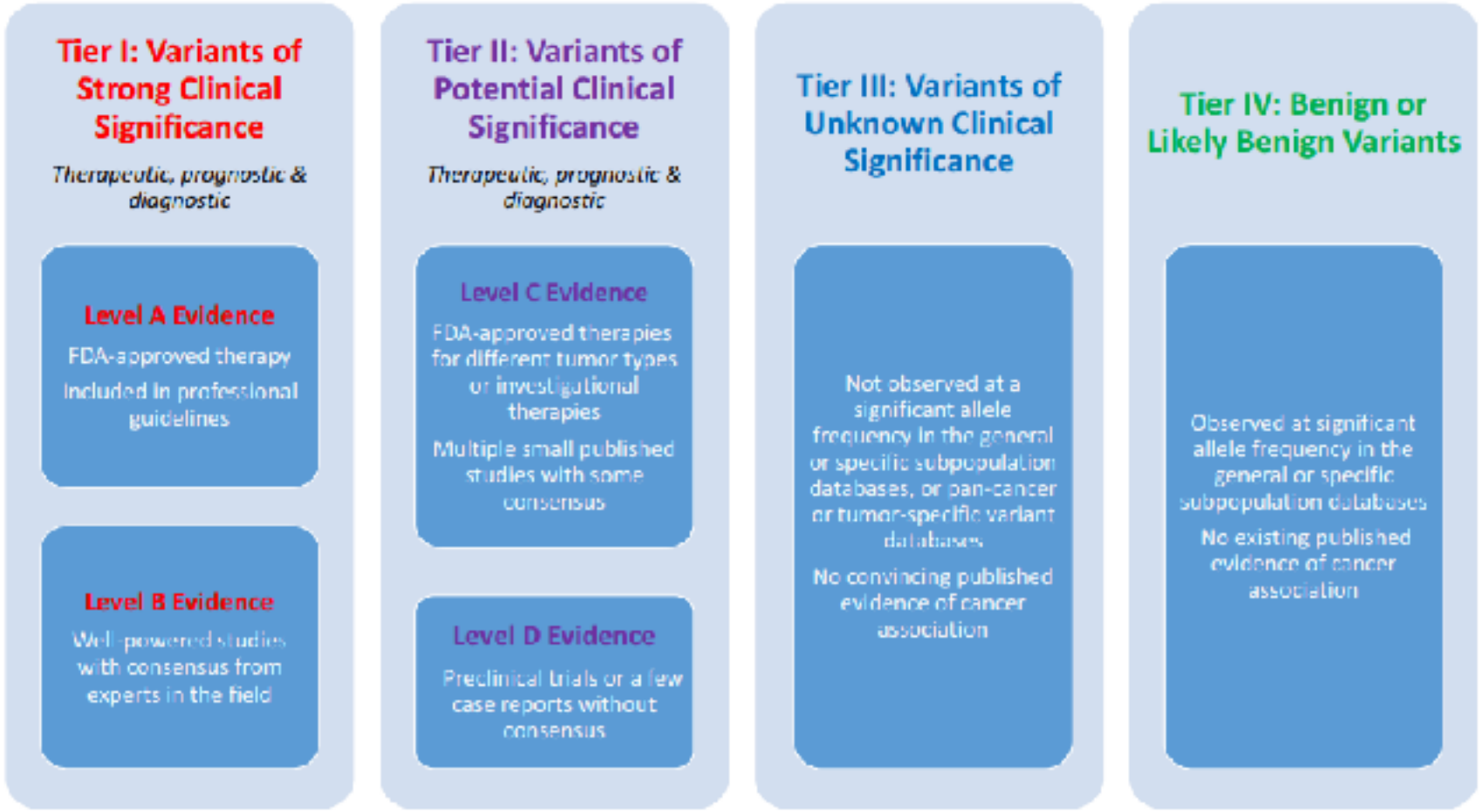



**Figure 2** Evidence-based variant categorization. Somatic variants are classified into four tiers based on their level of clinical significance in cancer diagnosis, prognosis, and/or therapeutics. Variants in tier I are of strongest clinical significance, and variants in tier IV are benign or likely benign variants. FDA, Food and Drug Administration.

# Concept ID Request

| Table | Column | Concept_name | Number of Concept_id |
|---|---|---|---|
| Target_gene | target_gene_concept_id | Approved Gene Symbols of HGNC Database | 41503 |
| Variant_occurrence | sequence_alteration | MNP | 1 |
| | variant_feature | Upstream, Downstream, Stop-loss, Inframe, 5_prime_UTR, 3_prime_UTR, Intron, Splice_donor, Splice_acceptor | 9 |
| Variant_annotation | annotation_database | Clinvar, PolyPhen, SIFT, SnpEff, 1000G, GNOMAD, ExAC | 7 |
| | variant_pathogeny | Low, Modifier, Moderate, High | 4 |
| | | Benign, Benign/Likely benign, Likely benign, Unknown significance, Likely pathogenic, Likely pathogenic/ Pathogenic, Conflict pathogenic, Pathogenic, Drug response | 9 |
| | | (Benign), Possibly damaging, Probably damaging | 2 |
| | | Tolerated, Tolerated (low confidence), Deleterious, Deleterious (low confidence) | 4 |
| | variant_class_level | Class 1~5 | 5 |
| | variant_tier_level | Tier 1~4 | 4 |
| | | | 41548 |

45

Thank you for Listening !