Comparing fields from variants knowledge bases and Beacon v2

May 21, 2020

Overview

Variant Knowledge bases & Case-level databases

Variant Knowledge bases: ClinVar

Case-level databases: Genomics England, Shariant, Clinical Variant Ark, DECIPHER

variant-level (aggregated) db: ClinVar

- (identification) Variant ID
- (identification) Assembly ID and version, assembly name and version; RefSeq ID; start and end coordinates e.g GRCh38.p12 (GCF_000001405.38) Chr 10 (NC_000010.11):87,852,589 - 87,982,779)
- Variant type e.g CNV, seq alteration, tandem dup, insertion, mobile element insertion, etc
- (features) Genes, transcripts
- (metadata) Molecular consequence
- (source) Source database
- (significance) Clinical significance (most severe one) e.g pathogenic
- MAF 3 values: 1000 Genomes, GO_ESP and ExAC
- (evidence type) Has publications

case-level db: Genomics England

- (identification) (Assembly ID, refseq, coordinates)
- (identification) Variant ID
- (provenance) patient ID
- (provenance) family ID
- (var metadata) Variant type
- (var metadata) Allele origin
- (var metadata) Mode of inheritance
- (var metadata) Zigosity in the tested individual
- (var metadata) Features
- (var metadata) Consequence type
- (significance) Deleteriousness prediction- prediction tools
- (significance) Clinical significance (GeCIP)
- (evidence type) Technical details of tests for variant detection eg. panel applied
- (phenotype) Disease group, disease
- (var metadata) Aligner
- (var metadata) Variant Caller

case-level db: Shariant

- (provenance) Lab record ID
- (identification) c.hgvs including gene symbol, ref, alt, transcript and build (assembly namealias e.g ghg19), location
- (significance) Clinical significance value (pathogenic to bening)
- (significance) ACGM criteria ACMGs scale all 28 criteria e.g BS1, BS2, PME, etc
- (evidence type) Structured evidence for clinical significance (database, website, bioinformatic prediction tools e.g "Population Data", Computational and Predictive Data , "Functional Data")
- (var metadata) Zigosity in the tested individual
- (evidence type) Technical details of tests for variant detection
- (evidence type) Details curating laboratory
- (phenotype) Condition. Details on conditions and phenotype
- (phenotype) Clinical phenotype : HPO, SNOMEDCT

case-level db: DECIPHER

- (provenance) patient ID
- (identification) Build ID, Transcript ID, location
- (var metadata) Molecular consequence
- (var metadata) Aminoacid Change
- (significance) Pathogenicity/contribution

case-level db: Clinical Variant Ark

- (identification) Assembly, transcript, location, ref alt)
- (provenance) patient ID
- (provenance) case ID
- (provenance) family ID
- (var metadata) Molecular consequence
- (var metadata) Aminoacid Change
- (var metadata) Features (Gene, protein)
- (significance) Pathogenicity
- (var metadata) Mode of inheritance
- (phenotype) Phenotype (HPO)
- (phenotype) Clinical indication
- (evidence type) Panel
- (evidence type) Interpretation service

Metadata fields

Field	Shariant	GEL	Others	Beacon
identification	X	X	X	Variant Identification
var type	X	X	X	var type
consequence	X	X	X	mol consequence
relation to genes			X	genomic reg, feat
features	X	X	X	feature id
aa change	X	X	X	aa change
aa subs score			X	-

Case-level fields

Field	Shariant	GEL	Others	Beacon
provenance	X	X	X	Bios ID, Ind ID, Pedigree ID
AF	X	X	X	Bios ID, Ind ID, frequency
evidence type	X	X	X	evidence type*
mode inheritance		X	X	allele origin
zigosity	X			zigosity
phenotype	X	X	X	phenotypic feat, disease
co-located vars				vars in sample?*

Variant-level fields (in Variant Annotation)

Field	**ClinVar**	**Others**	**Beacon**
external variant tds		X	alternative lds
source		X	-
penetrance		X	-
inheritance mode	X	X	allele origin
max AF	X		-
sinificance	X	X	var effect, clin sign*
haploinsufficiency score		X	-
conservation		X	-

Conclusions

- Most metadata and variant-level fields are already Beacon V2 schema (since it inherits from ClinVar)
- Some variant-level fields need to be added to Beacon V2 schema
- Most case-level fields are already Beacon V2 schema
- Some case-level fields need to be added to Beacon V2 schema
- Some case-level fields (*) in Beacon V2 schema need to be generalized to accept values from the different sources (eg. evidence type to include type e.g "Population Data" and source eg "gnomad" and technical details of test, clinical significance to add criteria such as ACGM and interpretation services eg. Exomizer)