Beacon v.2 default schemas v0.1

Variant Identification

assemblyId CATEGORICAL VALUE Genomic assembly accession as Genome Reference consortium Human name, e.g. "GRCh38"

chromosome CATEGORICAL VALUE Chromosome name, e.g. "9", "X", "MT"

start NUMERIC VALUE Start position of variant

end NUMERIC VALUE End position of variant

referenceBases ALPHANUMERIC VALUE Reference sequence in start-end coordinates

alternateBases ALPHANUMERIC VALUE Alternate sequence in start-end coordinates

variantType CATEGORICAL VALUE (ONTOLOGY LABEL) from Sequence Ontology describing the type of variant, e.g. "SNV" (SO:0001483), "structural variant" (SO:0001537)

Individual

individualId ALPHANUMERIC VALUE Individual identifier (external accession or internal ID)

sex CATEGORICAL VALUE (ONTOLOGY LABEL) Sex of individual. Value from NCIT General Qualifier (NCIT:C27993) ontology: "UNKNOWN" (not assessed or not available) (NCIT:C17998), "FEMALE" (NCIT:C46113), "MALE", (NCIT:C46112) or "OTHER SEX" (NCIT:C45908)

ethnicity CATEGORICAL VALUE (ONTOLOGY LABEL) Ethnic background of individual. Value from NCIT Race ontology (NCIT:C17049), e.g. "Latin American" (NCIT:C126531)

geographicOrigin CATEGORICAL VALUE (ONTOLOGY LABEL) Individual's country or region of origin (birthplace or residence place regardless of ethnic origin). Value from GAZ Geographic Location ontology (GAZ:00000448), e.g. "United States of America" (GAZ:00002459)

diseases (List of) disease(s) been diagnosed to the individual, defined by disease ID, age of onset, stage and the presence of family history

- **diseaseId** CATEGORICAL VALUE (DISEASE CODE/ONTOLOGY LABEL) Disease identifier. Value from ICD10 disease codes or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO, e.g. "lactose intolerance" (HP:0004789)
- **ageOfOnset** Individual's age at onset/ diagnosis of disease
  - **age** ALPHANUMERIC Age, in ISO8601 duration format
  - **ageGroup** CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)
- **stage** CATEGORICAL VALUE (ONTOLOGY LABEL) from Ontology for General Medical Science or Disease Stage Qualifier ontology (NCIT:C28108), e.g. "acute onset" (OGMS:0000119)
- **familyHistory** BOOLEAN indicating determined or self-reported presence of family history of the disease

pedigrees (List of) pedigree studi(es) in which the individual is part of

- **pedigreeID** ALPHANUMERIC VALUE Pedigree identifier

pedigreeRole CATEGORICAL VALUE (ONTOLOGY LABEL) Pedigree role, defined as relationship to proband. Value from HL7 code for family relationship or Relationship to Proband ontology (ERO:0002112), e.g. "self" (ERO:002036), "identical twin relationship" (ERO:0002041)

numberOfIndividualsTested NUMERIC VALUE Number of individuals in pedigree, including proband

diseaseId CATEGORICAL VALUE (DISEASE CODE/ONTOLOGY LABEL Disease identifier for disease focus of the pedigree. Value from ICD10 disease codes or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO, e.g. "lactose intolerance" (ICD10CM:E73). Affected individuals in pedigree will have Diseases.diseaseId (diagnosed disease) matching Pedigree.diseaseId

info

Biosample

biosampleId ALPHANUMERIC VALUE Biosample identifier (external accession or internal ID)

individualId Reference to Individual ID (Individual.individualId)

description FREE TEXT Any relevant info about the biosample that does not fit into any other field in the schema

biosampleStatus CATEGORICAL VALUE (ONTOLOGY LABEL) from Experimental Factor Ontology (EFO) Material Sample ontology (OBI:0000747). Classification of the sample in "abnormal sample "(EFO:0009655) or "reference sample" (EFO:0009654)

individualAgeAtCollection Individual's age at the time of sample collection

age ALPHANUMERIC VALUE Age, in ISO8601 duration format

ageGroup CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)

sampleOrigin Values specifying the origin of the biosample in organ, tissue and cell type/cell line

organ CATEGORICAL VALUE (ONTOLOGY LABEL) from Uber-anatomy ontology (UBERON) or BRENDA tissue / enzyme source (BTO) ontologies identifying the source organ of the biosample, e.g. "liver" (UBERON:0002107)

tissue CATEGORICAL VALUE (ONTOLOGY LABEL) from Uber-anatomy ontology (UBERON) or BRENDA tissue / enzyme source (BTO) ontologies identifying the source tissue of the biosample, e.g. "hepatic sinusoid" (UBERON:0001281)

cellType CATEGORICAL VALUE (ONTOLOGY LABEL) from BRENDA tissue / enzyme source (BTO) or Cell Ontology (CL) ontologies identifying the source cell type or cell line origin of the biosample, e.g. "Kupffer cell" (CL:0000091)

obtentionProcedure CATEGORICAL VALUE (ONTOLOGY LABEL) from NCIT Intervention or Procedure ontology describing the procedure for sample obtention, e.g. "Biopsy" (NCIT:C15189)

cancerFeatures Values specifying cancer-specific features, including progression and tumor grade

tumorProgression CATEGORICAL VALUE (ONTOLOGY LABEL) from Neoplasm by Special Category ontology (NCIT:C7062). Tumor progression category indicating primary, metastatic or recurrent progression, e.g. "Primary Malignant Neoplasm" (NCIT:C84509)

tumorGrade CATEGORICAL VALUE (ONTOLOGY LABEL) from Tumor Grading Characteristic ontology (Mondo Disease Ontology MONDO:0024488) General tumor grading

info

Variant Annotation

variantAlternativeIds (LIST OF) ALPHANUMERIC VALUE(S) Cross-referencing ID(s) (CURIE(s)) for the variant in the original databases or variant-level (aggregated) databases for previously described variants, e.g. "VCV000055583.1", "CA003602"

genomicHGVSId ALPHANUMERIC VALUE HGVSId descriptor at genomic level, referred to genome assembly defined in Variant Identification, e.g. "NC_000017.10:g.41199678C>A"

proteinHGVSId (LIST OF) ALPHANUMERIC VALUE HGVSId descriptor(s) at protein level (for protein-altering variants), e.g. "NP_009225.1:p.Glu1817Ter" or "LRG_199p1:p.Val25Gly"

variantGeneRelashionship CATEGORICAL VALUE (ONTOLOGY LABEL) classifying the variant according to the broadness of its effect in terms of genes: "intergenic", "single-gene" (exonic, intronic), "in overlapping genes" (exonic, intronic), "spanning multiple genes"

geneIds (LIST OF) ALPHANUMERIC VALUE(S) HGNC ID(s) for the gene(s) affected by the variant

transcriptIds (LIST OF) ALPHANUMERIC VALUE(S) ENSEMBML ID(s) for the transcript(s) affected by the variant

molecularConsequence (LIST OF) CATEGORICAL VALUE(S) (ONTOLOGY LABEL(S)) from Sequence Ontology describing the molecular consequence of the variant for protein-altering variants, e.g. "missense variant" (SO:0001583)

clinicalRelevance (List of) descriptor(s) of clinical relevance ascribed to the variant, including the variant classification, the disease identifier and references of studies supporting the association

> **diseaseId** Reference to DiseaseId from Disease object from Individual schema, e.g. "Hereditary breast ovarian cancer syndrome" (MONDO:0003582)
>
> **variantClassification** CATEGORICAL VALUE Value describing the effect of the variant on the disease: benign, likely benign, pathogenic, likely pathogenic or unknown
>
> **references** (List of) PUBMED ID(s) of studies describing the variant-disease association, e.g. "PMID:27153395"

info