

# Beacon v.2 default schemas v0.2

## Dataset

**datasetId** CATEGORICAL VALUE Dataset reference ID

**datasetSource** CATEGORICAL VALUE Reference to dataset source, e.g. "DECIPHER", "DisGenNET"

**datasetType** CATEGORICAL VALUE Type of dataset: "case-level" or "variant-level" (aggregated)

## Variant Identification

**variantId** ALPHANUMERIC VALUE Reference to variant ID

**assemblyId** CATEGORICAL VALUE Genomic assembly accession and version as RefSeq assembly accession (e.g. "GCG\_000001405.39") or a versioned assembly name or synonym such as UCSC Genome Browser assembly (e.g. "hg38") or Genome Reference consortium Human (e.g. GRCh38.p13) names

**refseqId** CATEGORICAL VALUE Reference sequence Refseq ID and version for genomic reference sequence in which variant coordinates are given, e.g. "NC\_000009" for human chromosome 9. Alternatively, names, synonymous or aliases e.g. "Chr9" when **assemblyId** is given. For organisms with a single reference sequence covering the genome, the versioned Refseq ID can be given here as an alternative to the assembly ID and version in **assemblyId**, e.g. "NC\_045512.2" for SARS-CoV2 full-length genome reference sequence.

**start** NUMERIC VALUE Start position of variant

**end** NUMERIC VALUE End position of variant

**ref** ALPHANUMERIC VALUE Reference sequence in start-end coordinates

**alt** ALPHANUMERIC VALUE Alternative sequence in start-end coordinates

**variantType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of variant. Value from [Sequence Ontology](#), e.g. "SNV" (SO:0001483), "structural variant" (SO:0001537)

## Variant Annotation

**variantId** ALPHANUMERIC VALUE Reference to Variant ID

**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. clinVarId, ClinGen, COSMIC), e.g. "VCV000055583.1", "CA003602"

**genomicHGVSId** ALPHANUMERIC VALUE HGVSId descriptor at genomic level (recommended, referred to genome assembly defined in Variant Identification), e.g. "NC\_000017.10:g.41199678C>A"

**transcriptHGVSIds** (LIST OF) ALPHANUMERIC VALUE(S) HGVSId descriptor at transcript level : "NC\_000023.10(NM\_004006.2):c.357+1G

**proteinHGVSIds** (LIST OF) ALPHANUMERIC VALUE(S) HGVSId descriptor(s) at protein level (for protein-altering variants), e.g. "NP\_009225.1:p.Glu1817Ter" or LRG\_199p1:p.Val25Gly (preferred)

**genomicRegions** (LIST OF) CATEGORICAL VALUE(S) (ONTOLOGY LABEL) Classification(s) of the variant according to the genomic region affected (all that apply, relative to each feature affected). Value from [Ensembl Glossary \(ENSGLOSS\) Variant consequence ontology \(ENSGLOSSARY:00000134\)](#), e.g. "3UTR" (ENSGLOSSARY:0000159), "coding sequence variant" (ENSGLOSSARY:0000159), "upstream gene variant" (ENSGLOSSARY:0000164), "intergenic variant" (ENSGLOSSARY:0000174), "intron variant" (ENSGLOSSARY:0000161), "non-coding transcript variant" (ENSGLOSSARY:0000163)

**genomicFeatures** Genomic feature(s) affected by the variant. (List of:)

**class** CATEGORICAL VALUE (ONTOLOGY LABEL) Class of feature affected by the variant. Value from [ENSGLOSS Biotype \(ENSGLOSSARY:0000025\) ontology](#), e.g. "protein coding gene", "non-coding RNA", "long non-coding RNA"

**featureID** (ALPHANUMERIC VALUE) ID /accession/name of feature affected by the variant, matching **class**, e.g. "TP53", "GeneID:43740578"

**molecularEffects** (LIST OF) CATEGORICAL VALUE(S) (ONTOLOGY LABEL) Predicted effect at nucleotide level for protein affecting variants. Value from [Sequence ontology \(SO\) coding sequence variant \(SO:0001580\) ontology](#), e.g. "synonymous variant" (SO:0001819), "nonsynonymous variant" (SO:0001992) (and classifications therein, such as "stop gained" (SO:0001587), "missense variant" (SO:0001583), "inframe indel" (SO:0001820))

**aminoacidChanges** (LIST OF) CATEGORICAL VALUE(S) Change at aminoacid level for protein affecting missense variants e.g. "V304\*"

## Subject

**subjectId** ALPHANUMERIC VALUE Subject reference ID (external accession or internal ID)

**datasetId** ALPHANUMERIC VALUE Reference to dataset ID

**taxonId** ALPHANUMERIC VALUE Taxon ID of subject

**sex** CATEGORICAL VALUE (ONTOLOGY LABEL) Sex of subject. Value from [NCIT General Qualifier ontology \(NCIT:C27993\)](#): "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 subject. Value from [NCIT Race ontology \(NCIT:C17049\)](#). e.g "Latin American" (NCIT:C126531)

**geographicOrigin** CATEGORICAL VALUE (ONTOLOGY LABEL) Subject'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)

**phenotypicFeatures** Phenotypic feature(s) observed in the subject, defined by phenotype, date, type or age of onset and level/ severity. (List of:)

**phenotypeId** CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic feature observed. Value from [Human Phenotype Ontology \(HPO\)](#)

**dateOfOnset** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Date of onset/observation of phenotype

**onsetType** CATEGORICAL VALUE (ONTOLOGY LABEL) Onset type. Value from [HPO Onset ontology \(HP:0003674\)](#), e.g. "congenital onset" (HP:0003577), "adult onset" (HP:0003581)

**ageOfOnset** Subject's age at onset/observation of phenotype

**age** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Age

**ageGroup** CATEGORICAL VALUE (ONTOLOGY LABEL) Age group. Value from [NCIT Age Group ontology](#), e.g. "NCIT:C27954" (Adolescent)

**level/severity** CATEGORICAL VALUE (ONTOLOGY LABEL) Level/severity when and as applicable to phenotype observed. Value from [Human Phenotype Ontology \(HPO\) Severity ontology \(HP:0012824\)](#), e.g. "severe" (HP:0012828)

**diseases** Disease(s) been diagnosed to the subject, defined by disease ID, date, type or age of onset, stage, level/severity and the presence of family history. (List of:)

**diseaseId** CATEGORICAL VALUE (DISEASE CODE /ONTOLOGY LABEL) Disease ID. Value from [ICD10 disease codes](#) or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO, e.g. "lactose intolerance" (HP:0004789, ICD10CM:E73)

**dateOfOnset** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Date of onset/diagnosis of disease

**onsetType** CATEGORICAL VALUE (ONTOLOGY LABEL) Onset type. Value from [HPO Onset ontology](#) ([HP:0003674](#)), e.g. "congenital onset" (HP:0003577), "adult onset" (HP:0003581)

**ageOfOnset** Subject's age at onset/ diagnosis of disease

**age** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Age

**ageGroup** CATEGORICAL VALUE (ONTOLOGY LABEL) Age group. Value from [NCIT Age Group ontology](#), e.g. "NCIT:C27954" (Adolescent)

**stage** CATEGORICAL VALUE (ONTOLOGY LABEL) Stage of disease. Value from [Ontology for General Medical Science](#) or [Disease Stage Qualifier ontology](#) ([NCIT:C28108](#)), e.g. "acute onset" (OGMS:0000119)

**level/severity** CATEGORICAL VALUE (ONTOLOGY LABEL) Level/severity when and as applicable to disease course. Value from [Human Phenotype Ontology \(HPO\) Severity ontology](#) ([HP:0012824](#)), e.g. "mild" (HP:0012825)

**familyHistory** BOOLEAN indicating determined or self-reported presence of family history of the disease

**pedigrees** Pedigree(s) to which the subject belongs (List of:)

**pedigreeId** ALPHANUMERIC VALUE Reference to pedigree ID (external accession)

**pedigreeDisease** CATEGORICAL VALUE (DISEASE CODE /ONTOLOGY LABEL) Disease ID. Value from [ICD10 disease codes](#) or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO, e.g. "lactose intolerance" (HP:0004789, ICD10CM:E73)

**pedigreeRole** CATEGORICAL VALUE (ONTOLOGY LABEL) Pedigree role, defined as relationship to proband. Value from [Family member ontology](#) ([NCIT:C41256](#)), e.g. "proband" (NCIT:C64435), "identical twin" (NCIT:C73429), "mother" (NCIT:C25189)

**affectedStatus** CATEGORICAL VALUE Affected status of subject in disease of pedigree: "affected" or "unaffected"

**numIndTested** NUMERIC VALUE Number of subjects in pedigree

## Biosample

**biosampleId** ALPHANUMERIC VALUE Biosample reference ID (external accession)

**subjectId** Reference to subject ID

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

**biosampleStatus** CATEGORICAL VALUE (ONTOLOGY LABEL) Classification of biosample based on their role in study. Value from [Experimental Factor Ontology \(EFO\) Material Sample ontology](#) ([OBI:0000747](#)) Classification of the sample in "abnormal sample" ([EFO:0009655](#)) or "reference sample" ([EFO:0009654](#))

**collectionDate** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Date of biosample collection

**subjectAgeAtCollection** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Subject's age at the time of biosample collection

**sampleOriginDescriptors** Descriptor(s) of origin of biosample. (List of:)

**sampleOriginType** CATEGORICAL VALUE (ONTOLOGY LABEL) Category of sample origin. Value from [Ontology for Biomedical Investigations \(OBI\) material entity \(BFO:0000040\) ontology](#), e.g. "specimen from organism" (OBI:0001479), "xenograft" (OBI:0100058), "cell culture" (OBI:0001876), "cell specimen" (OBI:0001468), "environmental swab specimen" (OBI:0002613)

**sampleOriginDetail** CATEGORICAL VALUE (ONTOLOGY LABEL) Specific instance of sample origin matching the category set in **sampleOriginType**. Value from [Uber-anatomy ontology \(UBERON\)](#) or [BRENDA tissue / enzyme source \(BTO\)](#), [Ontology for Biomedical Investigations \(OBI\)](#) or [Cell Line Ontology \(CLO\)](#), e.g. "cerebellar vermis" (UBERON:0004720), "HEK-293T cell" (BTO:0002181), "nasopharyngeal swab specimen" (OBI:0002606), "cerebrospinal fluid specimen" (OBI:0002502)

**obtentionProcedure** CATEGORICAL VALUE (ONTOLOGY LABEL) Ontology ID from [NCI Thesaurus \(NCIT\) Intervention or Procedure \(NCIT:C25218\) ontology](#). e.g. "biopsy" (NCIT:C15189)

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

**tumorProgression** CATEGORICAL VALUE (ONTOLOGY LABEL). Descriptor of tumor progression. Value 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 ID) from [Tumor Grading Characteristic ontology \(Mondo Disease Ontology MONDO:0024488\)](#) General tumor grading

## Run

**runId** ALPHANUMERIC VALUE Run reference ID (external accession or internal ID)

**biosampleId** ALPHANUMERIC VALUE Reference to sample ID

**runDate** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Date at which run was performed

**librarySource** CATEGORICAL VALUE Sequencing library source, e.g. "Metagenomic", "Viral RNA"

**libraryStrategy** CATEGORICAL VALUE Sequencing library strategy, e.g. "WGS"

**librarySelection** CATEGORICAL VALUE Selection method for sequencing library preparation, e.g. "RANDOM", "RT-PCR"

**libraryLayout** CATEGORICAL VALUE Sequencing library layout, e.g. "PAIRED", "SINGLE"

**platform** CATEGORICAL VALUE Sequencing technology, e.g. "Illumina", "Oxford Nanopore Technologies"

**platformModel** CATEGORICAL VALUE Sequencing platform model, e.g. "Illumina MiSeq", "GridION"

## Analysis

**analysisId** ALPHANUMERIC VALUE Analysis reference ID (external accession or internal ID)

**runId** ALPHANUMERIC VALUE Reference to run ID

**analysisDate** ALPHANUMERIC VALUE ([ISO8601 DURATION FORMAT](#)) Date at which analysis was performed

**pipelineName** CATEGORICAL VALUE Analysis pipeline and version if a standardized pipeline was used

**pipelineRef** Link to Analysis pipeline resource

**aligner** CATEGORICAL VALUE Mapping/Alignment software e.g. "bwa-0.7.8"

**variantCaller** CATEGORICAL VALUE Variant calling software/ pipeline, e.g. "GATK4.0"

## Variant in Sample

**variantId** ALPHANUMERIC VALUE Reference to Variant ID

**analysisId** Reference to analysis ID

**biosampleId** Reference to biosample ID

**variantFrequency** NUMERIC VALUE Variant frequency in biosample

**zigosity** CATEGORICAL VALUE (ONTOLOGY LABEL) Zigosity in which variant is present in the sample. Value from the [Zigosity Ontology \(GENO:0000133\)](#), e.g. "heterozygous" (GENO:0000135)

**alleleOrigin** CATEGORICAL VALUE (ONTOLOGY LABEL) Allele origin of variant in sample. Value from the [Variant Origin \(SO:0001762\)](#), e.g. "somatic variant", "germline variant", "de novo variant".

**phenotypicEffects** Observed effect of variant on phenotype. (List of:)

**phenotypeId** CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of phenotype found associated to variant in the present study. Value from [Human Phenotype Ontology \(HPO\)](#)

**phenotypeEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic effect classification determined in the present study. Value from [Sequence types and features ontology \(SO\) variant phenotype \(SO:0001769\)](#), e.g. "quantitative variant" (SO:0001774)

**evidenceType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of evidence supporting variant-phenotype association from the [Evidence & Conclusion Ontology \(ECO\)](#), e.g. "experimental evidence"

**clinicalRelevances** Observed effect of variant on disease. (List of:)

**diseaseId** CATEGORICAL VALUE (DISEASE CODE/ ONTOLOGY LABEL) Descriptor of disease associated. Value from [ICD10 disease codes](#) or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO, e.g. "lactose intolerance" (HP:0004789, ICD10CM:E73)

**clinicalEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Clinical effect classification. Value from [Sequence types and features ontology \(SO\) variant phenotype \(SO:0001769\)](#), e.g. "disease causing variant" (SO:0001772)

**evidenceType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of evidence supporting variant-disease association from the [Evidence & Conclusion Ontology \(ECO\)](#), e.g. "experimental evidence"

## Variant Interpretation

**variantId** ALPHANUMERIC VALUE Reference to Variant ID

**datasetId** ALPHANUMERIC VALUE Reference to dataset ID source of variant interpretation

**phenotypicEffects** Annotated effect(s) on any phenotypic feature other than a disease. (List of:)

**phenotypeId** CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of phenotype associated. Value from [Human Phenotype Ontology \(HPO\)](#)

**phenotypeEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic effect classification. Value from [Sequence types and features ontology \(SO\) variant phenotype \(SO:0001769\)](#), e.g. "benign variant" (SO:0001770)

**alleleOrigin** (LIST OF) CATEGORICAL VALUE(S) (ONTOLOGY LABEL) Annotation(s) on allele origins in which the variant has been found in association to phenotype. Value(s) from [Sequence types and features ontology \(SO\) variant origin \(SO:0001762\) ontology](#), e.g. "somatic variant" (SO:0001777), "germline variant" (SO:0001778), "de novo variant" (SO:0001781)

**references** (LIST OF) PUBLICATION IDENTIFIER(S) Reference(s) to publication(s) describing association, as PMID(s), e.g. "PMID:17024373"

**clinicalRelevances** Annotated effect on disease. (List of:)

**diseaseId** CATEGORICAL VALUE (DISEASE CODE/ ONTOLOGY LABEL) Descriptor of disease associated. Value from **ICD10 disease codes** or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO. e.g. "lactose intolerance" (HP:0004789, ICD10CM:E73)

**clinicalEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Clinical effect classification. Value from **Sequence types and features ontology (SO) variant phenotype (SO:0001769)**, e.g. "disease associated variant" (SO:0001771)

**alleleOrigin** (LIST OF) CATEGORICAL VALUE(S) (ONTOLOGY LABEL) Annotation(s) on allele origins in which the variant has been in association to condition. Value(s) from **Sequence types and features ontology (SO) variant origin (SO:0001762) ontology**, e.g. "somatic variant" (SO:0001777), "germline variant" (SO:0001778), "de novo variant" (SO:0001781)

**references** (LIST OF) PUBLICATION IDENTIFIER(S) Reference(s) to publication(s) describing association, as PMID(s), e.g. "PMID:19793655"

## Interactor

This is an organism/agent whose metadata/ phenotypic data is collected in association with the Subject, but which is not sequenced itself. It accounts for 'extended phenotype' of variants in other organisms/agents than the one harboring them.

**relationType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of relation with Subject. Value from **Infectious disease Ontology (IDO)**, e.g. "host" (IDO:0000531), "commensal" (IDO:0000525), "infectious agent" (IDO:0000596)

[... ] All the rest of objects from Subject