CERNER CORPORATION©

Clinical Bioinformatics Ontology™

This paper is a review of the architecture, standards, and scope of the Clinical Bioinformatics Ontology (CBO).

Table of Contents

Introduction

Purpose

Existing medical terminologies do not sufficiently support the description of molecular diagnostic findings due to significant gaps and inaccuracies. The CBO addresses the need for consistent descriptions and representations of molecular entities currently performed in clinical diagnostics in a standardized and machine readable format. Other notable ontologies of molecular biological interest are oriented toward the research setting and focus primarily on functional attributes, which do not meet the requirements of clinical use. The CBO focuses on this gap to meet the needs of clinical molecular diagnostics and cytogenetics.

The CBO is a semantically structured controlled vocabulary for clinical molecular diagnostics. It addresses the need for consistent representation of clinically relevant molecular biological and cytogenetic entities in a standardized and machine readable format. It combines the attributes of a controlled vocabulary by providing standardized naming conventions and coded values, with the characteristics of an ontology. Structuring the CBO in an ontology format provides a platform for advanced queries, inference logic, correlation, making assertions, and managing the complex data of current clinical practice.

Scope

Targets of current diagnostic analysis are considered clinically significant and are included in the CBO. However, the CBO also includes place holders for some domains primarily of research interest that are not currently a part of clinical diagnostic practice, which are denoted by asterisk symbols. Many clinically important concepts have not yet been incorporated in the CBO, future release will continue to broaden the depth of the content.

Methodology Overview

Content that meets the criteria for inclusion in the CBO is curated using a variety of resources, including public web sites and scientific literature. The CBO naming conventions utilize existing standards when possible and alternative methods when necessary to ensure consistently reproducible and traceable content. Quality assurance processes are used to verify new concepts prior to inclusion in the CBO. The standardized content is represented in codified values and structured in the CBO ontology.

2 Terminology Overview

Concept – A unique entity of knowledge.

Clinical Bioinformatics – The integration of genomic information and informatics for application in the clinical community.

Defining Relationship – Relationships that horizontally span hierarchies to create a semantic network of related concepts, a characteristic that differentiates an ontology from a controlled vocabulary.

Facet – An attribute associated to a concept to provide additional information and context.

Facet Definitions – Attributes of facets that define the instances and manner in which a facet is to be used.

Hierarchy – A grouping of concepts that maintains the parent-child relationship.

Ontology - A semantically structured controlled vocabulary.

Parent-Child Relationship – A relationship used in hierarchies to associate parent and child concepts. This relationship is also known as "Is A" and "Subsumes" relationship.

Relationship Definitions – Attributes of relationships that define the instances and manner in which a relationship is to be used.

Term – An attribute associated to a concept that encompasses common language, such as synonyms and abbreviations.

Terminology Axis – A defined level of granularity. Below a terminology axis, a true hierarchy exists.

3 Navigational Concepts

Navigational concepts are the highest levels of the ontology. They provide organization to concept hierarchies. At this level of the ontology, the true "IS A" relationship is not necessarily maintained until a terminology axis is reached. The highest level of navigational concepts in the CBO distinguishes between the human biology and nonhuman biology navigational concepts.

Human Biology

3.1. Human Proteome

- This refers to the protein products of genes.
- Includes concepts:

Human Protein

3.2. Human Transcriptome

- This refers to the transcripts derived from genes.
- Includes concepts:

Human cDNA

Human RNA
- Human mRNA
- Human tRNA
- Human rRNA
- [*] Human snRNA
- [*] Human SRP RNA

3.3. Human Variation

- This refers to variations within the human genome and proteome associated with a known clinical effect.
- Includes concepts:

Human Chromosomal Variation
- Human Chromosomal Aneuploidy
- Human Chromosomal Deletion
- Human Chromosomal Duplication
- Human Chromosomal Ring
- Human Chromosomal Marker
- Human Isochromosome
- Human Sex Determinant
- Human Chromosomal Rearrangement
  - Human Chromosomal Translocation
  - Human Chromosomal Inversion
  - Human Chromosomal Insertion

Human Amino Acid Variant

Human Nucleic Acid Variant
- Human Nucleotide Variant
- Human Allele
- [*] Human Haplotype
- Human Nucleotide Repeat

Human Karyotype
Human Uniparental Disomy

3.4. Human Genome

- This refers to structural features within the physical human genome.
- Includes concepts:

Human Locus

- This refers to a discrete structural position in the genome map.
- Includes concepts:

Human Repeat Markers
VNTR
STR
Human Gene
***Human Alu**

Human Chromosome

Human Cytogenetic Features

- This refers to chromosomal structures within a human chromosome.
- Includes concepts:

Human Chromosomal Arm
Human Chromosomal Arm p
Human Chromosomal Arm q
Human Centromere

Human Chromosomal Band
Human Chromosomal Band – 400 Res
Human Chromosomal Band – 550 Res
Human Chromosomal Band – 850 Res
Human Telomere
Human Telomere p
Human Telomere q

Human Sequence Features

- This refers to structural features within a gene sequence.
- Includes concepts:

Human Exon
Human Intron
Human 5-UTR
Human 3-UTR

Human Extrachromosomal Element

- This refers to human genomic elements not included within chromosomes.
- Includes concepts:

Human mtDNA

Human Epigenetic Status

- This refers to variation within an individual's genome not determined by Mendelian inheritance.
- Includes concepts:

Human Methylation Status

3.5. Descriptors

- This refers to classification systems to describe concepts within the Human Biology and Nonhuman Biology models.
- Includes concepts:

Mode of Inheritance
Autosomal
X-linked
Y-linked
Mitochondrial
[*] Gene Ontology (GO) ™ [6]
Cytogenetic Descriptor
Derivative Chromosome
Dicentric Chromosome
Constitutional Karyotype
Malignant Karyotype
Constitutional Mosaic Karyotype

Nonhuman Biology

3.6. Nonhuman Proteome

- This refers to the protein product of genes.
- Includes concepts:

Nonhuman Protein

3.7. Nonhuman Transcriptome

- This refers to the transcripts derived from genes.
- Includes concepts:

Nonhuman cDNA

Nonhuman RNA
Nonhuman mRNA
Nonhuman tRNA
Nonhuman rRNA

3.8. Nonhuman Variation

- This refers to variations within nonhuman genomes associated with a known clinical effect.
- Includes concepts:

Nonhuman Amino Acid Variant

Nonhuman Nucleic Acid Variant
Nonhuman Nucleotide Variant
Nonhuman Allele

3.9. Nonhuman Genome

- This refers to structural features within physical nonhuman genomes.
- Includes concepts:

Nonhuman Locus

- This refers to a discrete structural position in the genome map.
- Includes concepts:

Nonhuman Gene

Nonhuman Chromosome

Nonhuman Sequence Features

- This refers to structural features within a gene sequence.
- Includes concepts:

Nonhuman Exon
Nonhuman Intron

Nonhuman Extrachromosomal Element

- This refers to nonhuman genomic elements not included within chromosomes.

Bacterial Genome

- This refers to the complete genome of a bacterium.

Viral Genome

- This refers to the complete genome of a virus.
- Includes concepts:

ssRNA
dsRNA
ssDNA
dsDNA

Protozoan Genome

- This refers to the complete genome of a protozoan.

Fungal Genome

- This refers to the complete genome of a fungus.

4 Terminology Axes and Standards

Human Biology

Terminology Axes

A terminology axis defines the level of granularity in which a hierarchy, or grouping of related concepts, exists within an ontology. Concepts within a hierarchy maintain the true "is a" parent – child relational logic. Terminology axes allow definitional relationships to span multiple hierarchies, though a concept in one terminology axis can be associated to only one vertical level of another terminology axis to maintain relational logic integrity.

Human Proteome

Standards

The CBO has utilized existing standards when possible and alternative naming conventions when necessary. All concepts comply with the CBO standards, though terms provide the means to associate alternative naming conventions, synonyms, and common terminology to the primary concept. A complete listing of references is provided.

Human Transcriptome

Below is a listing of the CBO terminology axes and standards defined for the concepts of each axis.

4.1. Human Protein

- The Human Protein hierarchy includes protein concepts which are documented with the approved protein symbol. CBO standard: The naming convention for protein transcripts is the (AA) prefix, period, approved protein symbol, period, and alternative isoform number.
- Example: (AA).CFTR.0

4.2. Human RNA

- The Human RNA hierarchy includes RNA transcript type concepts: Human mRNA, Human SRP RNA, Human snRNA, Human tRNA, and Human rRNA.
- CBO standard: The naming convention for transcript concepts is the transcript type prefix [(mRNA), (srpRNA), (snRNA), (tRNA), and (rRNA)], period, approved gene symbol, transcript number.
- Example: (mRNA).CFTR.0

4.3. Human cDNA

- The Human cDNA hierarchy includes cDNA transcript concepts which are documented with the approved gene symbol. CBO standard: The naming convention for cDNA transcript concepts is the (cDNA) prefix, period, approved gene symbol, transcript number.
- Example: (cDNA).CFTR.0

The CBO differentiates between transcript types by applying a naming convention consistent with the HUGO Guidelines for Human Gene Nomenclature. When multiple RefSeqs exist, alternative transcripts are included in the CBO naming convention by sequential numerical characters corresponding with the appropriate protein, RNA, and cDNA transcripts. The reference transcript, in which variations are positionally referenced, is designated by the transcript number "0". The full length transcript is the designated reference transcript. In some cases a designated full length versus alternative shorter length transcripts does not exist, therefore, the reference transcript will be designated based on the correct sequential occurrence of exons. For example, two transcripts contain an equal number of exons,

however, the first transcript does not contain exon 3 and the second transcript does not contain exon 2. The first transcript is designated the reference transcript.

Table 1 - CBO molecule prefixes. See Human Gene Nomenclature Guidelines for further descriptions.

Gene Symbol Prefix	Description
(mRNA).CFTR	mRNA molecule
(gDNA).CFTR	Genomic molecule
(cDNA).CFTR	cDNA molecule
(AA).CFTR	Amino acid molecule
(tRNA).TRR	tRNA molecule
(rRNA).RNR1	rRNA molecule
(snRNA).BAX	snRNA molecule
(srpRNA).PAPOLG	SRP RNA molecule

4.4. Human Amino Acid Variant

- The Human Amino Acid Variant hierarchy includes concepts of variations within amino acid precursor transcripts. The naming convention for amino acid variants incorporates the Mutation Nomenclature Suggestions [1].
- CBO standard: The naming convention for amino acid variation concepts is the approved protein name and appropriate variant description. See Table 2 below.
- Example: CFTR.p.R117H

4.5. Human Nucleic Acid Variant

- The Human Nucleotide Variant hierarchy includes concepts of variations within genomic and cDNA sequences. The naming convention for nucleotide variations incorporates the Mutation Nomenclature Suggestions. Variations within the coding sequence are named and referenced according to the cDNA Reference Sequence (RefSeq), in which the position numbering is relative to the first nucleotide in the first codon. Variations outside of the coding sequence are named and referenced to the full length genomic sequence when possible. Often a genomic RefSeq does not exist for the gene, in which case intronic variations are named and referenced according to the cDNA RefSeq. All nucleotide variations are referenced to the primary RefSeq. The variation reference is relative to the "GI" number of the RefSeq documented at the time. When a genomic RefSeq exists, all mutations, intronic and coding, are named using positions relative to the genomic sequence. If the genomic RefSeq cannot clearly be reconciled with positions of previously adopted

mutation naming conventions, then an alternative genomic sequence will be referenced and the gi value will be documented, Documentation of up to ten nucleotides involved in the variation is permitted. Nucleotides are not documented if a deletion involves more than ten nucleotides. For example: 13_25del

- CBO standard: The naming convention for nucleotide variation concepts is the approved gene name and appropriate variation description. See Table 2 below.
- Example: CFTR.c.482G>A, CFTR.c.1318+3C>G
- For variants that involve both deletion and insertion of nucleotides, the deletion and insertion shall be described in detail up to 15 nucleotides.
- Example: BLM.c.2207_2212delATCTGAinsTAGATTC
    - In cases where the number of nucleotides exceeds 15, the number of nucleotides should be indicated, for example: BLM.c.2207_2212del[6bp]ins[7bp]

Table 2 – The CBO naming conventions for common sequence variation types. See Human Mutation Suggestions for further descriptions.

Type of Variation	**Description**	**Comments**
Substitution (Sequence description applies to all variation types)	g.12T>A	Change in reference to genomic sequence
	c.12T>A	Change in reference to cDNA sequence
	m.12T>A	Change in reference to mitochondrial sequence
	r.12t>a	Change in reference to RNA sequence
	p.T12A	Change in reference to amino acid sequence
	93-2A>G	Change in intronic sequence - splice acceptor site from the exon 93 nucleotide position in the cDNA transcript.
	15+1G>C	Change in intronic sequence - splice donor site from the exon 15 nucleotide position
Deletion	13_14delTT	Deletion of two nucleotides at positions 13 and 14.
	IVS1_IVS5	
	p.L3_H4del	Deletion of two amino acids at protein positions 3 and 4.
Duplication	10_11dupTG	

	p.L3_H4dup	
Insertion	14_15insT	Insertion of a "T" between nucleotides 14 and 15.
	p.W4_R5insK	Insertion of an amino acid (K) between position 4 and 5.
Frame shift	p.W4fsX	Frame shift causing a translational stop downstream. The codon in which the stop occurs is not documented.

4.6. Human Allele

- The Human Allele hierarchy includes concepts of alleles, or a unique combination of nucleotide variations, for a given gene. Currently, generalized standards for allele naming conventions do not exist; therefore CBO has proposed a new convention for alleles. Though the CBO allele naming convention resembles the OMIM™ allele naming system, CBO alleles do not match the OMIM allele system. OMIM and locus specific allele naming systems can be captured in terms to be incorporated in CBO.
- CBO standard: The naming convention for alleles is the approved gene name and a sequential numerical character beginning with 0001.
- Example: CFTR.0001

4.7. Human Nucleotide Repeat

- The Human Nucleotide Repeat hierarchy includes concepts of nucleotide repeats.
- CBO standard: The naming convention for nucleotide repeats is the approved gene name, the transcript type that is positionally referenced, the first nucleotide position of the first repeat, the last nucleotide position of the first repeat, the repeated sequence, and "n" to indicate the variable number of repeats.
- Example: HD.c.52_54(CAG)n

4.8. [*] Human Haplotype

- The Human Haplotype hierarchy includes concepts of haplotypes, or a unique combination of alleles.

4.9. Human Chromosomal Aneuploidy

- The Human Chromosomal Aneuploidy hierarchy includes concepts referring to variations in total chromosome number.
- CBO standard: The additional or missing chromosome is documented
- Example: +21

4.10 Human Chromosomal Deletion

- The Human Chromosomal Deletion hierarchy includes concepts referring to deletions within chromosomes.
- CBO standard: The ISCN is used as the naming convention.

- Example: del(15)(q11q13)

4.11 Human Chromosomal Duplication

- The Human Chromosomal Duplication hierarchy includes concepts referring to duplications within chromosomes.
- CBO standard: The ISCN is used as the naming convention
- Example: dup(1)(q22q25)

4.12 Human Chromosomal Ring

- The Human Chromosomal Ring hierarchy includes concepts referring to chromosomes in ring formation
- CBO standard: The ISCN is used as the naming convention
- Example: r(7)(p22q36)

4.13 Human Chromosomal Marker

- The Human Chromosomal Marker hierarchy includes concepts referring to unknown chromosomal material contained in markers
- CBO standard: The presence of the marker is documented
- Example: +mar

4.14 Human Isochromosome

- The Human Isochromosome hierarchy includes concepts referring to chromosomes with two identical arms.
- CBO standard: The ISCN is used as the naming convention
- Example: i(X)(q10)

4.15 Human Sex Determinant

- The Human Sex Determinant hierarchy includes concepts referring to sex chromosome complement
- CBO standard: The common sex chromosome complements are used as the naming convention
- Example: XY

4.16 Human Chromosomal Translocation

- The Human Chromosomal Translocation hierarchy includes concepts referring to translocations within chromosomes, at the chromosome, gene, and exon level. The level of granularity, either chromosomal, gene, or exon, is modeled through parent-child relationships. For example: Human Chromosomal Translocation {*SUBSUMES*} t(9;22)(q34;q11) {*SUBSUMES*} t(BCR;ABL1) {*SUBSUMES*} t(BCR.e.1;ABL1.e.2)
- CBO standard:
  Chromosomal translocation – The ISCN is used as the naming convention
  Translocation described at gene level – t(GeneSymbol1;GeneSymbol2)
  Translocation described at exon level – t(ExonName1;ExonName2)
- Example:
  Chromosomal translocation - t(9;22)(q34;q11)
  Translocation described at gene level - t(BCR;ABL1)
  Translocation described at exon level - t(BCR.e.1;ABL1.e.2)

4.17 Human Chromosomal Inversion

- The Human Chromosomal Inversion hierarchy includes concepts describing chromosomal inversions.

- CBO standard: The ISCN is used as the naming convention.
- Example: inv(3)(q21q26)

4.18 Human Chromosomal Insertion

- The Human Chromosomal Insertion hierarchy includes concepts describing chromosomal insertions.
- CBO standard: The ISCN is used as the naming convention.
- Example: ins(5;11)(q31;q13q23)

4.19 Human Chromosome

Human Locus

- The Human Chromosome hierarchy includes all twenty-four chromosome concepts.
- CBO standard: No abbreviations are used in the naming convention.
- Example: Chromosome 1

4.20 [*]Human mtDNA

- The Human mtDNA hierarchy includes the mitochondrial genome concepts.

4.21 Human Methylation Status

- The Human Methylation Status hierarchy includes the gene specific methylation status concepts.
- CBO standard: The approved gene name and methylation
- Example: FMR1 Methylation

4.22 [*] Human Alu

- The Human Alu hierarchy will include alu sequence concepts.

4.23 Human Repeat Markers

Human Cytogenetic Features

- The Human Repeat Markers hierarchy includes STR and VNTR sequence marker concepts.
- CBO standard: The marker ID will be used as the naming convention.
- Example: (gDNA).STR.TPOX

4.24 Human Gene

- The Human Gene hierarchy includes human gene concepts which are documented with the Human Gene Nomenclature Committee approved gene symbol. Full length gene names, alias gene names, alias gene symbols will be captured in terms.
- CBO standard: A (gDNA) prefix followed by a period will precede the gene symbol to distinguish between molecule types, as recommended by the Guidelines for Human Gene Nomenclature.
- Example: (gDNA).CFTR

4.25 Human Chromosomal Arm p

Human Genome

- The Human Chromosomal Arm p hierarchy includes all twenty-four chromosomal short arms (p), documented according to the ISCN standards.
- CBO standard: The chromosome number is followed by the abbreviation for short arm, p.
- Example: 7p

4.26 Human Chromosomal Arm q

- The Human Chromosomal Arm q hierarchy includes all twenty-four chromosomal short arms (q), documented according to the ISCN standards.
- CBO standard: The chromosome number is followed by the abbreviation for short arm, q.
- Example: 7q

4.27 Human Chromosomal Band - 400 Res

- The Human Chromosomal Band – 400 Res hierarchy includes all chromosomal bands visible at the 400-resolution level, documented according to the ISCN standards.
- CBO standard: To distinguish between resolution levels CBO includes a "400" prefix, followed by a period, chromosome number, chromosomal arm, band number.
- Example: 400.5q11.1

4.28 Human Chromosomal Band - 550 Res

- The Human Chromosomal Band – 550 Res hierarchy includes all chromosomal bands visible at the 550-resolution level, documented according to the ISCN standards.
- CBO standard: To distinguish between resolution levels CBO includes a "550" prefix, followed by a period, chromosome number, chromosomal arm, band number.
- Example: 550.5q11.1

4.29 Human Chromosomal Band - 850 Res

- The Human Chromosomal Band – 850 Res hierarchy includes all chromosomal bands visible at the 850-resolution level, documented according to the ISCN standards.
- CBO standard: To distinguish between resolution levels CBO includes an "850" prefix, followed by a period, chromosome number, chromosomal arm, band number.
- Example: 850.5q11.1

4.30 Human Tel p

- The Human Tel p hierarchy includes all twenty-four chromosomal telomeres for short arm p, documented according to the ISCN standards.
- CBO standard: The naming convention is the chromosome number followed by the abbreviation for short arm, p, and the abbreviation for telomere, tel.
- Example: 7ptel

4.31 Human Tel q

- The Human Tel q hierarchy includes all twenty-four chromosomal telomeres for short arm q, documented according to the ISCN standards.
- CBO standard: The naming convention is the chromosome number followed by the abbreviation for short arm, q, and the abbreviation for telomere, tel.
- Example: 7qtel

4.32 Human Centromere

- The Human Centromere hierarchy includes all twenty-four chromosomal centromeres, documented according to the ISCN standards.
- CBO standard: The naming convention is the chromosome number followed by the abbreviation for centromere, cen.
- Example: 7cen

4.33 Human 5-UTR

- The Human 5-UTR hierarchy includes concepts of 5 prime structures within nucleotide sequences.
- CBO standard: The approved gene name, period, 5-UTR
- Example: FMR1.5-UTR

4.34 Human 3-UTR

- The Human 3-UTR hierarchy includes concepts of 3 prime structures within nucleotide sequences.
- CBO standard: The approved gene name, period, 3-UTR
- Example: FY.3-UTR

4.35 Human Exon

Descriptors

- The Human Exon hierarchy includes concepts of exon structures within nucleotide sequences. The CBO assigns a naming convention for all exons. Exons are numbered sequentially relative to the RefSeq. CBO does not incorporate alpha-numeric exon numbering systems, such as 6a, 6b, 6c. The CBO determines the exon count from the primary RefSeq using UCSC RefSeq Tables (4).
- CBO standard: The approved gene symbol, period, e, period, exon number.
- Example: CFTR.e.4

4.36 Human Intron

- The Human Intron hierarchy includes concepts of intron structures within nucleotide sequences. The CBO assigns a naming convention for all introns. Introns are numbered sequentially relative to the RefSeq.
- CBO standard: The approved gene symbol, period, i, period, intron number.
- Example: CFTR.i.8

4.36 [*] Gene Ontology

- The Gene Ontology (GO) hierarchy is an incorporated controlled terminology and classification system for the molecular function of gene products.

4.37 Mode of Inheritance

- The Mode of Inheritance hierarchy includes X-linked, Y-linked, mitochondrial, and autosomal modes of inheritance concepts. Genes are related to these concepts.

Human Sequence Features

4.38 **Cytogenetic Descriptors**

- The Cytogenetic Descriptors hierarchy includes derivative and dicentric chromosome concepts; constitutional, malignant, and constitutional mosaic karyotype concepts; and chromosomal length variation concepts. Chromosome variations are related to these concepts.

Nonhuman Biology

Note: For nonhuman concepts within the CBO, the species of origin is specified for each concept. Eventually, we plan to make one clean linkage to SNOMED at the viral genome concept. We also specify the transcript type for each concept or gene. Capitalization is determined by the generally accepted standard, which is usually reflected in Entrez Gene.

4.39. Nonhuman Protein

- The Nonhuman Protein hierarchy includes precursor and mature protein concepts which are documented with the Entrez Gene adopted protein symbol.
- CBO standard: The naming convention for protein transcripts is adopted pathogen name, period, (AA) prefix, period, adotpted protein symbol
- Example: HIV-1.(AA).gag

Human Transcriptome

4.40. Nonhuman RNA

- The Nonhuman RNA hierarchy includes RNA transcript type concepts: Nonhuman mRNA, Nonhuman tRNA, and Nonhuman rRNA.
- CBO standard: The naming convention for transcript concepts is the adopted pathogen name, period, transcript type prefix [(mRNA), (tRNA), and (rRNA)], period, adopted gene symbol
- Example: HIV-1.(mRNA).env

4.41. Nonhuman cDNA

- The Nonhuman cDNA hierarchy includes cDNA transcript concepts which are documented with the adopted gene symbol.
- CBO standard: The naming convention for cDNA transcript concepts is the adopted pathogen name, period, (cDNA) prefix, period, adopted gene symbol
- Example: HIV-1.(cDNA).gag-pol

4.42. Nonhuman Amino Acid Variant

- The Nonhuman Amino Acid Variant hierarchy includes concepts of variations within amino acid precursor or mature protein transcripts depending upon the publicly adopted standard of documentation. A mature protein (mp) prefix is used to differentiate between the two transcript types. The naming convention for amino acid variants incorporates the Mutation Nomenclature Suggestions [1].
- CBO standard: The naming convention for amino acid variation concepts is the adopted pathogen name, period, adopted protein name, transcript type [mature protein (mp) or precursor protein (p)], period, and appropriate variant description.
- Example: HIV-1.pro.mp.K20R

4.43. Nonhuman Nucleic Acid Variant

- The Nonhuman Nucleotide Variant hierarchy includes concepts of variations within genomic sequences. The naming convention for nucleotide variations incorporates the Mutation Nomenclature Suggestions. All nucleotide variations are referenced to the primary RefSeq. The variation reference is relative to the "GI" number of the RefSeq documented at that time. All nonhuman mutations, intronic and coding are named relative to the genomic sequence. If the genomic RefSeq is not easily reconciled with positions of previously adopted mutation naming conventions, then an alternative genomic sequence will be referenced and the gi value will be documented, Documentation of up to ten nucleotides involved in the variation is permitted.

Nucleotides are not documented in the case that a variation ivolves an excess of ten nucleotides. For example: 13_25del

- CBO standard: The naming convention for nucleotide variation concepts is the adopted pathogen name, period, appropriate variation description. See Table 2 above.
- Example: HIV-1.c.**1857A>G Note: In this case a "c" is used instead of a "g" to designate transcript type because HIV is an RNA virus and mutations are documented from reverse transcription.**

4.44. [*] Nonhuman Chromosome

4.45. Nonhuman Gene

- The Nonhuman Gene hierarchy includes nonhuman gene concepts which are documented with the Entrez Gene adopted gene symbol. Full length gene names, alias gene names, alias gene symbols will be captured in terms.
- CBO standard: The adopted pathogen name, pereiod, (gDNA) prefix, period will precede the gene symbol to distinguish between molecule types.
- Example: HIV-1.(gDNA).gag

4.46. [*} Nonhuman Exon

4.47. [*] Nonhuman Intron

4.48. [*] Nonhuman Extrachromosomal Element

4.49. Bacterial Genome

- **The Bacterial Genome hierarchy includes concepts of bacterial genomes. The adopted naming convention is the abbreviated formal name followed by the transcript type.**
- **Example: M. bovis.(gDNA)**

4.50. Viral Genome

- **The Viral Genome hierarchy includes all types of viral genomes, including ssRNA, dsRNA, ssDNA, and dsDNA. The naming convention is the adopted virus name followed by the transcript type.**
- **Example: HIV-1.(ssRNA)**

5 Terms

Terms are attributes of concepts used to incorporate commonly used terminology or alternative naming conventions into the CBO.

5.1 Gene Concept Terms

- HUGO approved symbol
- Alias symbols
- HUGO approved gene name
- Alias gene names

5.2. Allele Concept Terms

- Locus specific naming conventions

5.3. Protein Concept Terms

- Approved symbol
- Approved protein name
- Alias protein names
- Alias symbols

5.4. Amino Acid Variant Terms

- Variation relative to mature protein

5.5. Nucleotide Variant Terms

- Widely adopted alternative naming convention

6 Facets

Facets are attributes of concepts used to incorporate external reference values for further traceability and additional information. Facet definitions are characteristics of facets that specify value restrictions or instances in which a facet may be used. See Table 3 below.

6.1. Gene Concept Facets

- OMIM ID number
- Entrez Gene ID
- Genomic RefSeq accession number
- Genomic RefSeq GI number at the time of documentation

6.2. cDNA Concept Facets

- cDNA [referred to mRNA in GenBank] RefSeq accession number
- cDNA [referred to mRNA in GenBank] RefSeq GI number at the time of documentation
- cDNA exon start position
- cDNA exon end position

6.3. Protein Concept Facets

- Amino acid RefSeq accession number
- Amino acid RefSeq GI number at the time of documentation

6.4. Chromosome Band Concepts Facets

- Chromosome Band Order

Table 3 - CBO facet definintions.

Name	Nickname	Type Restriction	Cardinality	Displayable	Inheritable	Immunitable	Applicability
Genomic RefSeq accession number	GEN_REFSEQ	string	multiple	yes	no	yes	concept
Genomic RefSeq GI number	GEN_GI	string	multiple	yes	no	yes	concept

Amino acid RefSeq Accession	AA_REFSEQ	string	multiple	yes	no	yes	concept
Amino acid GI number	AA_GI	string	multiple	yes	no	yes	concept
cDNA RefSeq Accession	CDNA_REFSEQ	string	multiple	yes	no	yes	concept
cDNA GI	CDNA_GI	string	multiple	yes	no	yes	concept
OMIM ID number	OMIM_ID	string	multiple	yes	no	yes	concept
Entrez Gene ID	CONCEPT_ID	string	multiple	yes	no	yes	Concept
CDS Exon Start Position	CDS_EXON_SRT	string	multiple	yes	no	yes	concept
CDS Exon End Position	CDS_EXON_END	string	multiple	yes	no	yes	concept
Chromosome Band Order	CHRM_BND_ORD	string	multiple	yes	no	yes	Concept

7 Relationship Types

A relationship associates two concepts. There are two types of relationships in the CBO: vertical parent-child relationships and horizontal defining relationships. Concepts may have multiple relationships. Defining relationships span multiple hierarchies to create a semantic network. The directionality and specificity of defining relationships are essential in defining concepts and maintaining acceptable performance of an ontology. Relationship definitions are attributes of relationships that specify instances of relationship use (Table 4.). The CBO retains biological accuracy and optimizes performance by combining the characteristics and attributes of defining relationships. Most relationships may be used for both human and nonhuman hierarchies, those relationships that can only be used for the nonhuman hierarchy are specified.

7.1. Subsumes

- This relationship refers to the universal "is a" relationship and is used in all parent-child hierarchies.
- Example: (gDNA).CFTR {*SUBSUMES}* Human Gene

7.2. Has Chromosomal Location

- Indicates chromosome location, used to relate locus concepts to chromosome concepts.
- Example: (gDNA).CFTR {*HAS_CHRM_LOC}* Chromosome 7

7.56. Arm of

- Indicates chromosomal arms, used to relate chromosomal arm concepts to chromosome concepts.
- Example: 7q {*ARM_OF}* Chromosome 7

7.57. Band of

- Indicates cytogenetic bands, used to interrelate band concepts, as well as to relate chromosomal arm concepts to chromosomal band concepts.
- Example:

  400.5q13 {*BAND_OF}* 5q

550.5q13.3 {*BAND_OF}* 400.5q13
850.2q32.2 {*BAND_OF}* 550.2q32.2

7.58. Has Arm Location

- Indicates chromosomal arm location, used to relate locus concepts to chromosomal arm concepts.
- Example: (gDNA).CFTR {*HAS_ARM_LOC}* 7q

7.59. Has Band Location

- Indicates chromosomal band location, used to relate locus concepts to chromosomal band concepts.
- Locus concepts that have multiple band locations are related to all band locations.
- Locus concepts must be related to the lowest applicable resolution, this is the 850 resolution. This is true even if the equivalent band exists at a higher resolution.
- Example: (gDNA).CFTR {*HAS_BAND_LOC}* 850.7q31.2

7.60. Exon of

- Indicates exon structure origin, used to relate exon concepts to gene concepts. Exons are related only to the genomic concept, thus excluding the relationship from cDNA and mRNA concepts.
- Example: CFTR.e.4 {*EXON_OF}* (gDNA).CFTR

7.61. Intron of

- Indicates intron structure origin, used to relate intron concepts to genomic concepts.
- Example: CFTR.i.6 {*INTRON_OF}* (gDNA).CFTR

7.62. Allele of

- Indicates a unique combination of nucleotide variants occurring within a gene, used to relate allele concepts to gene concepts.
- Example: CFTR.0001 {*ALLELE_OF}* (gDNA).CFTR

7.63. Has Constituent Variant

- Indicates the combination of nucleotide variants or single nucleotide variant that compose an allele, used to relate allele concepts to nucleotide variant concepts.
- Example: CFTR.0001 {*HAS_CONS_VAR}* CFTR.c.482G>A

7.64. Nucleotide Variant of

- Indicates nucleotide variants occurring within genes. Depending upon the level in which the nucleotide variant is documented (referenced to cDNA or genomic sequence), this relationship is used to relate nucleotide variant concepts to cDNA concepts, or nucleotide variant concepts to gene concepts.
- Example:

  CFTR.c.482G>A {*NUC_VAR_OF}* (cDNA).CFTR.0 (cDNA concept)
  CFTR.g.1680A>G {*NUC_VAR_OF}* (gDNA).CFTR (gene concept)

7.65. Has Effect

- Indicates the result of a nucleotide variant on a protein, used to relate nucleotide variant concepts to amino acid variant concepts.
- Example: CFTR.r.482g>a {*HAS_EFFECT}* CFTR.p.R117H

7.66. Transcript of

- Indicates the transcription product of a genomic sequence, used to relate gene concepts to RNA concepts.
- Example: (mRNA).CFTR.0 {*TRANSCRPT_OF}* (gDNA).CFTR

7.67. Reverse Transcription Product of

- Indicates the reverse transcription product of RNA transcripts, used to relate cDNA concepts to RNA concepts. cDNA concepts must be related to mRNA concepts with a corresponding suffix.
- Example: (cDNA).CFTR.0 {*RT_PROD_OF}* (mRNA).CFTR.0

7.68. Amino Acid Variant of

- Indicates amino acid variants occurring within a protein, used to relate amino acid variant concepts to protein concepts.
- Example: CFTR.p.R117H {*AA_VAR_OF}* (AA).CFTR

7.69. Encodes

- Indicates the protein product.
- Example: (mRNA).CFTR.0 {*ENCODES}* (AA).CFTR

7.70. [*] Has Phenotype

- Indicates the function of gene products, used to relate protein concepts to Gene Ontology concepts or RNA concepts to Gene Ontology concepts.
- Example: (AA).CFTR {H*AS_PHENOTYP}* Molecular Function

7.71. Has Location

- Indicates location of variant, used to relate nucleotide variant concepts to concepts within sequence features.
- Example: CYP2D6.g.4180G>C {*HAS_LOC}* CYP2D6.e.9

7.72. Telomere of

- Indicates chromosomal telomeres, used to relate telomere concepts to chromosome concepts.
- Example: 7ptel {*TELOMR_OF}* Chromosome 7

7.73. Centromere of

- Indicates chromosomal centromeres, used to relate centromere concepts to chromosome concepts.
- Example: 13cen {*CENTRMR_OF}* Chromosome 13

7.74. Has Telomere Location

- Indicates chromosomal telomere location, used to relate locus concepts to telomere concepts.
- Example: (gDNA).CFTR {*HAS_TELO_LOC}* 7q

7.75. Has Centromere Location

- Indicates chromosomal centromere location, used to relate locus concepts to centromere concepts.
- Example: (gDNA).CFTR {*HAS_CENT_LOC}* 7q

7.76. Mode of Inheritance

- Indicates mode of inheritance, used to relate locus concepts to mode of inheritance concepts.
- Example: (gDNA).CFTR {*MODE_OF_INH} A*utosomal

7.77. Species of Origin

- Indicates the species of the concept, used to relate CBO concepts to an external terminology, such as SNOMED CT™.
- Example: SPEC_OF_ORIG

7.78. Source Band

- Indicates the source band involved in a translocation, used to relate chromosomal translocation concepts to band concepts. Chromosomal translocation concepts are related to 400 resolution bands.
- Example: t(9;22)(q34;q11) {*SOURCE_BAND}* 400.22q11.2

7.79. Target Band

- Indicates the target band involved in a translocation, used to relate chromosomal translocation concepts to band concepts. Chromosomal translocation concepts are related to 400 resolution bands.
- Example: t(9;22)(q34;q11) {*SOURCE_BAND}* 400.9q34

7.80. [*] Target Centromere

- Indicates the target centromere involved in a translocation, used to relate chromosomal translocation concepts to chromosomal centromere concepts.
- Example:

7.81. [*] Source Centromere

- Indicates the source centromere involved in a translocation, used to relate chromosomal translocation concepts to chromosomal centromere concepts.
- Example:

7.82. [*] Target Telomere

- Indicates the target telomere involved in a translocation, used to relate chromosomal translocation concepts to chromosomal telomere concepts.
- Example:

7.83. [*] Source Telomere

- Indicates the source telomere involved in a translocation, used to relate chromosomal translocation concepts to chromosomal telomere concepts.
- Example:

7.84. Aneuploidy Addition of

- Indicates the addition of chromosomes, used to relate chromosomal aneuploidy concepts to chromosome concepts.
- Example: (47, XX, +21) {*ADDITION_OF}* Chromosome 21

7.85. Aneuploidy Deletion of

- Indicates the deletion of chromosomes, used to relate chromosomal aneuploidy concepts to chromosome concepts.
- Example:

7.86. Inverted Region Start Band

- Indicates the beginning band of an inverted region involved in chromosomal inversions, used to relate chromosomal inversion concepts to chromosomal band concepts. Chromosomal inversions are related to the 400 resolution bands. Bands within the inversion region are included by default as a result of the start band and end band relationships and band order facet values.
- Example: inv(3)(q21q26) { *INV_STRT_BAND* } 400.3q21

7.87. Inverted Region End Band

- Indicates the end band of an inverted region involved in chromosomal inversions, used to relate chromosomal inversion concepts to chromosomal band concepts. Chromosomal inversions are

related to the 400 resolution bands. Bands within the inversion region are included by default as a result of the start band and end band relationships and band order facet values.

- Example: inv(3)(q21q26) {*INV_END_BAND}* 400.3q26.3

7.88. [*] Inverted Centromere

- Indicates the centromere involved in a chromosomal inversion, used to relate chromosomal inversion concepts to chromosomal centromere concepts.
- Example:

7.89. [*] Inverted Telomere

- Indicates the telomere involved in a chromosomal inversion, used to relate chromosomal inversion concepts to chromosomal telomere concepts.
- Example:

7.90. [*] Deleted Telomere

- Indicates the telomere involved in a chromosomal deletion, used to relate chromosomal deletion concepts to chromosomal telomere concepts.
- Example:

7.91. [*] Deleted Centromere

- Indicates the centromere involved in a chromosomal deletion, used to relate chromosomal deletion concepts to chromosomal centromere concepts.
- Example:

7.92. Deleted Region Start Band

- Indicates the beginning band of a deleted region involved in chromosomal deletions, used to relate chromosomal deletion concepts to chromosomal band concepts. Chromosomal deletions must be related to the 850 resolution bands. Bands within the deleted region are included by default as a result of the start band and end band relationships and band order facet values.
- Example: del(15)(q11q13) {*DEL_STRT_BAND}* 400.15q11.1

7.93. Deleted Region End Band

- Indicates the end band of a deleted region involved in chromosomal deletions, used to relate chromosomal deletion concepts to chromosomal band concepts. Chromosomal deletions must be related to the 850 resolution bands. Bands within the deleted region are included by default as a result of the start band and end band relationships and band order facet values.
- Example: del(15)(q11q13) {*DEL_END_BAND}* 400.15q13

7.94. 5-UTR of

- Indicates the 5 prime UTR region of a gene, used to relate 5-UTR concepts to Gene concepts.
- Example: FY.5-UTR {*5_UTR_OF*} (gDNA).FY

7.95. 3-UTR of

- Indicates the 3 prime UTR region of a gene, used to relate 3-UTR concepts to Gene concepts.
- Example: FY.3-UTR {*3_UTR_OF*} (gDNA).FY

7.96. Nucleotide Repeat of

- Indicates nucleotide repeats within a locus, used to relate nucleotide repeat concepts to cDNA concepts.
- Example: HD.c.52_54(CAG)n {*NUC_RPT_OF*} (cDNA).HD.0

7.97. Source Gene

- Indicates the source gene involved in a translocation, used to relate chromosomal gene translocation concepts to gene concepts.
- Example: t(BCR;ABL1) {*SOURCE_GENE}* (gDNA).ABL1

7.98. Target Gene

- Indicates the target gene involved in a translocation, used to relate chromosomal gene translocation concepts to gene concepts.
- Example: t(BCR;ABL1) {*TARGET_GENE}* (gDNA).BCR

7.99. Source Exon

- Indicates the source exon involved in a translocation, used to relate chromosomal exon translocation concepts to exon concepts.
- Example: t(BCR.e.1;ABL1.e.2) {*SOURCE_EXON}* ABL1.e.2

7.100. Target Exon

- Indicates the target exon involved in a translocation, used to relate chromosomal exon translocation concepts to exon concepts.
- Example: t(BCR.e.1;ABL1.e.2) {*TARGET_EXON}* BCR.e.1

7.101. Has CH3 Status

- Indicates a locus known to have clinically significant methylation implications, used to relate locus concepts to methylation status concepts.
- Example: (gDNA).FMR1 {*HAS_CH3_STAT*} FMR1 Methylation

7.102. Feature of

- Indicates a cytogenetic entity that is part of a karyotype, used to relate chromosomal variation concepts to karyotype concepts.
- Example: del(X)(p22.3) {*FEATURE_OF*} 46,XX,del(X)(p22.3)

7.103. Has Maternal Uniparental Disomy

- Indicates maternal origin of a uniparental disomy concept, used to relate uniparental disomy concepts to karyotype concepts.
- Example: upd(15)mat {*HAS_MAT_UPD*} 46,XY,upd(15)mat

7.104. Has Paternal Uniparental Disomy

- Indicates paternal origin of a uniparental disomy concept, used to relate uniparental disomy concepts to karyotype concepts.
- Example: upd(14)pat {*HAS_PAT_UPD*} 46,XX,upd(14)pat

7.105. Isochromosome Origin

- Indicates origin of isochromosome concepts, used to relate isochromosome concepts to chromosomal band concepts
- Example: i(X)(q10) {*ISOCHRM_ORGN*} 400.Xq10

7.53 Sex Determinant

- Indicates sex chromosome complement, used to relate karyotype concepts to sex determinant concepts
- Example: 47,XX,+18 {*SEX_DTRMN*} XX

7.54 Karyotype Class

- Indicates type of karyotype (constitutional, malignant, mosaic), used to relate karyotype concepts to karyotype class concepts
- Example: 45,XX,dic(13;15)(q22;q24) {*KRYTYP_CLASS*}

7.55 Chromosomal Structural Variant

- Indicates type of chromosomal structural variation, used to relate chromosomal variation concepts to cytogenetic descriptor concepts.
- Example: der(13:14)(q10;q10) {*CHR_VAR*} Derivative Chromosome

7.56. Gene Of

- Relates a nonhuman gene with the viral/bacterial genome concept.
- Example: HIV-1.(gDNA).gag *{GENE_OF}* HIV-1.(ssRNA)

7.57. Has Gene Location

- Relates a nonhuman nucleotide variation concept to the gene of origin.
- Example: HIV-1.c.1857A>G {*HAS_GENE_LOC*} HIV-1.(gDNA).gag-pol

7.58. Constituent Amino Acid Of

- Relates a nonhuman mature protein concept to the precursor protein of origin.
- Example: HIV-1.(AA).MA {*CONS_AA_OF*} HIV-1.(AA).gag

Table 4 - CBO Relationship Definitions.

Name	Nickname	Type Restriction	Hierarchical	Inheritance Relation	Cardinality	Ordered	Transitive	Acyclic	Inheritable	Relationship Qualifier	Displayable	Immutable
Has Chromosomal Location	HAS_CHRM_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Arm of	ARM_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Band of	BAND_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Arm Location	HAS_ARM_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Band Location	HAS_BAND_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Exon of	EXON_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Intron of	INTRON_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Allele of	ALLELE_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Constituent Variant	HAS_CONS_VAR	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Nucleotide Variant of	NUC_VAR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Effect	HAS_EFFECT	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Transcript of	TRANSCRPT_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
RT Product of	RT_PROD_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Amino Acid Variant of	AA_VAR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Encodes	ENCODES	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Phenotype	HAS_PHENOTYP	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Location	HAS_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Includes Band	INCLUDS_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Telomere of	TELOMR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Centromere of	CENTRMR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Telomere Location	HAS_TELO_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Centromere Location	HAS_CENT_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Mode of Inheritance	MODE_OF_INH	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Species of Origin	SPEC_OF_ORIG	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Source Band	SOURCE_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Target Band	TARGET_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Source Centromere	SOURCE_CEN	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Target Centromere	TARGET_CEN	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Source Telomere	SOURCE_TELMR	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Target Telomere	TARGET_TELMR	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Aneuploidy Addition of	ADDITION_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes

Aneuploidy Deletion of	DELETION_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Inverted Region Start Band	INV_STRT_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Inverted Region End Band	INV_END_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Inverted Centromere	INV_CEN	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Inverted Telomere	INV_TEL	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Deleted Region Start Band	DEL_STRT_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Deleted Region End Band	DEL_END_BAND	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Deleted Centromere	DEL_CEN	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Deleted Telomere	DEL_TEL	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has CH3 Status	HAS_CH3_STAT	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
5-UTR of	5_UTR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
3-UTR of	3_UTR_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Trinucleotide Repeat Of	TRI_RPT_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Source Gene	SOURCE_GENE	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Target Gene	TARGET_GENE	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Source Exon	SOURCE_EXON	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Target Exon	TARGET_EXON	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Gene of	GENE_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Constituent Amino Acid of	CONS_AA_OF	none	no	no	multiple	no	no	TRUE	no	none	yes	yes
Has Gene Location	HAS_GENE_LOC	none	no	no	multiple	no	no	TRUE	no	none	yes	yes

8 References

1. White, J. A., McAlpine, P. J., Antonarakis, S., Cann, H., Eppig, J. T., Frazer, K., Frezal, J., Lancet, D., Nahmias, J., Pearson, P., Peters, J., Scott, A., Scott, H., Spurr, N., Talbot, C., and Povey, S. (1997). Guidelines for human gene nomenclature (1997). HUGO Nomenclature Committee. *Genomics* 45:468-71.

2. Antonarakis, S. E. (1998). Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat* 11:1-3.

3. den Dunnen, J. T. and Antonarakis, S. E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15:7-12.

4. An International System for Human Cytogenetic Nomeclature (ISCN). 1985. Report of the Standing Committee on Human Cytogenetic Nomenclature. Birth Defects Orig Artic Ser 21(1): 1-117.

5. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/

6. The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nature Genet 25: 25-29