

orphanet

orphan^{data}
USER'S GUIDE

November 2016

TABLE OF CONTENTS

RARE DISORDERS AND CROSS-REFERENCES	2
ORPHANET CLASSIFICATIONS	6
PHENTYPES ASSOCIATED TO RARE DISORDERS	8
DISORDERS WITH THEIR ASSOCIATED GENES	10
LINEARISATION OF DISORDERS	14

PART I

RARE DISORDERS AND CROSS-REFERENCES

This file contains the inventory of rare disorders listed in the Orphanet database. Each disorder is described by its preferred name and synonyms.

It is also defined by its ORPHA number, and indexed with :

- ICD-10 (10th International Classification of Diseases established by the World Health Organization - <http://www.who.int/classifications/icd/en/>),
- OMIM (Online Mendelian Inheritance in Man database - <http://www.omim.org/>),
- UMLS (Unified Medical Language System[®]) - <http://www.nlm.nih.gov/research/umls/>),
- MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed - <http://www.ncbi.nlm.nih.gov/mesh>,
- MedDRa (Medical Dictionary for Regulatory Activities - <http://www.meddra.org/>)

DEFINITIONS

Entries: Entries in the database are comprised of a heterogeneous typology of entities of decreasing extension, including: groups of disorders, disorders, sub-types. A disorder in the database can be a disease, a malformation syndrome, a clinical syndrome, a morphological or a biological anomaly or a particular clinical situation (in the course of a disorder). They are organised into groups, and further divided into clinical, etiological or histopathological sub-types.

Flags of entries: A flag is a numerical indication attached to an element of the database in order to allow retrieving information. Flags were created for entries :

- Head of classification: this is the top level of a given classification (For instance, Rare cardiac disease for Orphanet classifications of rare cardiac diseases)
- Historical entity: entries that were described long time ago, most of the time before the genetic era, and for which the princeps article is still available but no further literature exists that confirms their existence.
- Deprecated entity: these are entities formerly considered as distinct diseases and now recognised as being part of another disorder present in the database.

New relations between disorders: we identify “Moved to” relation between entries in the database when a deprecated disorder is part of another.

Characterization of the alignments between disorders and external terminologies or resources: OMIM, ICD10, MesH, UMLS, MedDRA and GARD (NIH-NCATS Genetic and Rare Diseases Center): these alignments are characterized in order to indicate if the terms are perfectly equivalent (exact mapping) or not. Possible values:

- E (exact mapping - the terms and the concepts are equivalent)
- NTBT (narrower term maps to a broader term)
- BTNT (broader term maps to a narrower term)

- NTBT/E (narrower term maps to a broader term because of an exact mapping with a synonym in the target terminology)
- BTNT/E (broader term maps to a narrower term because of an exact mapping with a synonym in the target terminology)
- ND (not yet decided/unable to decide)

Furthermore, for ICD10 a precision is added to indicate if a specific code exists in ICD10 for a disorder, or if it is listed in the tabular list or in the index in ICD10 without having a specific code, or if Orphanet has attributed the code together with the validation status for the attribution. Possible values:

- Specific code (the term has its own code in the ICD10)
- Inclusion term (the term is included under a ICD10 category and has not its own code)
- Index term (the term is included in ICD10 index pointing to a code that is not specific for the term)
- Attributed (the term does not appears at all in ICD10 and a code was attributed according to [these rules](#))

This file is available in 7 different languages: English, French, Italian, Spanish, German, Portuguese and Dutch. If terms are not yet translated from English, they remain in English by default.

The threshold applied to define a disease as rare is the one defined in Europe: any disease or condition with prevalence in the general population equal to or lower than 5 in 10,000.

Description of the XML content

DisorderList count: total number of disorders, group of disorders and subtypes in the XML file.

Orphanum: unique identifying number assigned by Orphanet to a given disease.

ExpertLink: stable URL pointing to the specific page of a given disease on the Orphanet website.

Lang: ISO 639 code for language names.

Name: preferred name of a given disorders, group of disorders and subtypes.

SynonymList: synonyms for a given disorders, group of disorders and subtypes name.

DisorderType: can be either Disease, Clinical syndrome, Malformation syndrome, Biological anomaly, Morphological anomaly, Group of phenomes, Etiological subtype,

Clinical subtype, Histopathological subtype or Particular clinical situation in a disease or syndrome

DisorderFlag: Most entries don't have a flag but for the other can be either head of classification, historical entity or deprecated entity

DisorderDisorderAssociation: Relationship "Moved to" between two disorders, group of disorders and subtypes – Please note that the hierarchical relationships are given in the Orphanet classifications file

ExternalReferenceList: list of cross-references for a given disorders, group of disorders and subtypes in the ICD-10, UMLS, MesH, MedDra, GARD and OMIM systems

Source: can be either OMIM, UMLS, MesH, MedDra, GARD or ICD-10

Reference: listed reference for a given source associated with a disorders, group of disorders and subtypes

DisorderMappingRelation: can be either E, NTBT, BTNT, NTBT/E, BTNT/E, ND (see explanations above)

DisorderMappingICDRelation: can be either Specific code, Inclusion term, Index term or Attributed (see explanations above)

DisorderMappingValidationStatus: can be either Validated or Not yet validated

Example

`<DisorderList count="9203">`

9203 is the total number of disorder, group of disorders and subtype presented in this XML file

`<Orphanum>558</Orphanum>`

The ORPHA number of this entry is 558

`<ExpertLink lang="en">http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=en&Expert=558</ExpertLink>`

This is the stable URL pointing to information on this entry

`<Name lang="en">Marfan syndrome</Name>`

The preferred name of the entry in English is "Marfan syndrome"

<SynonymList count="1"> <Synonym lang="en">MFS</Synonym>

This entry name has one synonym "MFS"

<DisorderType id="21394"><OrphaNumber>377788</Orphanumber><Name lang="en">Disease</Name></DisorderType>

The type of this entry is "disease"

<Source>MeSH</Source><Reference>D00863</Reference><DisorderMappingRelation id="21527"><Name lang="en">E (exact mapping (the terms and the concepts are equivalent))</Name></DisorderMappingRelation><DisorderMappingICDRelation/><DisorderMappingValidationStatus id="21611"><OrphaNumber>377819</OrphaNumber><Name lang="en">Validated</Name></DisorderMappingValidationStatus></ExternalReference>

This disorder is exactly mapped with MeSH reference "D00863" and the relation between reference and entry is "Validated".

<Source>ICD-10</Source><Reference>E77.1</Reference><DisorderMappingRelation id="21534"><OrphaNumber>377808</OrphaNumber><Name lang="en">NTBT (narrower term maps to a broader term)</Name></DisorderMappingRelation><DisorderMappingICDRelation id="21590"><OrphaNumber>377816</OrphaNumber><Name lang="en">Inclusion term (The term is included under a ICD10 category and has not its own code)</Name></DisorderMappingICDRelation><DisorderMappingValidationStatus id="21611"><OrphaNumber>377819</OrphaNumber><Name lang="en">Validated</Name></DisorderMappingValidationStatus></ExternalReference>

The ICD-10 reference E77.1, the orpha term is a narrower term that maps to a broader term. The term is included under a ICD10 category and has not its own code. The relation between reference and entry is "Validated".

<TextSectionType id="16907">

<Name lang="en">Disease definition</Name>

</TextSectionType>

<Contents>Marfan syndrome is a systemic disease of connective tissue characterized by a variable combination of cardiovascular, musculo-skeletal, ophthalmic and pulmonary manifestations.</Contents>

</TextSection>

This is the Orphanet entry's definition.

<Name lang="en"> Aurocephalosyndactyly </Name>

<Orphanum>1219</Orphanum><DisorderDisorderAssociation>

<Disorder1 id="1465" cycle="true"/>

<Disorder2 id="235">

<OrphaNumber>794</OrphaNumber>

<Name lang="en"> Saethre-Chotzen syndrome </Name>

</Disorder2>

<DisorderDisorderAssociationType id="21471">

<Name lang="en"> Moved to </Name>

</DisorderDisorderAssociationType>

</DisorderDisorderAssociation>

Aurocephalosyndactyly has been moved to Saethre-Chotzen syndrome.

PART II

ORPHANET CLASSIFICATIONS

This file contains the hierarchical representation, *i.e.* classification, of medical domains. There is one file per classification. This file is available in English only.

Rare disorders in Orphanet, depending on their clinical presentation, are included in as many classifications as needed. Classifications are based on published scientific articles and reviewed by experts.

Description of the XML content

Orphanum: unique identifying number assigned by Orphanet to a given classification or entry.

ExpertLink: stable URL pointing to the specific page of a given entry or classification on the Orphanet website.

Lang: ISO 639 code for language names.

Name: preferred name of a given disorders, group of disorders, sub-type or classification.

ClassificationNodeList count: number of entry present at the same level of the classification.

ClassificationNodeChildList count: number of disorders, group of disorders, sub-type classified under a given entry

Example

```
<Orphanum>68334</Orphanum>
```

The ORPHA number of this entry is 68334

```
<ExpertLink lang="en">http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=en&Expert=68334</ExpertLink>
```

The stable URL pointing to information on this entry is

http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=en&Expert=68334

<Name lang="en">Rare hemorrhagic disorder due to a coagulation factors defect</Name>

The name of this entry in English is "Rare hemorrhagic disorder due to a coagulation factors defect"

<ClassificationNodeChildList count="2">

Two entries are present at this classification level

<Orphanum>162949</Orphanum>

The classification ORPHA number is 162949

<Name lang="en">Orphanet classification of rare hematological diseases</Name>

This is the name of the classification

<ClassificationNodeChildList count="15">

There are 15 entries classified under the group of "Rare hemorrhagic disorder due to a constitutional coagulation factors defect"

PART III

PHENOTYPES ASSOCIATED WITH RARE DISORDERS

The Orphanet inventory of rare disorders is now annotated with the Human Phenotype Ontology (HPO) terms, a standardised and controlled terminology covering phenotypic abnormalities in human diseases. This new product contains two different files. The first one contains rare disorders listed in Orphanet annotated with HPO phenotypes. The alignment is characterised by frequency (obligate, very frequent, frequent, occasional, very rare or excluded) and whether the annotated HPO term is a major diagnostic criterion or a pathognomonic sign of the rare disease. The second file contains the source, the date and the validation's status of the association between the rare disease and HPO terms.

Definitions

Frequencies:

- Obligate: the phenotype is always present and the diagnosis could not be achieved in its absence;
- Very frequent: the phenotype is present in 80 to 99% of the patient population ;
- Frequent: the phenotype is present in 30 to 79% of the patient population ;
- Occasional: the phenotype is present in 5 to 29% of the patient population ;
- Very rare: the phenotype is present in 1 to 4% of the patient population ;
- Excluded: the phenotype is always absent AND is an exclusion criteria for diagnosing the disorder.

Diagnostic criterion: A diagnostic criterion is a phenotypic abnormality used consensually to assess the diagnosis of a disorder. Multiple sets of diagnostic criteria are necessary to achieve the diagnosis. Orphanet indicates only diagnostic criteria that are consensually accepted by the experts of the medical domain AND published in medical literature. Depending of the medical consensus, they could be further qualified as minor, major, etc. This level of precision is yet not included in the Orphanet dataset.

Pathognomonic sign: A pathognomonic phenotype is a feature sufficient by itself to establish definitively and beyond any doubt the diagnosis of the disease concerned (i.e. heliotrope erythema for dermatomyositis).

Files are available in 7 different languages (disease's name, frequency, diagnostic criterion and pathognomonic sign are translated): English, French, Italian, Spanish, German, Portuguese and Dutch. If terms are not yet translated from English, they remain in English by default.

Description of the XML content

File 1

DisorderList count: total number of disorders, group of disorders and subtypes in the XML file.

Orphanum: unique identifying number assigned by Orphanet to a given entry

Lang: ISO 639 code for language names.

Name: preferred name of a given entry.

HPODisorderAssociationList count: number of HPO phenotypes associated with a given entry.

HPOID: unique identifying number assigned by HPO to a given phenotype.

HPOTerm: preferred name of HPO phenotype.

HPOFrequency: estimated frequency of occurrence for a given phenotype in a given entry. Five different frequency groups have been defined.

DiagnosticCriteria: indicate if the given phenotype is a pathognomonic sign or a diagnostic criterion in a given entry.

Example

`<DisorderList count="2214">`

2214 is the total number of disorder, group of disorders and subtype presented in this XML file

`<Name lang="en">Kawasaki disease</Name>`

The main name of the entry is "Kawasaki disease"

`< HPODisorderAssociationList count="6">`

There are 6 HPO phenotypes associated with this entry

`<DiseaseSign>`

`<Name lang="en"> Kawasaki disease </Name>`

`<HPOID>HP:0001945</HPOID>`

`<HPOTerm>Fever</HPOTerm>`

`<DiagnosticCriteria>Pathognomonic sign</ DiagnosticCriteria >`

Fever is a pathognomonic sign seen in patients with Kawasaki disease.

PART IV

DISORDERS WITH THEIR ASSOCIATED GENES

This product contains the inventory of rare disorders listed in the Orphanet database and the associated genes (pathogenic and susceptibility genes). Each gene is described by its official name, synonyms, type and location, defined by the HUGO Gene Nomenclature Committee (HGNC). Genes are mapped with their *HGNC ID*, *OMIM number*, *GenAtlas* symbol, *UniProtKB/Swiss-Prot references*, *Ensembl*, *Reactome* and *IU-PHAR*.

The relationship between a gene and a disease is qualified according to the role that the gene plays in the pathogenesis of a disease. Genes are annotated as causative, modifiers (both from germline or somatic mutations), major susceptibility factors or playing a role in the phenotype (for chromosomal anomalies). Candidate genes are included if a genetic test exists in the clinical setting.

HGNC (<http://www.genenames.org/>) is a committee jointly funded by the US National Human Genome Research Institute (NHGRI) and the Wellcome Trust (UK). It operates under the auspices of HUGO, with key policy advice from an International Advisory Committee. It is in charge of approving gene names and symbols (short-form abbreviations). All approved symbols are stored in the HGNC database. Each symbol is unique and each gene is only given one approved gene symbol.

OMIM, Online Mendelian Inheritance in Man (<http://www.omim.org/>), is the database of human genes and genetic phenotypes.

GenAtlas (<http://www.genatlas.org/>) is a database of genes and phenotypes. Only the objects with a known cytogenetic location are retained.

UniProtKB (<http://www.uniprot.org/>) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation.

Ensembl (<http://www.ensembl.org/>) is an EBI database that maintains automatic annotation on selected eukaryotic genomes.

Reactome (<http://www.reactome.org/>) is an EBI open-source, open access, manually curated and peer-reviewed pathway database.

IUPHAR (<http://www.iuphar.org/>) is The International Union of Basic and Clinical Pharmacology.

Genetic entries in the database will have new information attached:

The type of the entry: either gene with protein product, locus, non-coding RNA

The chromosomal location

Two new gene-disease relationships: gain of function and loss of function germline disease causing mutation.

This product is available in English only.

Description of the XML content

DisorderList count: total number of disorders, group of disorders and subtypes in the XML file.

Orphanum: unique identifying number assigned by Orphanet to a given entry (disorder, group of disorders, subtype or gene).

Name: preferred name of a given entry (disorder, group of disorders, subtype or gene).

GeneList count: number of genes associated with a given entry.

Symbol: official HGNC-approved gene symbol.

Synonym list: list of synonyms for a given gene, including past symbols

GeneType: can be either gene with protein product, locus or non-coding RNA

GeneLocus: gene chromosomal location

DisorderGeneAssociationType: gene-disease relationships. They can be either Role in the phenotype of, Disease-causing germline mutation(s) (loss of function) in, Disease-causing germline mutation(s) (gain of function) in, Disease-causing somatic mutation(s) in, Modifying somatic mutation in, Part of a fusion gene in, Major susceptibility factor in and Candidate gene tested in.

DisorderGeneAssociationStatus: can be either Validated or Not validated

External Reference List: list of references in HGNC, OMIM, GenAtlas and UniProtKB, *Ensembl*, *Reactome* and *IU-PHAR* associated with a given gene.

Source: HGNC, OMIM, GenAtlas or UniProtKB.

Reference: listed reference for a given source associated with a gene.

Example

`<DisorderList count="9203">`

9203 is the total number of disorder, group of disorders and subtype presented in this XML file

`<Orphanum>61</Orphanum>`The ORPHA number of this entry is 61

`<GeneList count="1">`

The entry is associated with one gene

`<Orphanum>123328</Orphanum>`

The gene ORPHA number is 123328

<Symbol>**MAN2B1**</Symbol>

<Name lang="en">**Mannosidase, alpha, class 2B, member 1**</Name>

Its official symbol and name are MAN2B1 and Mannosidase, alpha, class 2B, member 1, respectively

<SynonymList count="2">

<Synonym lang="en">**LAMAN**</Synonym>

<Synonym lang="en">**MANB**</Synonym>

There are two synonyms for this gene: LAMAN and MANB

<GeneType id="24110">

<Name lang="en">**gene with protein product**</Name></GeneType>

The type of the given entry is a gene with protein product

<Locus id="7379">

<GeneLocus>**7q31-q32**</GeneLocus >

The given gene is located on 7q31-32

<DisorderGeneAssociationType id="17949">

<Name lang="en">**Disease-causing germline mutation(s) in**</Name>

</DisorderGeneAssociationType>

The relationship between a given entry and the given gene is Disease-causing germiline mutation(s) in

<DisorderGeneAssociationStatus id="17991">

<Name lang="en">**Validated**</Name>

The status of the relationship between a given entry and the given gene is Validated

<ExternalReferenceList count="4">

This gene is mapped with 4 other references

<Source>**GENATLAS**</Source>

<Reference>**MAN2B1**</Reference>

The reference for this gene in GenAtlas is MAN2B1

<Source>**HGNC**</Source>

<Reference>**6826**</Reference>

The reference for this gene in HGNC is 6826

<Source>**OMIM**</Source>

<Reference>**609458**</Reference>

The reference for this gene in OMIM is 609458

<Source>**UNIPROTKB/SWISSPROT**</Source>

<Reference>**O00754**</Reference>

The reference for this protein in UniProtKB is O00754

PART V

LINEARISATION OF DISORDERS

This file contains linearization of disorders. For analysis purposes, each disorder is attributed to a preferred classification by linking it to the head of classification entity. As some decisions could be arbitrary, we have established a set of rules to make attributions consistently. The methodology is described [here](#).

This file is available only in English.

Description of the XML content

DisorderList count: total number of disorders, group of disorders and subtypes in the XML file.

Orphanum: unique identifying number assigned by Orphanet to a given entry.

ExpertLink stable URL pointing to the specific page of a given entry on the Orphanet website.

DisorderDisorderAssociation: Relationship “Preferential parent” between one entry and a head of classification

Example

```
<DisorderList count="9203">
```

9203 is the total number of disorder, group of disorders and subtype presented in this XML file

```
<Name lang="en"> Alpha-mannosidosis </Name>
<Orphanum>61</Orphanum><DisorderDisorderAssociation>
  <Disorder1 id="3" cycle="true"/>
  <Disorder2 id="10507">
    <OrphaNumber>68367</OrphaNumber>
    <Name lang="en"> Inborn errors of metabolism </Name>
  </Disorder2>
  <DisorderDisorderAssociationType id="21485">
    <Name lang="en"> Preferential parent </Name>
  </DisorderDisorderAssociationType>
</DisorderDisorderAssociation>
```

The preferential parent of **Alpha-mannosidosis (Orphanum 61)** is **Inborn errors of metabolism (Orphanum 68367)**.