WHAT IS THE ORPHANET RARE DISEASE ONTOLOGY (ORDO)?

SEPTEMBER 2019

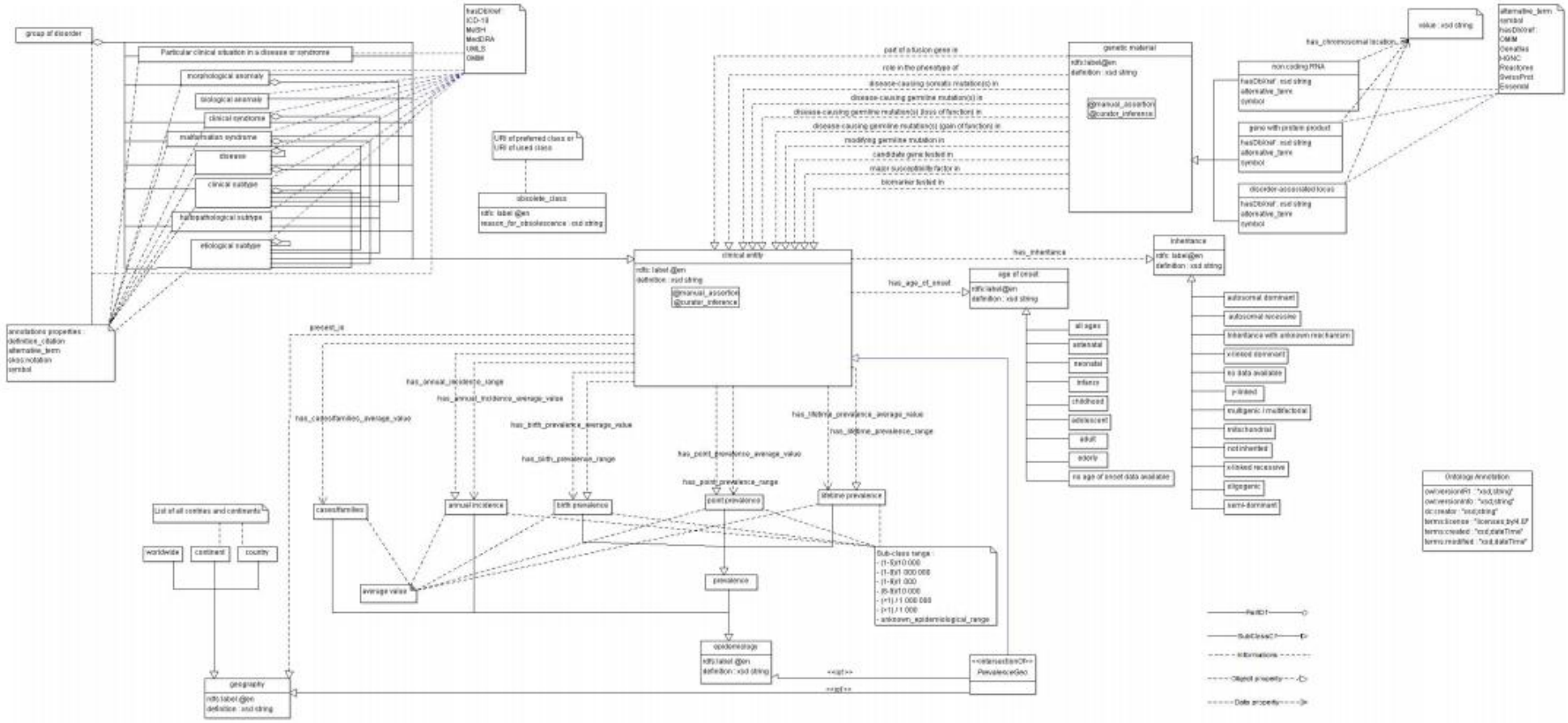
www.orphadata.org **www.orpha.net**

The Orphanet Rare Disease ontology (ORDO) was jointly developed by Orphanet and the EBI to provide a structured vocabulary for rare diseases capturing relationships between diseases, genes and other relevant features which will form a useful resource for the computational analysis of rare diseases. It derived from the Orphanet database (www.orpha.net), a multilingual database dedicated to rare diseases populated from literature and validated by international experts. It integrates a nosology (classification of rare diseases), relationships (gene-disease relations, epidemiological data) and connections with other terminologies (MeSH, SNOMED CT, UMLS, MedDRA), databases (OMIM, UniProtKB, HGNC, ensembl, Reactome, IUPHAR, Geantlas) or classifications (ICD-10). The ontology will be maintained by Orphanet and further populated with new data. Orphanet classifications can be browsed in the OLS view. The Orphanet Rare Disease Ontology is updated every six months and follows the OBO guidelines on deprecation of terms. It constitutes the official ontology of rare diseases produced and maintained by Orphanet (INSERM, US14).

ORDO is accessible on several websites:

- Link to ORDO on Orphadata website
- Link to SPARQL Endpoints for ORDO on Orphadata website
- Link to ORDO on Bioportal
- Link to ORDO on EBI website
- Link to ORDO on Orphanet website

ORDO MODEL


group of disorder
Particular clinical situation in a disease or syndrome
morphological anomaly
biological anomaly
clinical syndrome
malformation syndrome
disease
clinical subtype
histopathological subtype
etiological subtype
hasDbXref
ICD-10
MeSH
MedDRA
UMLS
OMIM
URI of preferred class or URI of used class
obsolete_class
rdfs: label @en
reason_for_obsolescence : xsd string
clinical entity
rdfs: label @en
definition : xsd string
@manual_assertion
@curator_inference
annotations properties :
definition_citation
alternative_term
class notation
symbol
part of a fusion gene in
role in the phenotype of
disease-causing somatic mutation(s) in
disease-causing germline mutation(s) in
disease-causing germline mutation(s) (loss of function) in
disease-causing germline mutation(s) (gain of function) in
modifying germline mutation in
candidate gene tested in
major susceptibility factor in
biomarker tested in
genetic material
rdfs:label@en
definition : xsd string
@manual_assertion
@curator_inference
has_chromosomal location
value : xsd string
non coding RNA
hasDbXref : xsd string
alternative_term
symbol
gene with protein product
hasDbXref : xsd string
alternative_term
symbol
disorder-associated locus
hasDbXref : xsd string
alternative_term
symbol
alternative_term
remind
hasDbXref :
OMIM
Genatlas
HGNC
Reactome
SwissProt
Ensembl
has_inheritance
inheritance
rdfs: label@en
definition : xsd string
has_age_of_onset
age of onset
rdfs:label@en
definition : xsd string
all ages
antenatal
neonatal
infancy
childhood
adolescent
adult
elderly
no age of onset data available
autosomal dominant
autosomal recessive
inheritance with unknown mechanism
x-linked dominant
no data available
y-linked
multigenic / multifactorial
mitochondrial
not inherited
x-linked recessive
oligogenic
semi-dominant
prevalence
has_annual_incidence_range
has_annual_incidence_average_value
has_cases/families_average_value
has_birth_prevalence_average_value
has_birth_prevalence_range
has_point_prevalence_average_value
has_point_prevalence_range
has_lifetime_prevalence_average_value
has_lifetime_prevalence_range
cases/families
annual incidence
birth prevalence
point prevalence
lifetime prevalence
average value
prevalence
Sub-class range :
-]1-f(x)1.000
-]1-f(x)1.000.000
-]1-f(x)1.000
-]0-f(x)1.000.000
-]>1[/ 1.000.000
-]>1[/ 1.000
- unknown_epidemiological_range
List of all contries and continents
worldwide
continent
country
geography
rdfs:label @en
definition : xsd string
epidemiology
rdfs:label @en
definition : xsd string
<<extensionOf>>
PrevalenceGeo
<<get>>
<<get>>
Ontology Annotation
owl:versionIRI : "xsd:string"
owl:versionInfo : "xsd:string"
dc:creator : "xsd:string"
terms:license : "licenses.byNd.EP"
terms:created : "xsd:dateTime"
terms:modified : "xsd:dateTime"
SubClassOf
isEquivalentTo
Object property
Data property

Briefly, In ORDO, there are 7 SuperClass : "**Clinical entity", which is central and "Age of onset", "Epidemiology", "Genetic material", "Geography", "Inheritance", "Obsolete"**.

SuperClass	Definition
clinical entity	A set of phenotypic abnormalities.
age of onset	Age of onset of clinical manifestations related to a clinical entity.
epidemiology	Number of cases in a population during a given period. The number of cases may be expressed by annual incidence (Annual incidence), prevalence (Prevalence) and / or number of cases / families (Cases/families) published in the literature.
genetic material	DNA or RNA sequence (gene with protein product, non-coding RNA and disorder-associated locus).
geography	Refers to a country, a continent or the whole world.
inheritance	Pattern in which a clinical entity of genetic origin is being passed down to the offspring.
obsolete	ORPHA number no more in use. Deprecated entities are part of them. They are clinical entities thought to be unique in the past but now considered, thanks to the evolution of the knowledge, part of another clinical entity.

The "**Clinical entity**" SuperClass contains classes representing the classification of rare diseases. These classes have poly-parental relations.

SubClassOF "Clinical Entity"	Definition
biological anomaly	Clinical entity defined by a set of physiological abnormalities without clear associated clinical manifestations.
clinical subtype	Subdivision of a disease, malformation syndrome, morphological anomaly, biological anomaly, clinical syndrome or particular clinical situation in a disease or a syndrome further defined by its particular clinical presentation.
clinical syndrome	Clinical entity defined by a set of phenotypic abnormalities with a homogeneous evolution and homogeneous therapeutic possibilities, regardless of the physiopathological mechanism.

disease	Clinical entity defined by a set of phenotypic abnormalities resulting from a common physiopathological mechanism with a homogeneous evolution and homogeneous therapeutic possibilities. Excludes developmental anomalies.
etiological subtype	Subdivision of a disease, malformation syndrome, morphological anomaly, biological anomaly or particular clinical situation in a disease or a syndrome further defined by its aetiology.
group of disorders	Clinical entity defined by a set of phenotypic abnormalities shared by several diseases, malformation or clinical syndromes, morphological or biological anomalies, and particular clinical situations in a disease or a syndrome and used to group them together.
histopathological subtype	Subdivision of a disease, malformation syndrome, morphological anomaly, biological anomaly, clinical syndrome or particular clinical situation in a disease defined by the histological abnormalities in affected tissues.
malformation syndrome	Clinical entity defined by a set of morphological abnormalities resulting from a developmental anomaly involving more than one morphogenetic field. Includes sequences and associations.
morphological anomaly	Clinical entity defined by an alteration of the normal morphology resulting from a development anomaly involving a single morphogenetic field.
particular clinical situation in a disease or syndrome	Clinical entity defined by a set of phenotypic abnormalities occurring in particular circumstances.

The other SuperClasses ("**Age of onset", "Epidemiology", "Genetic material", "Geography", "Inheritance", "Obsolete**") have SubClass. There are describe below.

SuperClass	SubClasseOf	Definition
age of onset	adolescent	From 12 to 18 years.
	adult	From 19 to 65 years.
	all ages	From birth to adulthood without peak of onset.
	antenatal	Before birth.
	childhood	From 2 to 11 years.
	ederly	After 65 years.
	infancy	From the end of the fourth week to the 23rd month of life.

	neonatal	From birth to the fourth week of life.
	no age of onset data available	No information is available in the scientific literature on the age of onset of the first clinical manifestations.
genetic material	disorder-associated locus	Chromosomal region associated with a hereditary disorder but without any precision on the possible associated gene.
	gene with protein product	DNA sequence translated into protein.
	non coding RNA	RNA transcript not translated into protein.
inheritance	autosomal dominant	Pattern of inheritance in which a single mutated allele located on one of the 22 autosomes (non-sex chromosomes) is sufficient to express the phenotype.
	autosomal recessive	Pattern of inheritance in which two mutated alleles of the same gene located on one of the 22 autosomes (non-sexual chromosomes) are needed to express the phenotype.
	mitochondrial	Pattern of inheritance in which a mutation in one of the mitochondrial genes is sufficient to express the phenotype. The transmission is exclusively maternal.
	multigenic/multifactorial	The combination of one or more genes and/or environmental factors contributes to the expression of the phenotype.
	no inheritance data available	No information is available in the scientific literature on heredity of the clinical entity.
	not genetically inherited	clinical entity without genetic inheritance.
	oligogenic	The combination of mutated alleles of two or more genes is necessary to express the phenotype.
	semi-dominant	Pattern of inheritance in which a single mutated allele located on one of the 22 autosomes (non-sex chromosomes) suffices to express the phenotype, the phenotype of the homozygous individual being more severe, when both alleles are mutated.
	unknown inheritance	Hereditary clinical entity whose mode of inheritance is unknown.
	X-linked dominant	Pattern of inheritance in which a single mutated allele on the X chromosome is sufficient to express the phenotype. The phenotype is more consistently and severely expressed in hemizygous boys (having only one copy of the gene) than in heterozygous girls.
	X-linked recessive	Pattern of inheritance in which two mutated alleles on the X chromosome are needed to express the phenotype. The phenotype is expressed in hemizygous boys (having only one copy of the gene) and homozygous girls.
	Y-linked	Pattern of inheritance in which a single mutated allele on the Y chromosome is

		sufficient to express the phenotype. The transmission is exclusively paternal.

SuperClass	SubClasseOf	Definition	SubClasseOf	Definition	SubClasseOf	Definition
epidemiology	annual incidence	Number of newly diagnosed cases in a population in 1 year.				
	cases/ families	Number of cases or family (ies) published in the literature.	case	Number of cases published in the literature.		
			family	Number of family(ies) published in the literature.		
	prevalence	Number of cases scaled up to the general population at a given time or during a given period. Prevalence can be observed at birth (prevalence at birth), at a point in time (point prevalence), or in a lifetime (lifetime prevalence).	birth prevalence	Number of cases observed at birth relative to the number of children born alive at a given moment.	<1 / 1 000 000	Interval of prevalence or annual incidence of less than 1 case per 1,000,000 in the population.
					>1 / 1000	Interval of prevalence or annual incidence greater than 1 case per 1,000 in the population.
					1-5 / 10 000	Interval of prevalence or annual incidence of between 1 and 5 cases per 10,000 in the population.
					1-9 / 1 000 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 1,000,000 in the population.
					1-9 / 100 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 100,000 in the population.
					6-9 / 10 000	Interval of prevalence or annual incidence of between 6 and 9 cases per 10,000 in the population.
					Unknown_ epidemiological_range	No information is available in the scientific literature to inform prevalence or annual incidence.
			lifetime prevalence	Number of cases presenting or having presented the clinical entity during their lifetime scaled up to the general population.	<1 / 1 000 000	Interval of prevalence or annual incidence of less than 1 case per 1,000,000 in the population.
					>1 / 1000	Interval of prevalence or annual incidence greater than 1 case per 1,000 in the population.
					1-5 / 10 000	Interval of prevalence or annual incidence of between 1 and 5 cases per 10,000 in the population.
					1-9 / 1 000 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 1,000,000 in the population.
					1-9 / 100 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 100,000 in the population.
					6-9 / 10 000	Interval of prevalence or annual incidence of between 6 and 9 cases per 10,000 in the population.
					Unknown_	No information is available in the scientific literature to inform

					epidemiological_range	prevalence or annual incidence.
			point prevalence	Number of cases scaled up to the general population at a given time.	<1 / 1 000 000	Interval of prevalence or annual incidence of less than 1 case per 1,000,000 in the population.
					>1 / 1000	Interval of prevalence or annual incidence greater than 1 case per 1,000 in the population.
					1-5 / 10 000	Interval of prevalence or annual incidence of between 1 and 5 cases per 10,000 in the population.
					1-9 / 1 000 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 1,000,000 in the population.
					1-9 / 100 000	Interval of prevalence or annual incidence of between 1 and 9 cases per 100,000 in the population.
					6-9 / 10 000	Interval of prevalence or annual incidence of between 6 and 9 cases per 10,000 in the population.
					Unknown_ epidemiological_range	No information is available in the scientific literature to inform prevalence or annual incidence.

Definition of properties in ORDO

Property	Definition
biomarker tested in	A gene in which a variation is used to monitor disorder activity and/or patient outcome.
candidate gene tested in	A gene in which a mutation is suspected, but not yet proven, to be responsible for a disorder, but for which a genetic test (s) is (are) available
disease-causing germline mutation(s) (gain of function) in	Relationship between clinical entity and age of onset.
disease-causing germline mutation(s) (loss of function) in	Relationship between the clinical entity and the mean value of its annual incidence.
disease-causing germline mutation(s) in	Relationship between clinical entity and annual incidence range.
disease-causing somatic mutation(s) in	Relationship between the clinical entity and the mean value of its birth prevalence.
has_age_of_onset	Relationship between clinical entity and birth prevalence range.
has_annual_incidence_average_value	Relationship between clinical entity and number of cases/families.
has_annual_incidence_range	Relationship between a gene with protein product, non-coding RNA or disorder-associated locus and its cytogenetic location on the chromosome.
has_birth_prevalence_average_value	Relationship between a clinical entity and modes of inheritance.
has_birth_prevalence_range	Relationship between the clinical entity and the mean value of its lifetime prevalence.
has_cases/families_average_value	Relationship between clinical entity and lifetime prevalence range.
has_chromosomal location	Relationship between the clinical entity and the mean value of its point prevalence.
has_inheritance	Relationship between clinical entity and point prevalence range.
has_lifetime_prevalence_average_value	RNA transcript not translated into protein.
has_lifetime_prevalence_range	A coding or regulatory DNA sequence from a gene that has fused with another coding and/or regulatory DNA sequence from a different gene.
has_point_prevalence_average_value	Relationship between a clinical entity and the geographical area for which epidemiological data (Epidemiology) is available.
has_point_prevalence_range	A gene included in a chromosomal rearrangement, and proved to have a major influence in the phenotype of the chromosomal rearrangement.
major susceptibility factor in	A gene mutation in a germ cell that predisposes to the development of a

	disorder, and that is necessary but not sufficient to develop the disorder.
non coding RNA	A mutation of a gene in a germ cell that results in a new function of the corresponding protein is sufficient to cause the disorder and can be transmitted to the offspring.
part of a fusion gene in	A mutation of a gene in a germ cell that alters the function of the corresponding protein is sufficient to cause the disorder and can be transmitted to the offspring.
present_in	A mutation of a gene in a germ cell that is sufficient to cause the disorder and can be transmitted to the offspring.
role in the phenotype of	A mutation of a gene in a somatic cell that is sufficient to cause the disorder but can not be transmitted to the offspring.
modifying germline mutation in	A gene mutation in a germ cell that modifies the clinical presentation of the disorder and that can be passed on to offspring.
part_of	Relation between two clinal entities, one being included in the other. Ex : clinical subtype part_of disease.

To exploit ORDO, SPARQL queries can be use on SPARQLendpoint, virtuoso, blazegraph or other tools who permit SPARQL queries.

These are examples of queries that you can use:

Example 1:

Getting a concept label from Orphanumber.

```
PREFIX ordo:<http://www.orpha.net/ORDO/>
PREFIX w3: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?label
WHERE {
ordo:Orphanet_558 w3:label ?label
}
```

Example 2:

Getting a concept label and alternative term from Orphanumber.

```
PREFIX ordo:<http://www.orpha.net/ORDO/>
PREFIX ebi: <http://www.ebi.ac.uk/efo/>
PREFIX w3: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?label ?alternativeterm
WHERE {
ordo:Orphanet_1187 w3:label ?label.
ordo:Orphanet_1187 ebi:alternative_term ?alternativeterm
}
```

Example 3:

Getting the genetic material linked to more than 10 disorders

```
PREFIX ordo:<http://www.orpha.net/ORDO/>
PREFIX w3: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

SELECT ?gene ?geneLab ?nbD
WHERE{
```

```
  {
    SELECT ?g  (COUNT(?d) as ?nbD)
    WHERE {
      ?r owl:onProperty ?rel.
      ?g w3:label ?gLabel.
      ?g w3:subClassOf ?r.
      ?g w3:subClassOf ?class.
      ?class w3:subClassOf ?sc.
      filter (?sc = ordo:Orphanet_C010)
      ?r owl:someValuesFrom ?d.
    }
    GROUP BY ?g
  }
  filter (?nbD > 10)
  BIND (?g as ?gene)
  ?gene w3:label ?geneLab.
}
ORDER BY DESC(?nbD)
```

Example 4:

Getting the label and mapping information about a concept.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX oboInOwl: <http://www.geneontology.org/formats/oboInOwl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX Orphanet_: <http://www.orpha.net/ORDO/Orphanet_#>
PREFIX ORDO: <http://www.orpha.net/ORDO/>
```

```
select ?s ?p ?o WHERE {
?s ?p ?o.
optional {?o ?m ?v.
?v rdfs:label ?e}.
filter (?s = ORDO:Orphanet_558)
}
ORDER BY ?o)
```

LEXICON

EMBL-EBI : European Bioinformatics Institute

HGNC : *HUGO Gene Nomenclature Committee*

HIPBI-RD: Harmonising phenomics information for a better interoperability in the RD field

ICD-10 : International Classification of Diseases 10th

INSERM : Institut national de la santé et de la recherche médicale

IUPHAR : International Union of Basic and Clinical Pharmacology

MedDRA : Medical Dictionary for Regulatory Activities

MeSH : Medical Subject Headings

OBO :

OLS : Ontology Lookup Service

OMIM : *Online Mendelian Inheritance in Man*

ORDO: Orphanet Rare Disease Ontology

RD-Action: Rare Diseases Action

SNOMED CT: SNOMED Clinical Trial

UMLS : *Unified Medical Language System*

UniProtKB : UniProt Knowledgebase

For any questions or comments, please contact the Orphadata team: data.orphanet@inserm.fr

The correct form when quoting this document is :
" What is the Orphanet Rare Disease Ontology (ORDO)? " – September 2019
http://www.orphadata.org/cgi-bin/img/PDF/WhatIsORDO.pdf

The activities described in this document are part of the Direct Grant N°831390 which has received funding from the European Union's Health Programme (2014-2020).

HOOM is part of the HIPBI-RD project, this project (HIPBI-RD) has received funding from the European Union's Horizon 2020 Research and Innovation program under the ERA-NET Cofund Action n° 643578. It was supported by the French National Research Agency (ANR), the German Federal Ministry of Education and Research (BMBF), German Research Foundation (DFG), and the Canadian Institutes for Health Research (CIHR) under the frame of E-Rare-3, the ERA-NET for research on rare diseases.

Inserm
La science pour la santé
From science to health

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

Co-funded by the Health Programme of the European Union