

Procedural document Inventory of genes related to rare diseases

www.orpha.net

www.orphadata.org

Table of content

I.	Introduction	3
1.	Purpose/objectives	3
2.	Range of application	3
3.	Disclaimer.....	3
4.	References	3
5.	Definitions	4
a.	Orphanet gene-related dataset.....	4
b.	Other definitions.....	5
6.	Filing and updates.....	6
II.	Methodology.....	7
1.	Flowchart	7
2.	Description.....	11
a.	Requests for modification of the Orphanet inventory of genes	11
b.	Analysis of scientific resources.....	11
c.	Identification of the related rare disease	11
d.	Identification of the gene-disease relationship.....	11
e.	Update of the Orphanet database	13
f.	Quality control	13
III.	Availability of data	14

I. Introduction

1. Purpose/objectives

Orphanet has developed and maintains the Orphanet nomenclature of rare diseases (RDs), a unique and multilingual standardised system aimed at providing a specific terminology for RDs. Each clinical entity within the Orphanet nomenclature of RDs is enriched by scientific and textual information. In order to better define RDs of genetic origin, Orphanet provides information on genes related to a RD. This information is structured around the international nomenclature of genes, provided by the HUGO Gene Nomenclature Committee (HGNC), and comprises the gene typology, the chromosomal location, cross-mappings with international genetic databases and semantic relationships between genes and their related RDs, along with the evidence-based publications supporting this relationship.

This procedural document describes the production, validation and update process of the Orphanet inventory of genes related to RDs.

2. Range of application

The present procedure applies to all genes listed in the Orphanet database.

The Orphanet inventory of genes is managed by an information scientist with a genetic background, under responsibility of the Orphanet nomenclature manager. The Orphanet Advisory Board on Genetics and international experts are regularly consulted to ensure the accurate representation of the current knowledge. The Orphanet Advisory Board on Genetics is also consulted on this procedure.

3. Disclaimer

- This publication is part of the OrphaNetWork Direct Grant (831390) which has received funding from the European Union's Health Programme (2014-2020).
- The content of this publication represents the views of the author only and is his/her sole responsibility; it can not be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.
- Candidate genes and biomarkers are excluded from the inventory unless a diagnostic test is made available in one of the Orphanet network's countries.

4. References

For more information on the management of the inventory and classification of RDs, see the procedural document: [The Orphanet nomenclature and classification of rare diseases](https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_Genes_inventory_R1_Ann_gen_EP_02.pdf).

5. Definitions

a. Orphanet gene-related dataset

Gene dataset	Definition	Data coming from
Gene nomenclature		
Gene name	HGNC-approved name	HGNC
Symbol	HGNC-approved symbol	HGNC
Synonyms	HGNC alias symbols and names	HGNC
Previous symbols and names	HGNC previous symbols and names	HGNC
Gene information		
Gene Type:		
Gene with protein product ¹	Basic unit of heredity, consisting of a segment of DNA arranged in a linear manner along a chromosome that is transcribed in RNA and translated in a protein.	HGNC
Disorder-associated locus ²	Chromosomal region associated with a single heritable disorder. The heritable disorder may be mapped to a chromosome but generally has not been associated to a specific gene.	HGNC
Non-coding RNA ³	RNA encoded by a gene but not translated in protein. ie: Transfer RNA.	HGNC
Chromosomal location	Cytogenetic location on the chromosome of a gene with protein product, non-coding RNA or disorder-associated locus.	HGNC
Genetic database cross-mappings		
HGNC	Non-profit making body which is jointly funded by the US National Human Genome Research Institute (NHGRI) and the Wellcome Trust (UK). The HGNC is responsible for approving unique symbols and names for human loci.	Orphanet curation
OMIM	Online Mendelian Inheritance in Man is a compendium of human genes and genetic phenotypes.	HGNC
UniProtKB	Central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation.	HGNC
Ensembl	An EBI database that maintains automatic annotation on selected eukaryotic genomes.	HGNC
IUPHAR	The International Union of Basic and Clinical Pharmacology portal.	HGNC
Reactome	An open-source, open access, manually curated and peer-reviewed pathway database.	HGNC
GenAtlas	A database of genes and phenotypes. Only the objects with a known cytogenetic location are retained. This cross reference is still present in Orphanet database but no longer updated.	HGNC
LOVD	Leiden Open Variation Database is an online gene-centered collection and display of DNA variants.	HGNC

¹ Matches with Gene with protein product; immunoglobulin gene; T cell receptor gene; Protocadherin; Readthrough (from HGNC locus type)

² Matches with RNA, long non-coding; Pseudogene; Virus integration site; Endogenous retrovirus; Immunoglobulin pseudogene; RNA, micro; RNA, ribosomal; RNA, transfer; RNA, cluster; RNA, misc; RNA, small nuclear; RNA, small nucleolar; RNA, small cytoplasmic; RNA, Y; T cell receptor pseudogene; RNA, vault; unknown (from HGNC locus type)

³ Matches with region; Complex locus constituent; Fragile site (from HGNC locus type)

Gene dataset	Definition	Data coming from
Gene-disease relationship		
Disease-causing germline mutation(s) in	A gene mutation in a germ cell that is sufficient to produce the disorder and that can be passed on to offspring.	Orphanet curation
Disease-causing germline mutation(s) (gain of function) in	A gene mutation in a germ cell that provides a new function of the corresponding protein and that is sufficient to produce the disorder and that can be passed on to offspring.	Orphanet curation
Disease-causing germline mutation(s) (loss of function) in	A gene mutation in a germ cell that impairs the function of the corresponding protein and that is sufficient to produce the disorder and that can be passed on to offspring.	Orphanet curation
Disease-causing somatic mutation(s) in	A gene mutation in a somatic cell that is sufficient to produce the disorder but that cannot be passed on to offspring.	Orphanet curation
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.	Orphanet curation
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.	Orphanet curation
Part of a fusion gene in	A coding or regulatory DNA sequence from a gene that has fused with another coding and/or regulatory DNA sequence from a different gene.	Orphanet curation
Role in the phenotype of	A gene included in a chromosomal rearrangement of which the mutation results in a phenotype related to the given chromosomal rearrangement, therefore proving its influence in a particular manifestation.	Orphanet curation
Biomarker tested in	A gene in which a variation is used to monitor disease activity and/or patient outcome.	Orphanet curation
Candidate gene tested in	A gene in which a mutation is suspected, but not yet proven, to be responsible for a disorder, and that is tested for in a clinical setting.	Orphanet curation
Reference	Reference(s) for a given source associated with a disease-gene relationship.	Orphanet curation
Validation status	Indication that the curation is done accordingly to Orphanet procedures and is scientifically valid.	Orphanet curation

b. Other definitions

Experts: a medical doctor or researcher with prominent experience in a rare disease or a group of rare diseases, and identified by Orphanet based on published articles (particularly reviews and guidelines), involvement in expert centers, expert networks, and/or in dedicated research activities including clinical trials.

Familial cases: At least two patients in the same family affected by a rare disease suspected to be of genetic origin.

Genetic functional study: Genetic functional studies aim to explain the physiological and pathophysiological roles of a gene. The functional studies used to validate the association of a gene with a disorder can be carried out in animal models like knock-out, cell experiments such as restoration of a gene/protein activity and/or *in vitro* experiments like testing the gene expression level and/or protein activity.

Information scientist: a member of the Orphanet team with a scientific and/or medical background in charge of collecting, producing and updating information provided in the Orphanet database.

Orphanet advisory board on genetics: This board is composed either by geneticists, members of the Orphanet management board and joining the advisory board on a volunteer basis, or by geneticists not

belonging to the Orphanet management board and invited under proposal of the Orphanet Management Board. The membership, mandate and role of the Orphanet Advisory Board on Genetics is available [here](#).

Orphanet gene manager: an information scientist of the Orphanet coordinating team in charge of updating gene information and their relationships to the Orphanet nomenclature of rare diseases.

Orphanet nomenclature manager: an information scientist of the Orphanet coordinating team in charge of producing and updating the nomenclature and classification of rare diseases.

Orphanet national team: An Orphanet team based in one of the member countries of the Orphanet Network as per the Orphanet Network Agreement, and responsible for the collection of data on national expert resources. Some of the national teams are also in charge of the translation of the Orphanet nomenclature into one of the languages of translation of the database (German, Italian, Spanish, Portuguese, Polish, Czech and Dutch).

Orphanet Rare Disease Ontology (ORDO): an open access, structured, and machine-readable vocabulary for rare diseases derived from the Orphanet database, capturing relationships between diseases, genes and other relevant features, and forming a useful resource for the computational analysis of rare diseases. ORDO was jointly developed by Orphanet and the European Bioinformatics Institute (EMBL-EBI).

Orphadata: a platform developed by Orphanet to provide the scientific community with comprehensive, high-quality and freely accessible datasets related to rare diseases and orphan drugs, in a reusable format.

Panel of genes: Collection of targeted genes thought to be relevant for particular diseases or conditions that are analysed together in a single diagnostic test. Genes present in a panel are usually linked by common biological pathways, or known disease associations.

Penetrance: the proportion of individuals that expresses an associated phenotype within a population carrying a particular variant of a gene.

Rare disease: a disease that affects less than five in 10,000 persons in Europe, as defined by the European Regulation on orphan medicinal products (Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products). In order to be registered in Orphanet, the disease must be described in at least two independent individuals, confirming that it is not an incidental association of clinical signs.

6. Filing and updates

This procedural document is updated annually and as often as necessary by the data manager. The most up-to-date version is available on the Orphanet website:

https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_Genes_inventory_R1_Ann_gen_EP_02.pdf

II. Methodology

1. Flowchart

Figure 1: Flowchart of the procedure

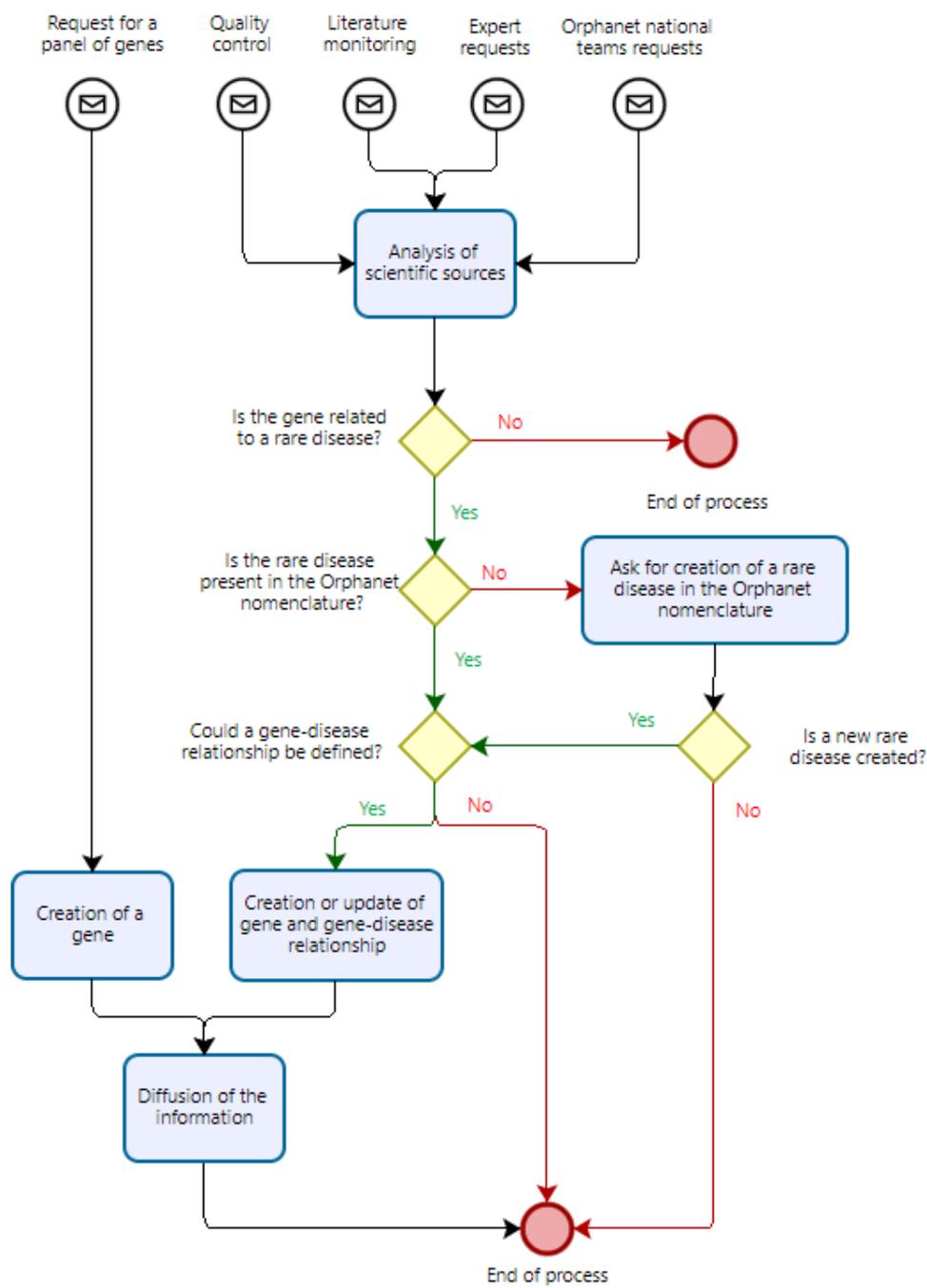
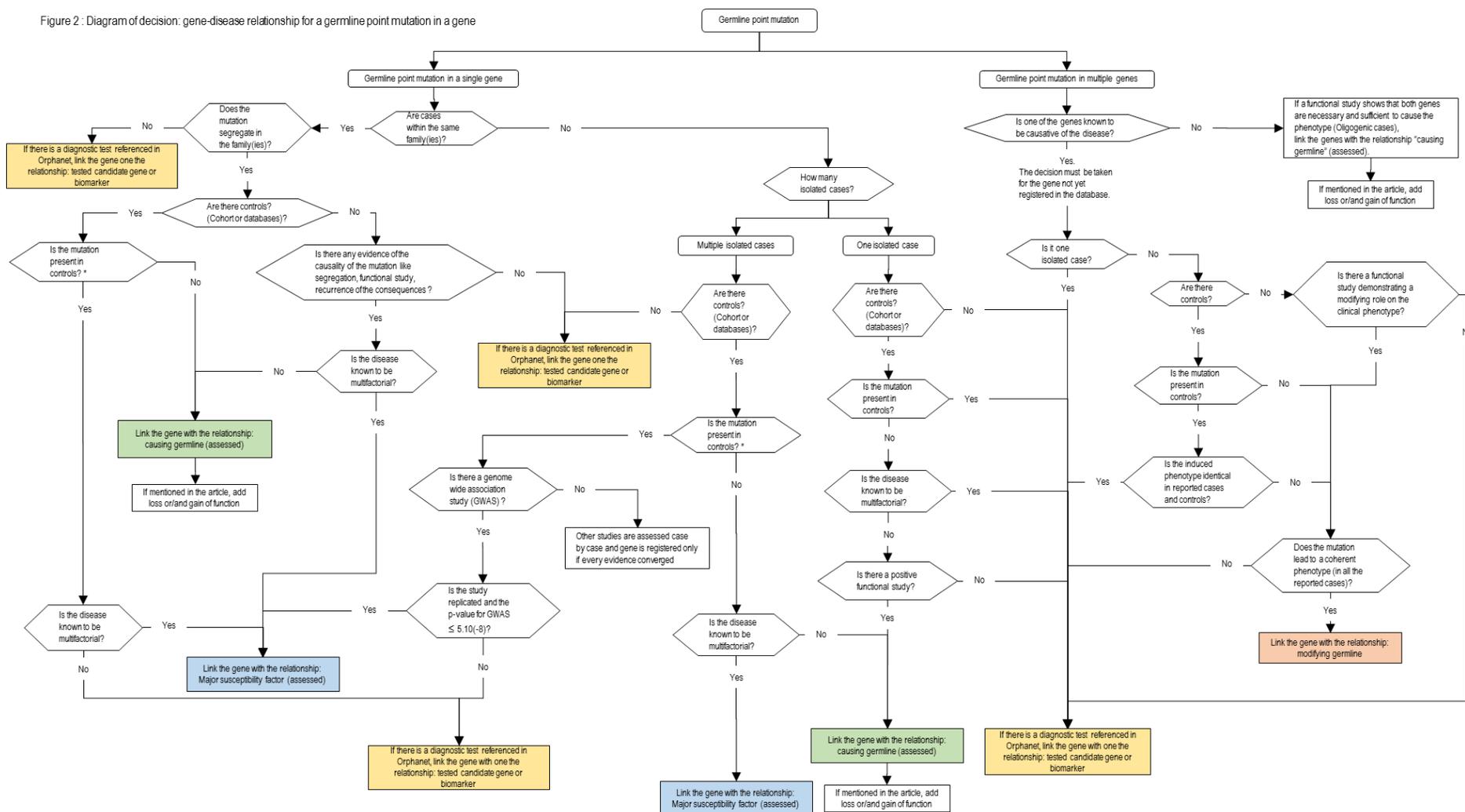


Figure 2 : Diagram of decision: gene-disease relationship for a germline point mutation in a gene



*Taking to account the evaluation of the penetrance and the frequency of the allele. See description in §2d for « disease-causing germline mutation ».

Figure 3 : Diagram of decision: gene-disease relationship for a somatic point mutation in a gene.

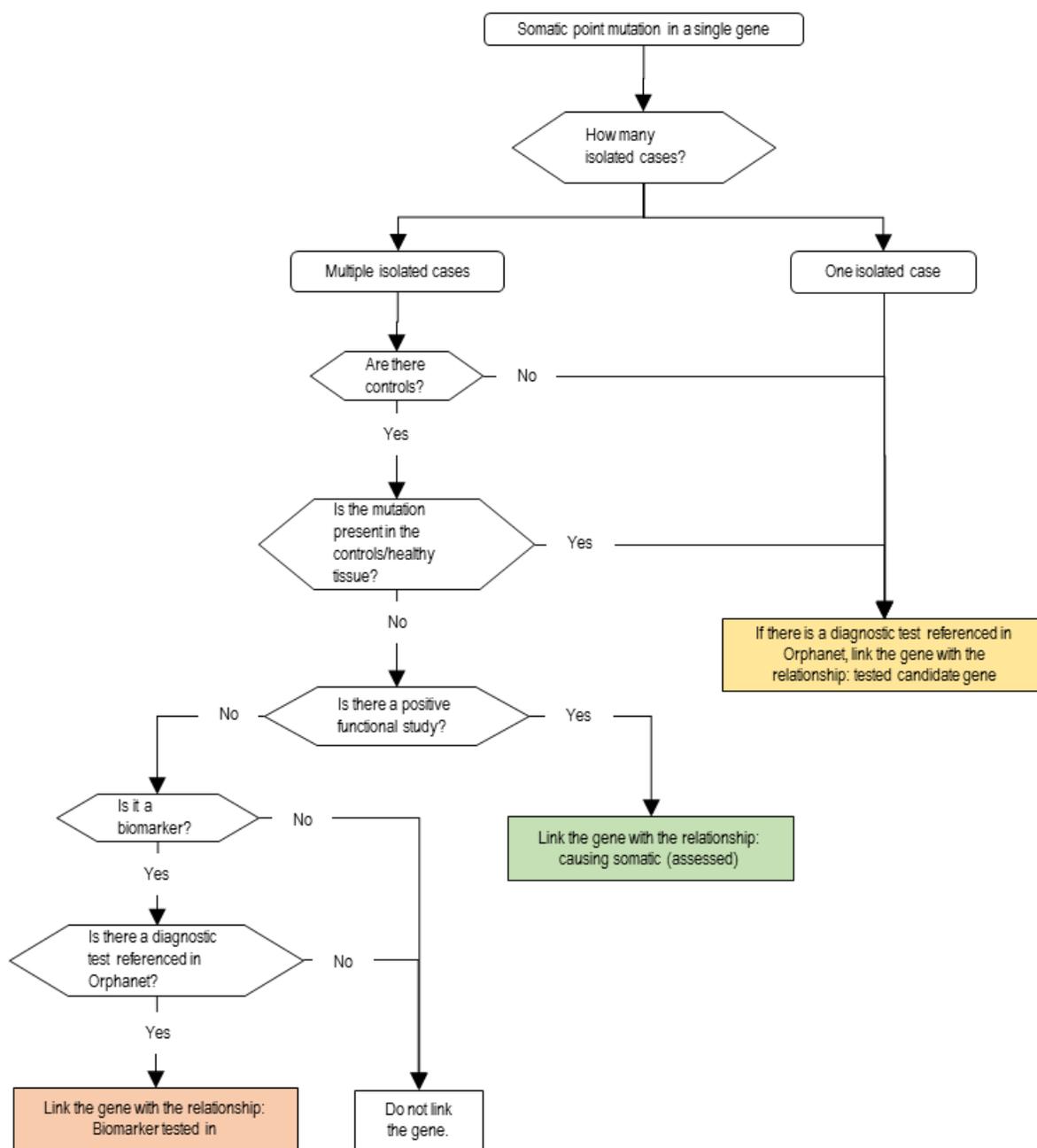
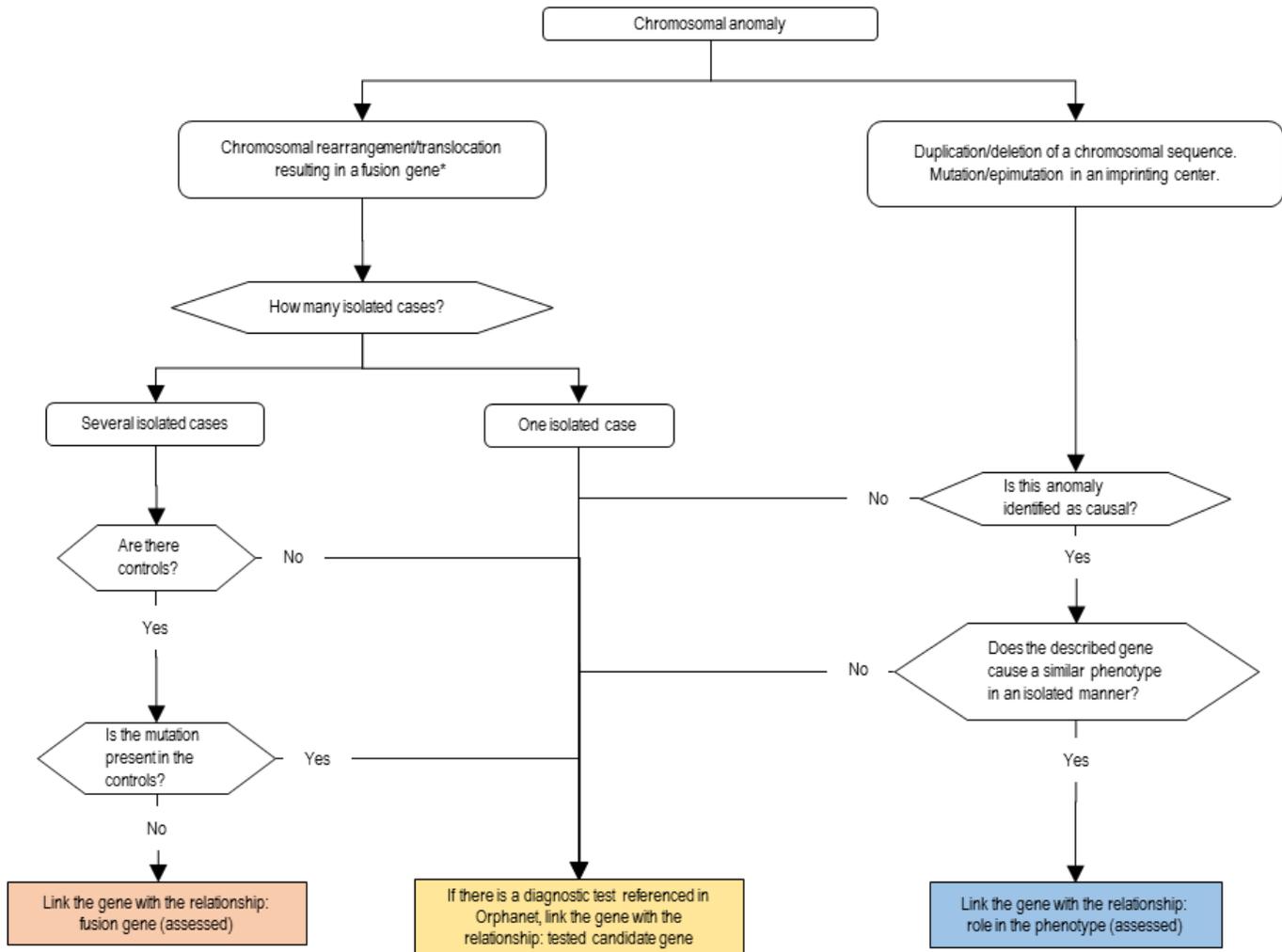


Figure 4: Diagram of decision: gene-disease relationship for a chromosomal anomaly



*see definition part of the procedure

2. Description

The update process of the Orphanet inventory of genes related to RDs is subject to a validation workflow designed to ensure that the information and data provided by Orphanet have a satisfactory level of exhaustivity, accuracy, pertinence, and reliability.

a. Requests for modification of the Orphanet inventory of genes

Different sources point the need for creation or update of a gene entry:

- literature monitoring;
- requests from experts;
- requests from national information scientists of the Orphanet consortium;
- cross-referencement with medical and/or genetic databases (such as OMIM).

b. Analysis of scientific resources

Regarding every update request, the Orphanet gene manager gathers information through a review of the scientific literature in order to determine the most appropriate action in view of the current state of knowledge and the Orphanet database rules (see below). Only peer-reviewed publications are consulted. Decisions are based only on the publications establishing the gene-disease relationship.

As an exception, candidate genes and biomarkers can be recorded without peer-reviewed publication upon an expert demand regarding a diagnostic test in use.

c. Identification of the related rare disease

Based on the publication(s), the RD is identified in the Orphanet rare disease nomenclature. Orphanet rare diseases are clinically defined. Accordingly, the decision to establish a gene-disease relationship is clinically oriented. It can differ from other genetically defined database for which Orphanet provides cross referencement in order to enhance interoperability of the dataset.

If needed, addition of a new disease to the Orphanet rare disease nomenclature is studied according to [The Orphanet nomenclature and classification of rare diseases](#) procedure.

d. Identification of the gene-disease relationship

The gene-disease relationship is qualified according to the pathogenesis of the described mutations. Thus, genes are annotated as causative (from germline or somatic mutations), modifiers, major susceptibility factors or playing a role in the phenotype (for chromosomal anomalies). When causative mutations are of germline origin, loss or gain of protein function are documented if available.

Two relationships are specifically dedicated to candidate genes and biomarkers, that are registered only if a diagnostic genetic test is made available in one of the Orphanet network's countries.

Criteria to link a gene to a disease differ from a relation to another. Decisional diagrams (figure 2, 3 and 4), validated by the Orphanet Advisory Board on Genetics, help the data manager to standardise this step. However, the scientific knowledge couldn't be easily represented in a simple diagram and number of the decisional steps depends on the data manager expertise.

Scientific validation is sometimes necessary to confirm the identification of the related disease and/or its gene

relationship. These particular cases are first submitted for scientific validation to the Orphanet nomenclature manager. If necessary, experts on a specific disease or members of the Orphanet Advisory Board on Genetics are also consulted.

The main decisional criteria are described below:

Disease-causing germline mutation(s) in

- The mutation segregates in one or more family(ies) or is identified in several isolated cases. Segregation of the putative pathogenic mutation is assessed case by case. The structure of the family, the mode of inheritance and the penetrance of the disease are carefully examined.
- The mutation is not found in the general population (absent from control cohort and/or genetic variants databases). In case of low penetrance, mutation's occurrence in control is permitted. Then, causality is assessed according to the degree of penetrance, the frequencies of the mutation in patient's and control's populations. In that case, functional studies may help to confirm the relationship.
- If there is only one isolated case, functional studies are needed to confirm the association.
- If the inheritance is digenic (or oligogenic more generally), functional studies are needed to confirm the association.

Functional studies are assessed case by case.

If available, loss or gain of the protein function is documented.

Disease-causing somatic mutation(s) in

- The mosaic mutation is found in several patients.
- The mosaic mutation is absent from controls (persons or healthy tissues).
- Functional studies confirm the association.

Functional studies are assessed case by case.

If there is only one isolated case, the gene is not registered in the database.

Concerning cancers, the causality of the mutation is considered if the mutation is an early event in oncogenesis and if an *ex vivo* or *in vivo* study confirms its oncogenicity.

Major susceptibility factor in

- The variant segregates in a family or is present in several isolated cases.
- The variant is usually found in the general population and/or the disease is known to be multifactorial.
- If the variant has been identified by a Genome wide association study (Gwas), only meta-analysis and a p-value less than $5.10(-8)$ are considered.

Currently, there is no accepted consensus for other types of association studies (ie: Whole exome sequencing). Consequently, these studies are assessed case by case and gene is registered only if every evidence converged (replicated studies, functional studies...).

Isolated cases are not considered.

Modifying germline mutation in

A modifying germline mutation is considered only if the causative gene is already identified in the patients described in the study which fulfill the following criteria:

- The variant is present in several cases.
- The variant is absent from the general population or don't lead alone to a similar phenotype.
- The consequences of the mutation are identical from a patient to another.

Isolated cases are not considered.

Part of a fusion gene in

Genes fused by chromosomal rearrangement are registered only if the molecular study show evidence of the causality of the fusion and if they fulfill the following criteria:

- The fusion is observed in several cases.
- The fusion is absent from controls.

Role in the phenotype of

A gene with a major role in the phenotype of a chromosomal rearrangement is linked to a disease only if:

- A mutation hitting only this gene gives rise to similar phenotypic consequences when it occurs in isolation.

Mutations or epimutations in an imprinting center are also considered as having a major role in a phenotype as the phenotypic consequences are similar to the one of a chromosomal rearrangement.

Candidate gene tested in or Biomarker tested in

Candidate genes or biomarkers are registered only if a genetic test is available in at least one of the Orphanet network's countries.

Depending on the purpose of the test, either for diagnosis or to monitor disease activity, the candidate gene or biomarker relationship is attributed.

If any of these relationships can't be implemented, the gene is not introduced in the Orphanet inventory of genes. As an exception, genes claimed in a Panel indicated for a diagnostic test can be registered in the Orphanet inventory of genes without any relationship to rare disease.

e. Update of the Orphanet database

When a request fulfills all the inclusion criteria, the gene is introduced in the Orphanet inventory of genes. Computational cross-referencing with international genetic databases is completed and allows to inform on:

- Gene nomenclature, including main name, symbol, synonyms and previous names and symbols according to HGNC.
- Typology of the genetic entity (gene with protein product, disorder-associated locus, non-coding RNA).
- Chromosomal location.
- Mappings with HGNC, OMIM, UniProtKB, GenAtlas, Ensembl, Reactome and IUPHAR-DB, LOVD.

Every gene-disease relationships that have been curated and fulfill our inclusion criteria is mentioned as «  assessed ». Without this note, the gene-disease relationship is still waiting for expert validation.

Candidate genes are never « assessed » since there is no scientific evidence to attest the gene-disease relationship.

f. Quality control

To guarantee that the Orphanet genetic database is consistent with the rules and process described in this document, quality controls are performed regularly.

An automatic mapping with HGNC database is monthly performed to ensure the interoperability with other genetic international databases.

III. Availability of data

Orphanet inventory of gene is released at a variable frequency depending on the channel of dissemination, daily for the website (note that references are not displayed on the website) - www.orpha.net -, monthly for the Orphanet download platform - www.orphadata.org - and annually for the Orphanet Rare Disease Ontology – [ORDO](#).

For any questions or comments, please contact us: contact.orphanet@inserm.fr

Editor of this procedural document: Antoine Marmignon - This procedural document has been approved by: Annie Olry -
Quality control: Charlotte Gueydan

The correct form when quoting this document is:

Procedural document on the Orphanet inventory of genes related to rare diseases, Orphanet, June 2021, Number 02

https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_Genes_inventory_R1_Ann_gen_EP_02.pdf