The SysNDD Documentation

Bernt Popp, Christiane Zweier

2022-06-07

Contents

Preface

This documentation is intended to describe the SysNDD[1] project and provide instructions for regular user- show to use the too and for curator status users how to perform reviews and how to enter data.

History of SysID and SysNDD

SysNDD is based on its predecessor SysID[2] which had been published in 2016 (Kochinke et al., 2016). Christiane Zweier has been involved in establishing and updating SysID from its start in 2009. She has since performed and coordinated curation and regular updates.

The PHP based SysID web tool (Yii 2 framework) was however not further developed and maintained besides necessary bugfixes. After the maintenance agreement for the original server at the CMBI at Radboud University in Nijmegen ran out, the installation was moved to a virtual server at the Department for BioMedical Research (DBMR) at the University Bern. The former link from the initial publication is re-directed so it still works. The legacy code base was updated to allow installation and security fixes and to be uploaded to a GitHub repository (SysID)[3].

In 2019 the chance arose to integrate the SysID curation effort with the Orphanet resource, supported by ERN ITHACA. In the process of aligning the curation and naming conventions for genes, diseases and phenotypes we decided to redesign the database and web tool.

The SysNDD concept

SysNDD contains a manually curated catalogue of published gene-disease-associations implicated in neurodevelopmental disorders (NDD).

To allow interoperability and mapping between gene-, phenotype- or disease-oriented databases, we center our approach around curated gene-inheritance-disease units, so called entities. These entities are classified into different confidence status (categories: “Definitive”, “Moderate”, “Limited”, “Refuted”) according to the degree of underlying scientific evidence. Furthermore, manually curated information on associated phenotypes is provided.

The entries in SysNDD are currently updated every 3-4 months and can be utilized for a broad spectrum of tasks from both research and diagnostics.

One of our goals is to incorporate the SysNDD data[4] into other gene/ disease-relationship databases like the Orphanet ontology (first results: id-genes.orphanet.app[5]).

Bernt Popp (scientist at the Institute of Human Genetics at the University Hospital Leipzig, Germany) developed and programmed the SysNDD tool and will be integrating further functionality including variants associated with entities in future updates.

Acknowledgments

The current SysNDD database development is supported by:

[1] https://sysndd.dbmr.unibe.ch/
[2] https://www.sysid.dbmr.unibe.ch/
[3] https://github.com/berntpopp/SysID
[4] https://sysndd.dbmr.unibe.ch/
[5] https://id-genes.orphanet.app/ithaca/

- DFG (Deutsche Forschungsgemeinschaft) grant PO2366/2-1 to Bernt Popp[6].
- DFG (Deutsche Forschungsgemeinschaft) grant ZW184/6-1 to Christiane Zweier[7].
- ERN ITHACA[8] through Alain Verloes[9].

The previous SysID database and data curation was supported by:

- The European Union's FP7 large scale integrated network GenCoDys (HEALTH-241995) to MA Huynen and Annette Schenck.
- VIDI and TOP grants (917-96-346, 912-12-109) from The Netherlands Organisation for Scientific Research (NWO) to Annette Schenck.
- DFG (Deutsche Forschungsgemeinschaft) grants ZW184/1-1 and -2 to Christiane Zweier[10].
- the IZKF (Interdisziplinäres Zentrum für Klinische Forschung) Erlangen to Christiane Zweier[11].
- ZonMw grant (NWO, 907-00-365) to Tjitske Kleefstra.

1 Curating gene-disease relationships

As the name implies a rare disease affects only very few individuals. However, there are many unique causes of rare diseases, thus many individuals are affected by such a disease. Due to the rarity of each single entity, effective management, surveillance and treatment is challenging So is finding the correct diagnosis, which is often described as the "diagnostic odyssey".

Rare diseases often have a genetic cause, making high-throughput sequencing (next-generations sequencing; NGS) a central part of finding the molecular diagnosis.

1.1 Neurodevelopmental disorders

Neurodevelopmental disorders (NDD) affect about 2% of children. They represent a clinically and genetically extremely heterogeneous disease group comprising amongst other developmental delay (DD), intellectual disability (ID) and autism spectrum disorder (ASD) and developmental and epileptic encephalopathies (DEE).

1.2 Genetic heterogeneity

The huge genetic heterogeneity is evident when looking at the published gene-disease associations over time.

Thus the question arises:

> How can we keep track of this fast development and have the information at hand when we need it in the clinic or when analyzing sequencing data?

While the answer to this question is easy:

> We need curated databases to catalogue and summarize the wealth of published information.

The task at hand is not only laborious but also requires expertise and consistence.

[6] https://orcid.org/0000-0002-3679-1081
[7] https://orcid.org/0000-0001-8002-2020
[8] https://ern-ithaca.eu/
[9] https://orcid.org/0000-0003-4819-0264
[10] https://orcid.org/0000-0001-8002-2020
[11] https://orcid.org/0000-0001-8002-2020

1.3 Expert curation

In our opinion, the curation of gene-disease relationships in rare disease such as NDDs requires clinical and scientific proficiency in the respective field. This implies that clinician scientists involved in counseling, diagnostics, and research of NDD are predestined for this task.

To reduce workload and dependence on single experts, a distributed effort in larger consortia and collaboration between different work groups is needed.

In the course of updating SysID we had the great chance to contribute our data to Orphanet to create a European ID/NDD specific reference list. With support from the „ITHACA Workgroup: intellectual disability“ (id-genes.orphanet.app[12]) in 2019 we started working with the Orphanet team which is part of the Gene Curation Coalition (GenCC).

Additionally, we are able to recruit expert curators from ERN ITHACA[13] to contribute to re-curation of old data and updating new data in SysNDD.

1.4 Technical concepts

In addition to a pool of experts, the right tools are needed.

We defined “gene-inheritance-disease” units as “entities” which represent the central curation effort. The components of these entities are normalized using widely used and standardized ontology terms (e.g. HGNC identifier for genes, OMIM or MONDO for disease and inheritance from HPO). This allows interoperability and linking to other data sources.

Based on this concept we developed a new database scheme, which allows entities to be systematically and reproducibly cataloged. The database is abstracted into a JSON API, which allows structured programmatic access to the underlying data.

Finally, the API feeds the web tool which can be used to easily search, filter, download and visualize the database contents in modern webbrowsers.

1.5 Outlook

- The SysNDD database will improve the understanding and curation of rare NDD entities.
- SysNDD will enable systems biology and network analyses.
- Our long-term goal is incorporation of the high-quality, manually curated SysNDD data into European and international gene disease relationship databases,
- thus, improving diagnostics and care for individuals with rare NDDs.

2 Web tool

3 API

[12] https://id-genes.orphanet.app/ithaca%5D
[13] https://ern-ithaca.eu/

4 Database structure

---

5 Curation criteria

5.1 Definitions

Intellectual disability (ID) and neurodevelopmental disorders (NDD) are defined in the scope of SysNDD as follows:

- Early onset neurodevelopmental delay and cognitive impairment (severe ID to learning difficulties)
- Regression/ neurodegeneration in the first years of life with or without prior developmental delay
- Disorders with cognitive impairment in a significant (ca. >10%) fraction of individuals

5.2 NDD Definitive entities

Inclusion criteria for Category 1 ("Definitive"):

1. Publication required (no grey literature like conference abstracts or personal communication; manuscripts on preprint servers can be considered individually but only entered through their DOI in the comment field until published with PMID, when they should be updated)

AND

2. Clear-cut frequency (no further criteria needed)

- >= 10 cases with *de novo* variants
- >= 5 autosomal-recessive families
- >= 3 families with X-chromosomal variants

OR

3. Cumulative evidence

- 1 strong frequency criterium

PLUS

- 1 strong genetic or 1 strong clinical criterium

OR

- 2 further strong (genetic and/or clinical) criteria in case of only 2 families with recessive inheritance

OR

- >= two moderate criteria

5.2.1 Strong criteria

Strong frequency criteria:

- >= 3 patients with *de novo* variant
- >= 2 families with bi-allelic truncating variants
- >= (2-)3 families with bi-allelic missense variants
- >= 2 families with X-chromosomal variants

Strong genetic criteria:

- recurrence of a variant
- clustering of variants
- *de novo* truncating variants in a gene intolerant to loss-of-function variants (gnomAD constraint score)

Strong clinical criteria:

- Homogeneous phenotype
- Presence of specific/distinct clinical aspects (e.g., recognizable facial gestalt; rare specific malformations; pattern of multiple malformation; characteristic MRI anomalies; specific metabolic/enzymatic anomalies)

5.2.2 Moderate criteria

- Multigenerational segregation of variants
- Functional tests
- Gene involved in a pathway/complex where variants in other subunits are associated with a similar phenotype
- *De novo* missense variants in a gene intolerant to missense variants (gnomAD constraint scores)

5.2.3 Possible negative criteria

These should be included into consideration in borderline cases.

- Age of first publication(s) without further confirmatory reports in the meantime
- Publication quality and journal or genetics expertise "doubtful"
- New evidence against gene and/or variants: e.g., constraint scores, frequencies in gnomAD

5.3 NDD Moderate and Limited entities

These categories include the previous category of "candidate genes" and are now split into criteria for entity categories 2 ("Moderate") and 3 ("Limited"):

1. Must be published (no private, in-house candidate lists)

AND

2. ID indicated, but criteria not sufficient for category 1, examples:

a. Limited genetic evidence

- < 3 cases with *de novo*, different variants and non-specific NDD phenotype

- 1 recessive family with truncating variant or <= 2 recessive families with missense variants (category 2 or 3 depending on number of affected and tested individuals per family, functional evidence and homogeneity of phenotype etc.)
- candidate gene from translocation or larger deletion
- reports of enzymatically confirmed patients with specific metabolic disorders but without genetic mutation confirmed

b. Limited clinical evidence

- not much evidence for ID, e.g. reported as ADHD or ASD or neurological disorder without clearly reported low IQ and ID
- known disorder, but only single patients reported with ID
- motor developmental delay without evidence for cognitive impairment
- clear neurodegenerative course without ID or cognitive delay present in the first years
- lethal before ID might be evident, although e.g. brain malformations or metabolic abnormalities might point to ID
- ID reported in other, similar disorders caused by mutations in the same pathway/complex but not (yet) in association with this particular gene (e.g. Fanconi anemia)

c. Limited combined genetic and clinical evidence

- Gene enriched for de novo or rare deleterious variants in large NDD cohorts or meta-studies, no further details

5.3.1 Exclusion criteria

1. Published as candidate gene only based on function or experimental results but without variants reported in humans

AND/OR

2. Only 1 *de novo* case from longer ago without further evidence and gene tolerant towards missense and/or loss-of-function variants according to gnomAD constraint scores

AND/OR

3. Only 1 sporadic case with bi-allelic variants and without any further supporting evidence such as segregation in other family members, functional tests, similar phenotypes in other patients with variants in genes from the same pathway, etc.

5.3.2 When to choose category 2 ("Moderate")?

Too good for category 3 ("Limited") but not good enough for category 1 ("Definitive")

Examples:

- Recurrent de novo variant in 2 individuals with a similar phenotype
- Bi-allelic or X-chromosomal truncating variant segregating in >= two generations of a large family
- Convincing functional evidence
- 1-2 patients with convincing variants in a gene which is in the same complex/pathway with other known disease genes and phenotype fits (e.g. CDG syndrome)

5.3.3 Special case: non-NDD entities

Some genes are associated with multiple entities. Among these entities there might be some without ID as a clinical feature. These non-NDD entities will be included in SysNDD but they will not be classified to any of the categories. Instead, they are tagged with "n.a." (not applicable).

6 Re-review instructions

The goal of the SysNDD "Re-Review" effort is to update and standardize the SysID entities collected during the past years to enable better integration into and interoperability international with gene curations.

6.1 Re-review tool usage

We created Reviewer status accounts for participating scientists.

6.1.1 Login

You can log into your account by pointing your browser to https://sysndd.dbmr.unibe.ch/ and then clicking the "Login" link on the right side of the menu:

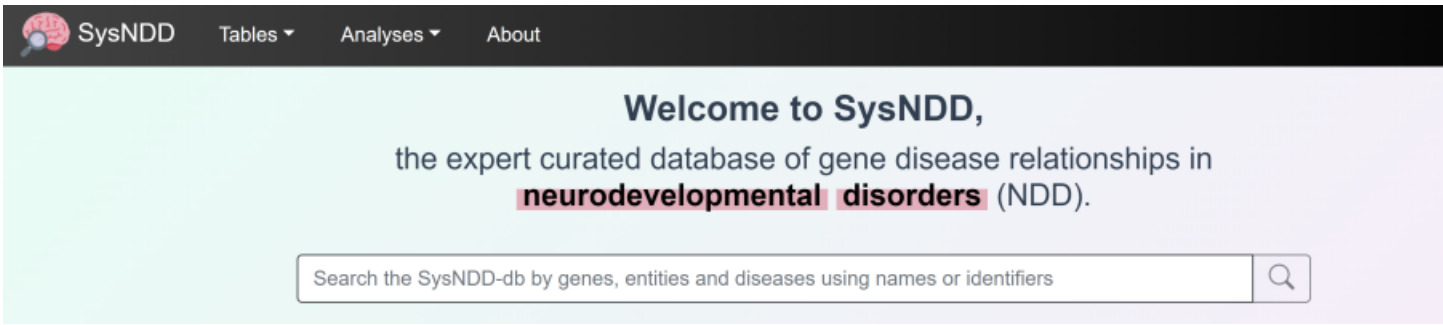


Figure 1: Login menu

On the Login page enter your credentials and press the Login button:

After successful login, you will be redirected to the start page and the navigation bar will show new links depending on your account privileges:

Your login token (JWT; JSON Web Token) is valid for 1 hour, after which you will be logged out. You can however always refresh the time by clicking the link in the user menu. The website will warn you at 5, 3 and 1 minutes before log out.

6.1.2 Review page

Click the "Review" link to your personal "Re-Review" site:

The "Re-Review" page is structured as a table enriched with information and controls.

These show you the number of entities assigned to your account

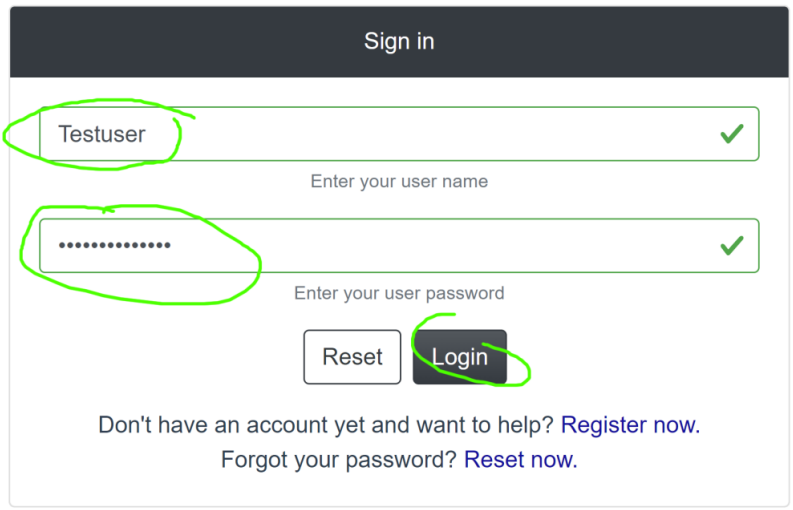


Figure 2: Login page

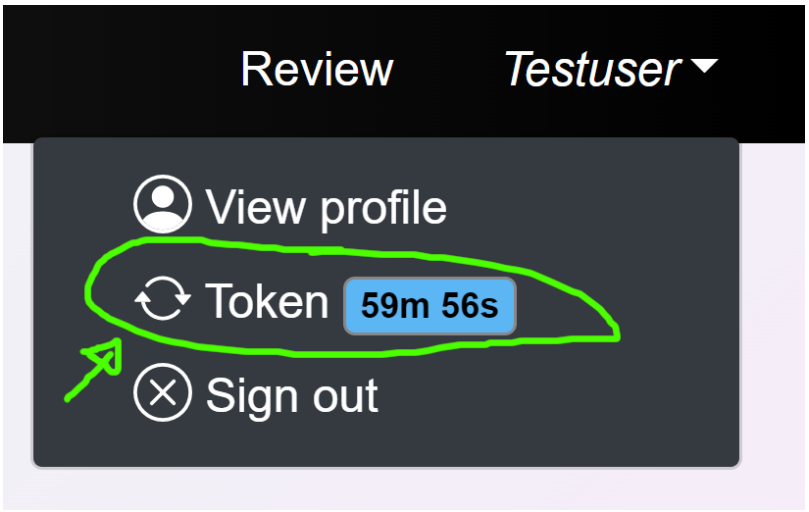


Figure 3: Login token menu

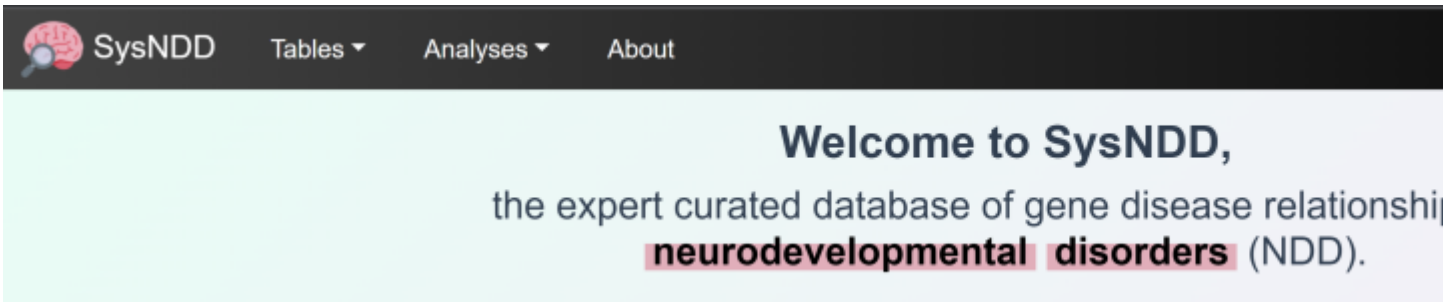


Figure 4: Review page menu

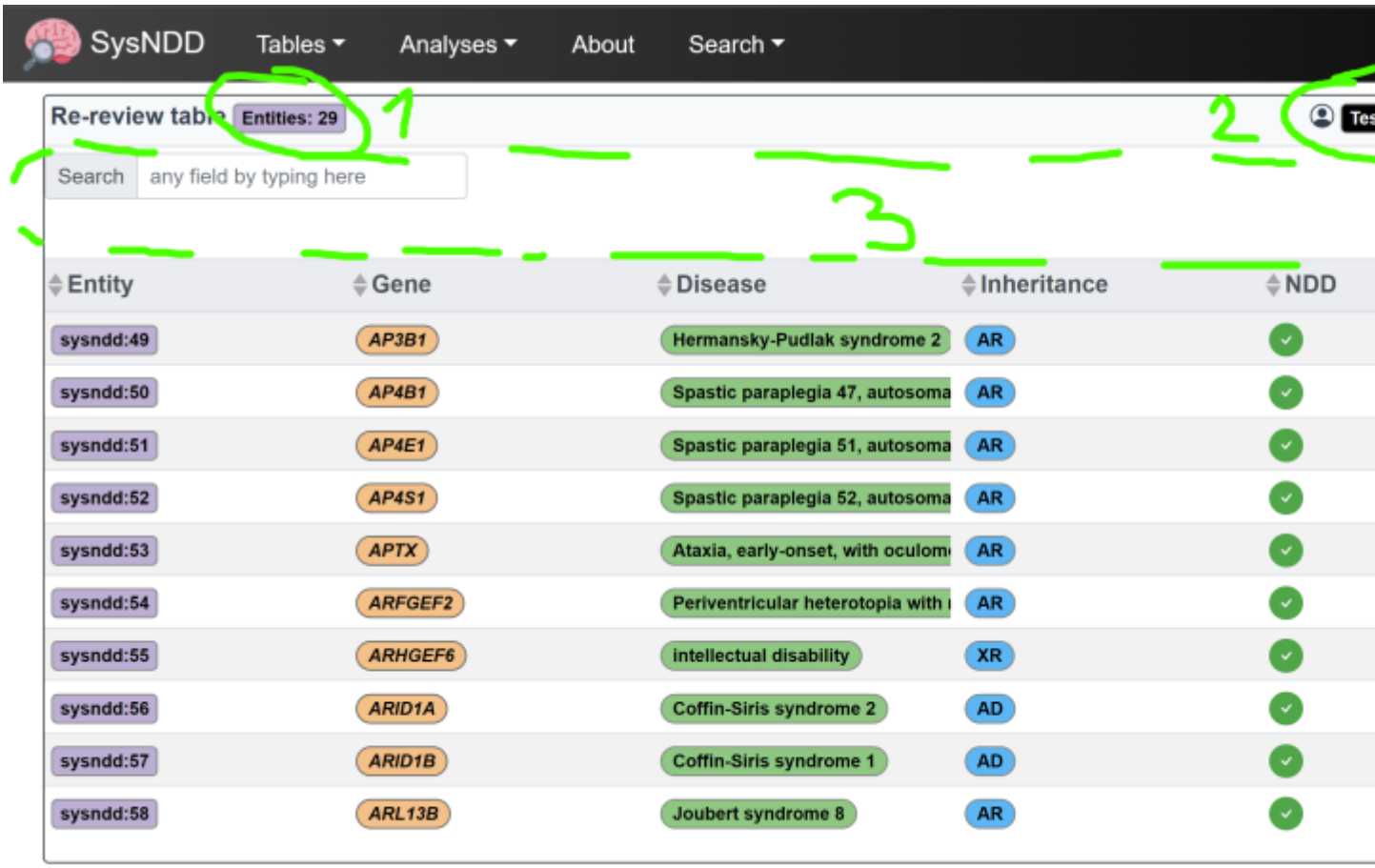


Figure 5: Review page

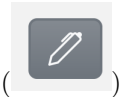
(1) your account information status specific controls (e.g. switching to "Curator" mode, applying for a new batch of entities)

(2) menu items to filter/ navigate the table

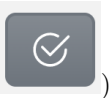
(3) and finally, the table with the entity information and

(4) controls to review and change the information:

By clicking the action buttons, you can open 3 different windows to change the entities review:

(1) entities review ()

(2) the status ()

(3) and to submit your work ()

6.1.3 New Review edit

In this window you have - the possibility to change/adapt or completely rewrite the current synopsis **(1)**, - add, or remove phenotype associations **(2)**, - add or remove publications from the review by PMID **(3)** - and add/ edit fitting GeneReviews articles by PMID **(4)**. - Finally, you can add a comment to your review for the Curator later approving this entities changes **(5)** and - save your review **(6)**.

By clicking on the little question marks you can show help messages for each item:

These help instructions are:

Synopsis: Short summary for this disease entity. Please include information on: a) approximate number of patients described in literature, b) nature of reported variants, b) severity of intellectual disability, c) further phenotypic aspects (if possible with frequencies) d) any valuable further information (e.g. genotype-phenotype correlations).

Examples:

> *de novo* truncating or missense variants in > 20 individuals: variable ID (mild to severe), 50% short stature and microcephaly, 30% seizures, non-specific facial dysmorphism, variable cardiac and renal anomalies in some
>
> bi-allelic truncating variants in 7 individuals from 3 families: severe ID, microcephaly, seizures in 3/7, MRI anomalies

Phenotypes: Add or remove associated phenotypes. Only phenotypes that occur in 20% or more of affected individuals should be included. Please also include information on severity of ID where available and applicable.

Publications: No complete catalogue of entity-related literature required! If information in the clinical synopsis is not only based on OMIM entries, please include PMID of the article(s) used as a source for the clinical synopsis.

GeneReviews: Please add PMID for GeneReview article if available for this entity.

Comment: Additionally add information about your review potentially helpful to the curator approving the entity later.

6.1.4 New Status edit

In this window you can propose

- to change the entities association confidence category **(1)**,
- suggest it's overall removal **(2)**,
- add a comment for your change suggestions for the Curators to better understand the proposal **(3)** and
- save your work **(4)**:

Figure 6: Submit re-review modal

6.1.5 Submit Re-review

The last action window is just to confirm that you are satisfied with your work and would like to submit it for curation:

After clicking this button, the entity will disappear from your list. And you can proceed with the remaining entries until no entity is left in your list.

6.2 Re-review curation

6.2.1 Definitive association status

1. Check if category 1 ("Definitive") is correct or shift status to category 2 ("Moderate") or 3 ("Limited"), where appropriate

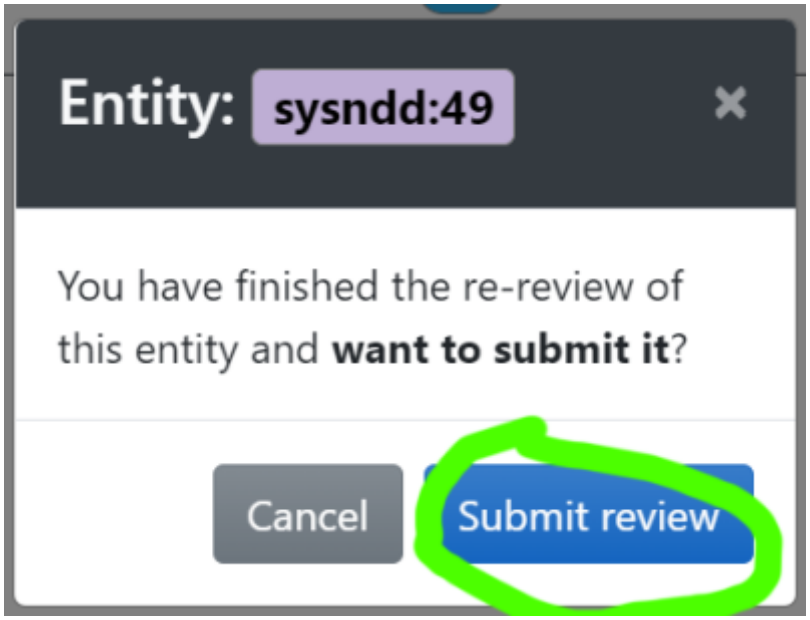


Figure 7: Submit re-review modal

2. Check and revise gene-related entities regarding diseases/inheritance patterns (ID and non-ID disorders) –> non-ID disorders will not go into any of the categories but will be tagged with "n.a." (not applicable)
3. Check and revise associated phenotypes: select HPO terms from the list, only use HPO term if this specific aspect is present in approximately >= 20% of patients. Please also check and revise severity of ID using HPO terms. If ID is very variable, select all appropriate ID terms (e.g. severe, moderate, mild, borderline)
4. Check references (OMIM, PMID, GeneReviews). References do not have to be complete but should be sufficient to give a good impression on the mutational and clinical spectrum. Add references where it would add to the picture.
5. Check and revise clinical synopsis: it does not have to contain everything that is known but should give a short and comprehensive picture on:

- which data the gene and disease category were chosen on and
- the molecular and clinical picture.

Please include information on:

a) approximate number of patients described in literature,
b) nature of reported variants,
c) severity of intellectual disability,
d) further phenotypic aspects (if possible with frequencies),
e) any valuable further information (e.g. genotype-phenotype correlations)

Examples:

> *de novo* truncating or missense variants in > 20 individuals: variable ID (mild to severe), 50% short stature and microcephaly, 30% seizures, non-specific facial dysmorphism, variable cardiac and renal anomalies in some
>
> bi-allelic truncating variants in 7 individuals from 3 families: severe ID, microcephaly, seizures in 3/7, MRI anomalies

6.2.2 Moderate and Limited association status

- Check if inclusion criteria for candidate genes are still fulfilled or if it should be deleted from the list ("Refuted")
- Check if candidate status is still correct and sort it into Category 2 ("Moderate") and 3 ("Limited") (or reclassify to 1 ("Definitive"), if applicable)
- Check, if associated phenotype still fits
- Check, if references are correct, if there is any new published information and modify clinical synopsis where appropriate
- Clinical synopsis can be very short for candidate genes
- no associated phenotypes (HPO terms) and frequencies are needed for candidate genes, but could be helpful

Examples:

de novo missense variants in 2 individuals: autism, ID in 50%

bi-allelic missense variant in 2 affected individuals from 1 family: moderate ID, MRI anomalies

6.2.3 Refuted association status

- Check if there is current evidence against this gene association (e.g. few truncating variants described in old publications before gnomAD constrain scores and the gene now has a pLI of 0; genes reported in a family with later report of another cause etc.)

References

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M. A. W., Kleefstra, T., Kramer, J. M., Webber, C., Huynen, M. A., and Schenck, A. (2016). Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *American Journal of Human Genetics*, 98(1):149–164.