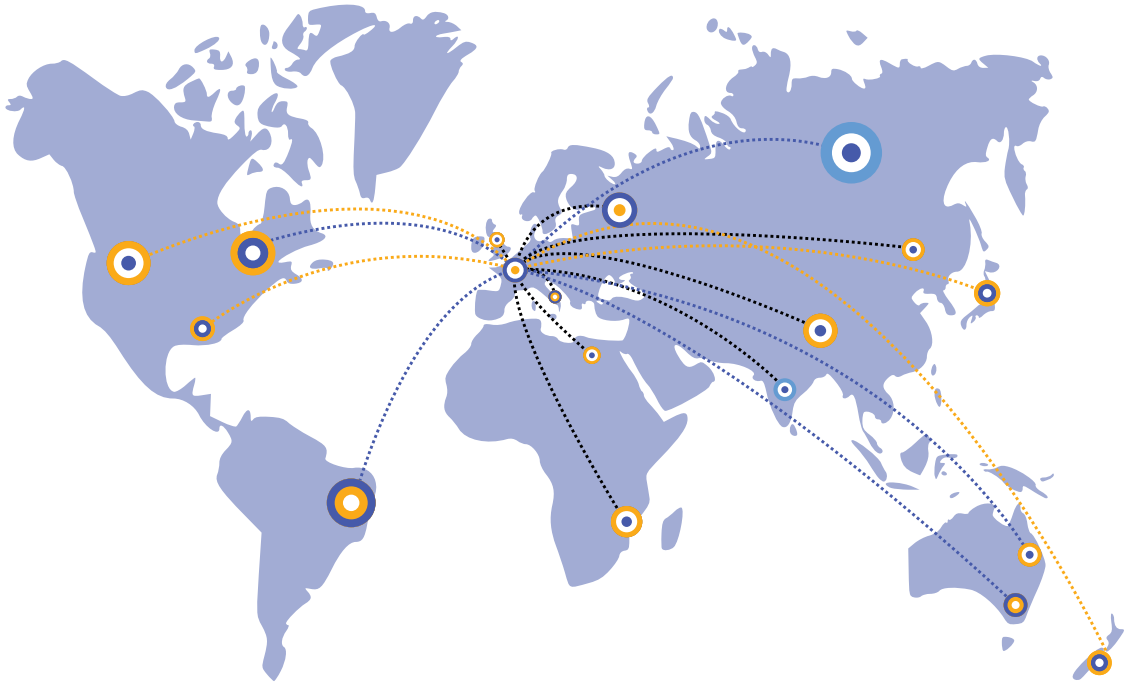
Guidance for Registering hPSC Lines Associated with Rare Disease

2023-12-22

Version 0.1



The Human Pluripotent Stem Cell Registry (hPSCreg®) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 101074135

Contents

Introduction

The human pluripotent stem cell registry (hPSCreg®) is a not-for-profit on-line resource, supported by multiple rounds of research funding from the European Commission since 2007[1]. The resource is freely accessible to all at https://hpscreg.eu/, and it includes a registry for human pluripotent stem cell (hPSC) lines, a database of projects where registered hPSC lines are used in EC-funded research, an overview of the legal status of embryonic stem cell research in different countries, and a database of clinical studies that use hPSC-derived cell products for interventional clinical trials.

The cell line registry collects data on the biological properties of the cell line, its derivation and the ethical provenance of the biological material used to generate the cell line. Based on this body of data collected on each cell line, registered cell lines may be eligible to receive certificates from hPSCreg that attest to the biological properties and ethical provenance of the cell lines. These certificates are recognized by the European Commission funding programmes to fulfil the ethical requirements for EC-funded research. Although the registry is focused on data, it provides information where the physical line is and its distributor.

Within the European Union (EU), a rare disease (RD) is defined to be a disease that affects less than 1 in 2000 persons. An estimated 30 million people in the EU are affected by more than 6000 rare diseases[2]. The EU has a strategic objective to improve the lives of RD patients by increasing their access to diagnosis, information, and care[3]. Due to their rarity, resources and information about rare diseases can be fragmented and scattered between multiple online resources, making them difficult to find and access.

hPSC lines derived from RD patients or other lines associated to RD, such as those gene-edited to contain genetic mutations associated to RD, can be valuable starting materials for research into the disease mechanism and help researchers to identify new potential avenues of treatment. For example, hPSC-derived neuronal cells from gene-edited lines can be used to model biological processes in these lines, compared to their unedited isogenic counterparts. Some major aspects of RD have been uncovered only by use of patient-derived hPSC lines, while corresponding RD animal models (e.g. genetically modified mice and rats) failed to either sufficiently recapitulate the human disease mechanisms or were a “dead end” in the development of new treatments[4]. Alternatively, hPSC lines derived from RD patients can be differentiated into specialized cell types, such as neurons or cardiomyocytes, to develop cell-based assays for screening drug treatment candidates. More recent developments in hPSC culturing technology also allows to grow them into miniature organs, also known as organoids, which resemble a more *in vivo*- like tissue architecture[5] .

To make RD hPSC lines accessible resources for the scientific community, the cell line data needs to be recorded, stored and made easily searchable. hPSCreg provides the infrastructure for cell line generators to assign unique identifiers to their lines and to record basic anonymous information on the donor and data on the derivation and

biological properties of the line, as well as information on the ethical provenance and usage of the cell line.

In this document, we provide guidance for registering hPSC lines that are associated with rare disease, in order to make these valuable resources more easily findable and accessible to the scientific community at large, thus accelerating rare disease research.

Enhancing the FAIRness of hPSC lines associated with Rare Disease

The Registry collects a number of data fields using metadata, ontologies and identifiers, to increase the Findability, Accessibility, Interoperability and Reusability of the cell line data[6] . An overview of the cell line data that is collected upon cell line registration is depicted in Figure 1. For specific information on how to register cell lines at hPSCreg, please consult the Cell Line Registration Quick-Start Guide and the Mandatory Fields documents available from the hPSCreg website.

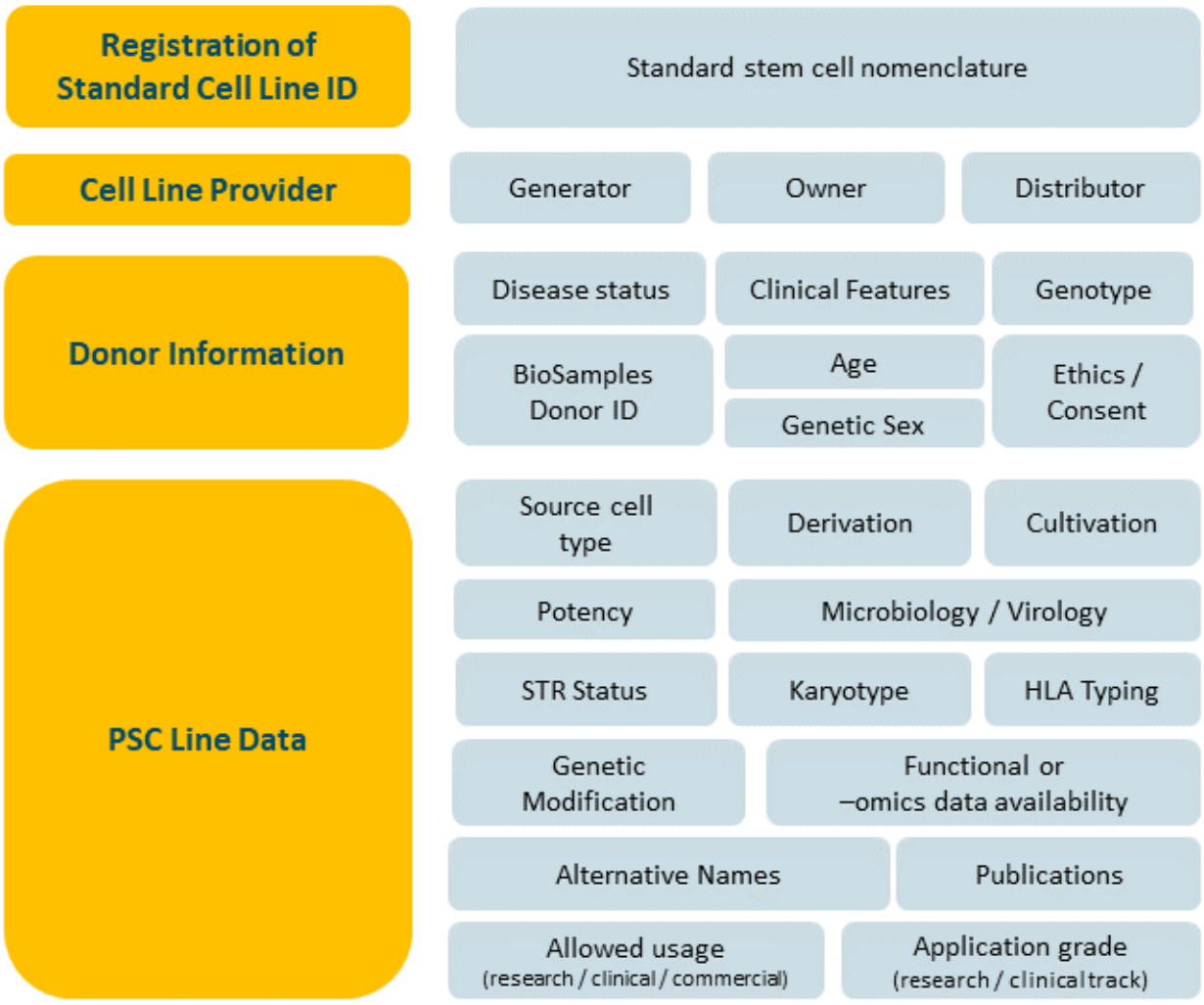


Figure 1. hPSCreg data fields for stem cell line registration.

Although not all data fields in Figure 1 are mandatory for cell line registration, please make the effort to provide as much stem cell data as possible, since a complete dataset makes the cell lines more easily findable by others. It also helps to display the uniqueness of the research material and the amount of characterization that has been completed, maintaining a high quality standard within the RD field.

To maximize the FAIRness of cell lines associated to rare disease, special attention should be paid to the following data fields.

In the “Donor information” tab of the cell line data entry, under the heading “Phenotype and Disease related information (Donor)”, there is the opportunity to enter the disease and phenotypes associated to the donor (Figure 2).

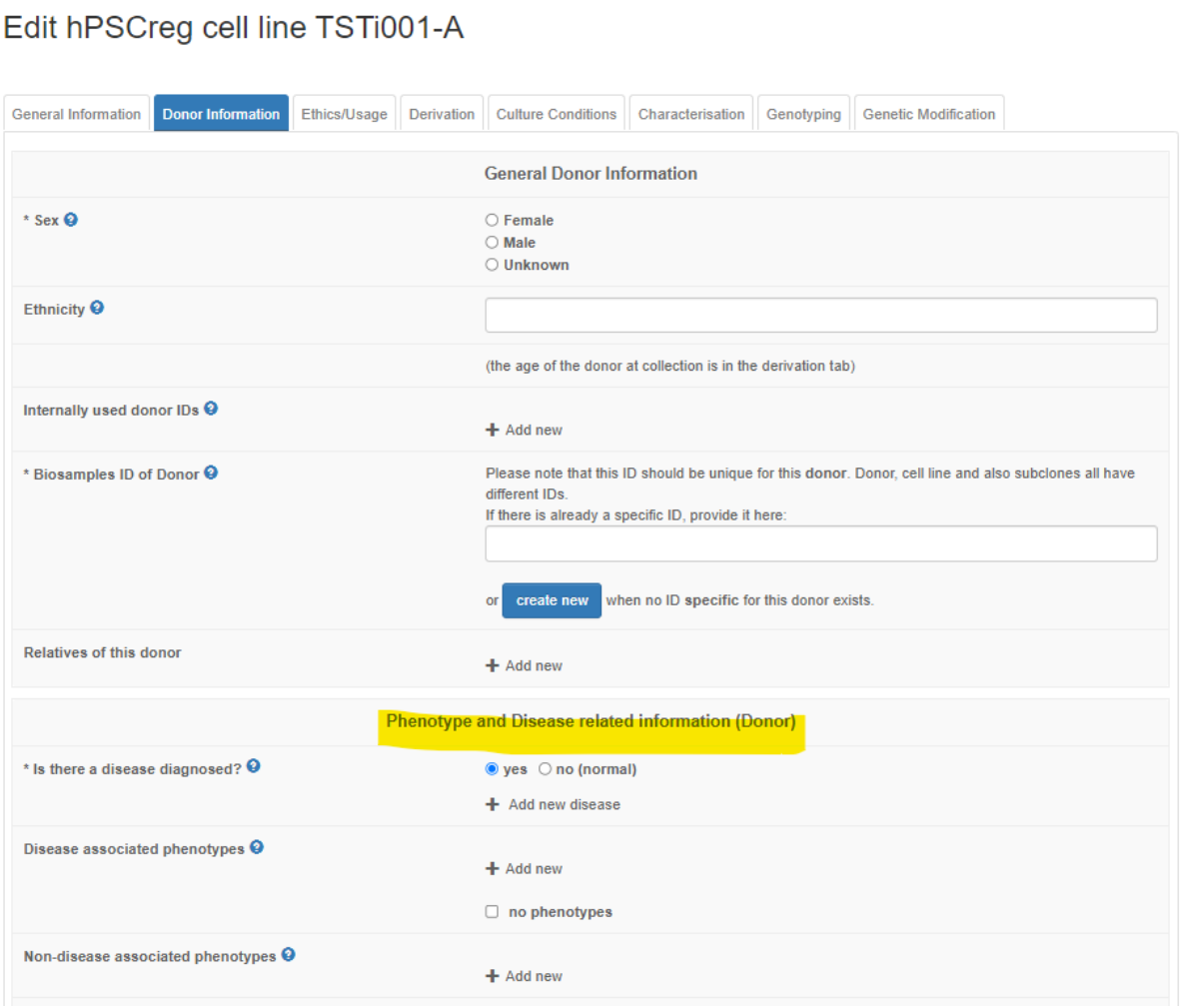


Figure 2. hPSCreg Donor Information Tab.

Recording the rare disease associated with the donor

In order to make diseases findable, the Registry uses controlled vocabularies, like ontologies, to record the diseases. In this way, the disease is recorded using standardized terms, and this facilitates search and discovery of this data. To the question “Is there a disease diagnosed?”, answer “yes”. A new box of questions will appear (Figure 3). Choose an ontology term for the rare disease by first clicking on the “enter disease” button.

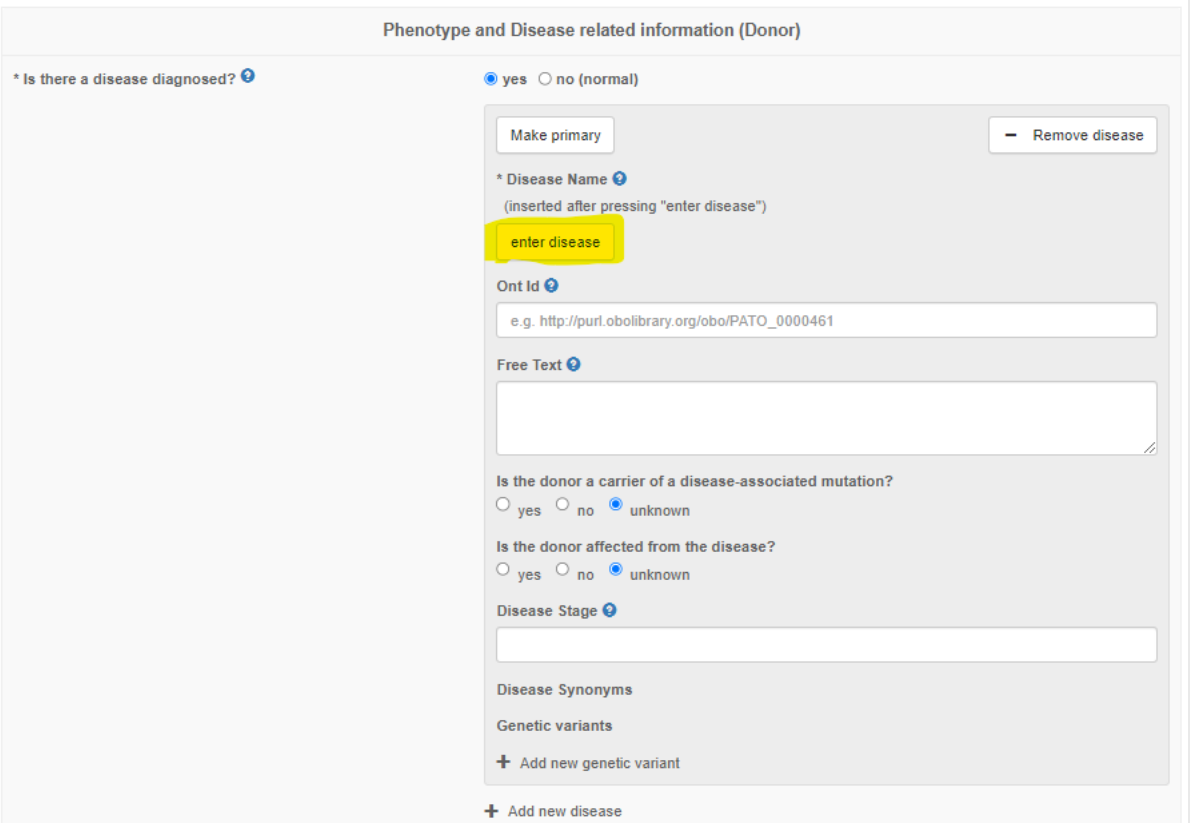


Figure 3. Entering disease information

Then search for the most specific rare disease term that applies to the donor. In the example, “Dravet syndrome” was typed into the search box as the query disease, and a number of terms for Dravet syndrome are suggested from different ontologies, namely from NCIT, MONDO, Orphanet and DOID (Figure 4). Choose the Orphanet rare disease ontology term. The rare disease ontology term from Orphanet provides the most crosslinks to additional expertly reviewed key information on rare disease, such as clinical summaries in multiple languages and links to other standardized vocabularies for disease-related terms such as OMIM, GARD, ICD-10, ICD-11, UMLS and MedDRA.

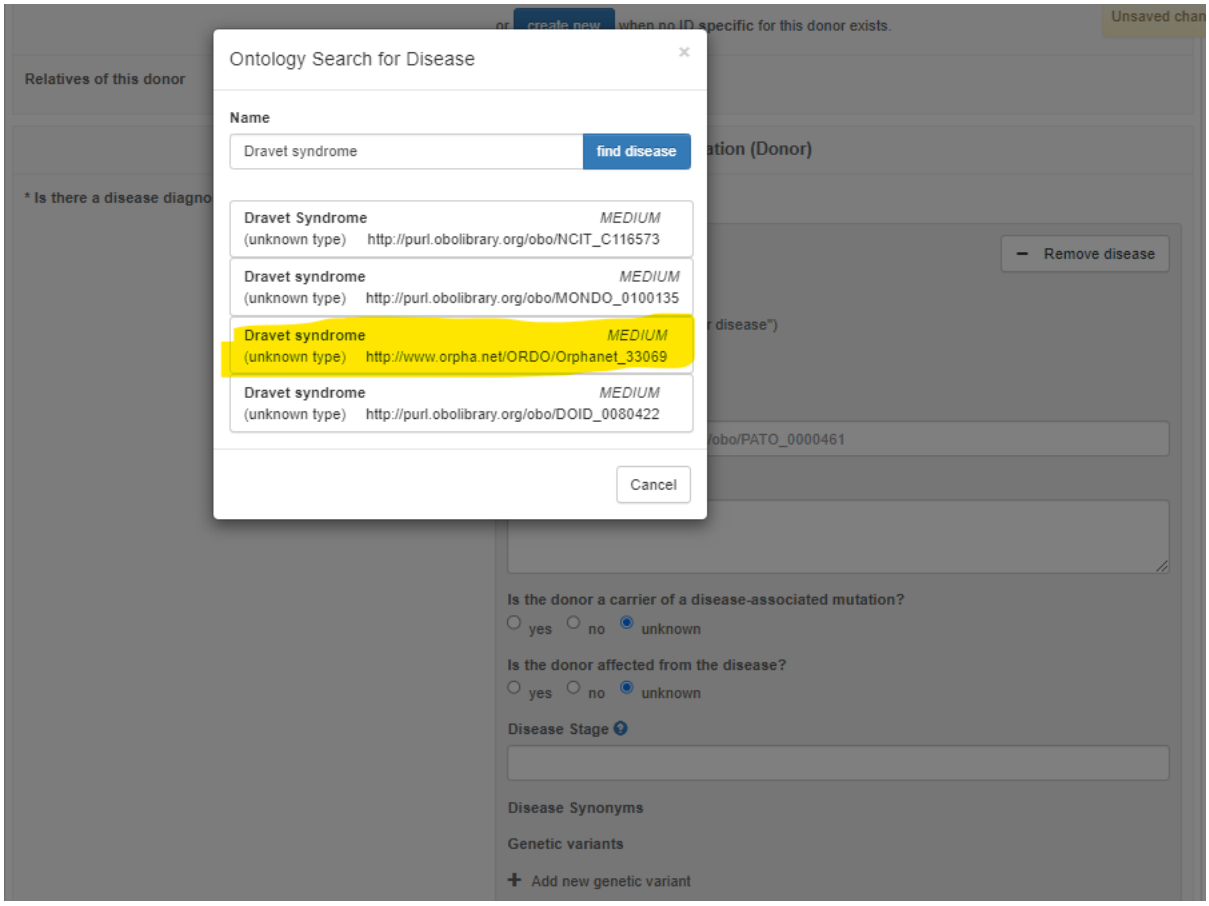


Figure 4. Selecting the rare disease ontology term from a list.

Recording Non-/Disease associated phenotypes

Phenotypes or clinical features associated with the donor are critical for reaching a rare disease diagnosis. Phenotypes can be entered as free text phrases in the sections “Disease associated phenotypes” or “Non-disease associated phenotypes” (Figure 5). Currently, hPSCreg does not have an automated drop-down suggestion list for phenotypes. However, to facilitate future digitalization of phenotype information, please enter the human readable labels from the standardized vocabulary at Human Phenotype Ontology (HPO). For example, “low birth weight” has the equivalent term HPO name “Small for gestational age”.

Please be aware that free text entries will always be difficult for efficient online searching and thus we require users to play the utmost attention to free text entries and potential spelling and typo mistakes.

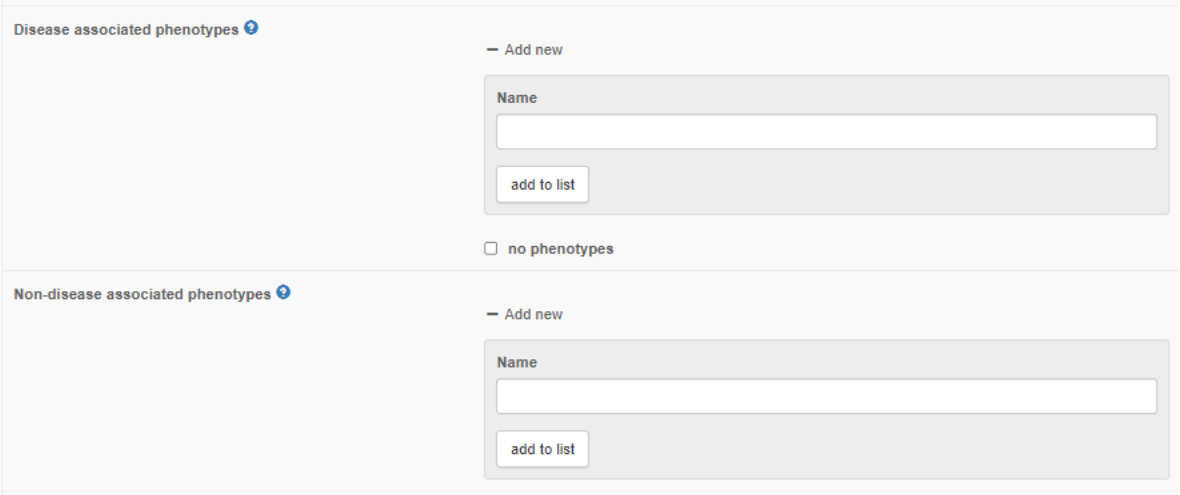


Figure 5. Disease-associated or non-disease associated phenotypes free text boxes.

Recording genetic variants associated with Rare Disease

If a donor has been typed for genetic variants that are associated with the rare disease, please choose the rare disease first (as in the section “Recording the rare disease associated with the donor”), and then a gene variant can be added by clicking on: + Add new genetic variant (see bottom of Figure 3).

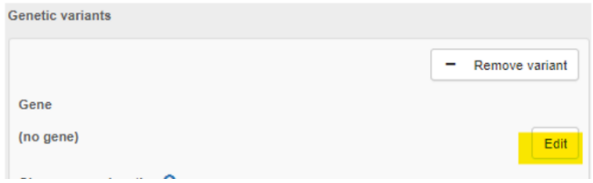
A new set of fields for the genetic variant will appear (Figure 6). Please look up the official gene symbol and NCBI Entrez Gene ID here. As an example, please see the gene “sodium voltage-gated channel beta subunit 1”, which has the official gene symbol “SCN1B” and Gene ID “6324” (Figure 7A). The gene symbol and gene ID can be recorded in hPSCreg® as follows:

- Click Edit (see Figure 7 B)
- Select Entrez from the drop-down menu (see Figure 7 C)
- Enter the Entrez Gene ID, which is an integer (see Figure 7 D)
- Enter the official gene symbol in **all capital letters** (see Figure 7 D)
- Click Add Marker to record the gene
- The gene is now displayed in hPSCreg (Figure 7 E)

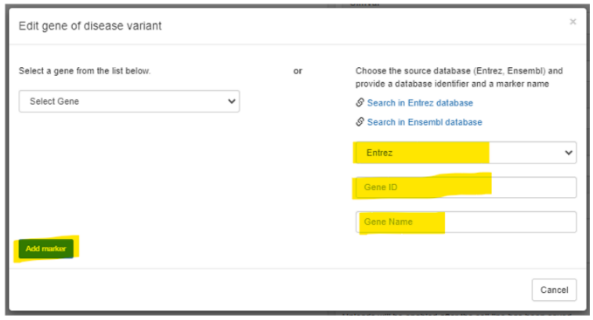
Figure 6. Genetic variant fields

A

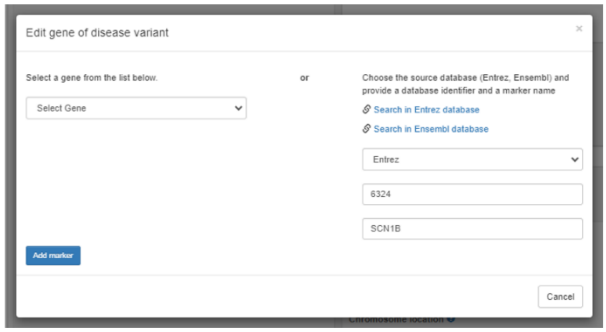
B



C



D



E

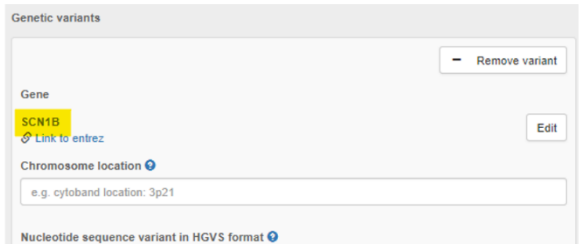


Figure 7. Find the official gene symbol and Entrez Gene ID at NCBI, and enter official gene symbol and Entrez GeneID into hPSCreg

Details of the gene variant can be entered in the remaining fields, such as a description of the variant in HGVS format, the heterozygosity of the variant, and any database links to the variant in ClinVar, dbSNP, dbVar, or PubMed. Finally, a detailed description can be entered in the free text box to supplement the database links.

For a detailed guidance on recording genetic information in hPSCreg, please consult the Guidance for Recording Genetic Information in hPSC Lines.

Summary

Proper annotation of the Rare Disease will help to maximize the discovery of these valuable resources in the research community and encourage their application to understand and treat rare diseases.

We do understand that your time is valuable and that filling out fields in our registry can be time consuming. However, by taking the time and effort to correctly annotate information about your cell lines, you will get more visibility for your cell line and ensure valuable research tools that you have generated can benefit other researchers. This supports the FAIR principles, which are important to improve research infrastructure in the RD field.

Funding acknowledgement

This guidance document was created with support from the Horizon 2020 project European Joint Programme for Rare Diseases (EJP-RD) under grant agreement 825575.

Contact

If you have questions or comments about rare disease cell line registration, please contact us using our contact form.

References

1. Kurtz, A., Mah, N., Chen, Y., Fuhr, A., Kobold, S., Seltmann, S., and Müller, S.C. (2022). Human pluripotent stem cell registry: Operations, role and current directions. Cell proliferation *55,* e13238. 10.1111/cpr.13238.
2. Directorate-General for Research and Innovation at the European Commission. Rare diseases: EU research on rare diseases. https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en.
3. Directorate-General for Health and Food Safety at the European Commission. Rare diseases. https://health.ec.europa.eu/european-reference-networks/rare-diseases_en.
4. Sánchez-Duffhues, G., and Hiepen, C. (2023). Human iPSCs as Model Systems for BMP-Related Rare Diseases. Cells *12.* 10.3390/cells12172200.

5. Fuhr, A., Kurtz, A., Hiepen, C., and Müller, S. (2022). Organoids as Miniature Twins—Challenges for Comparability and Need for Data Standardization and Access. Organoids *1,* 28–36. 10.3390/organoids1010003.
6. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L.B., and Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data *3,* 160018. 10.1038/sdata.2016.18.