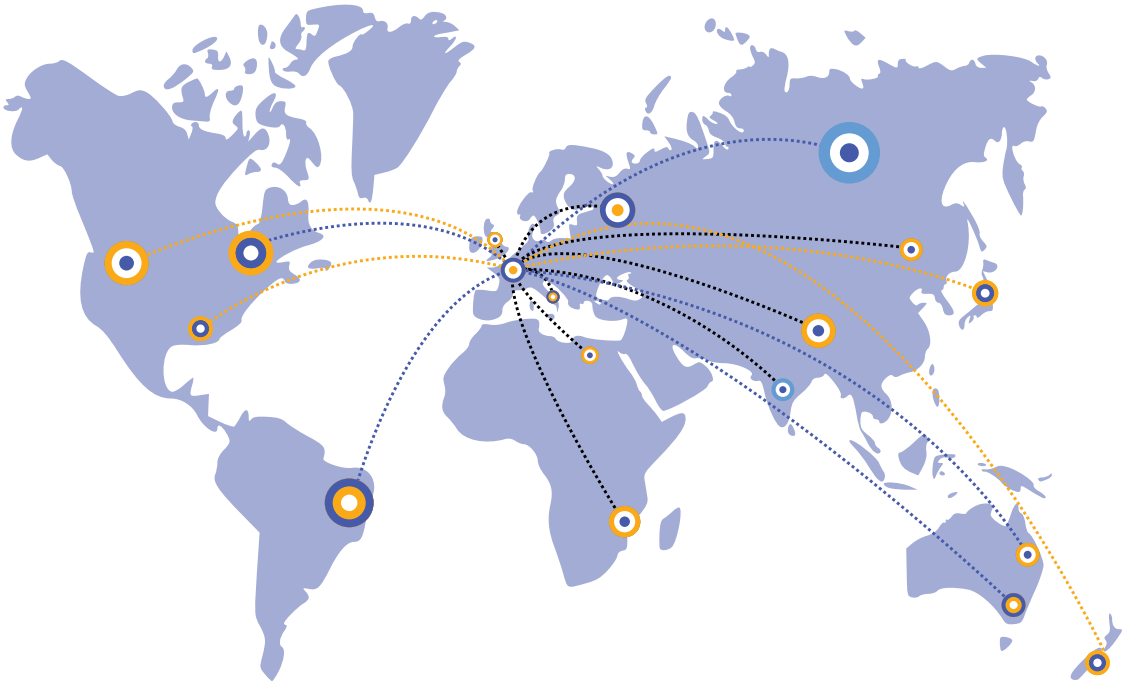
Guidance for Recording Genetic Information in hPSC Lines

2022-03-25

Version 1.1



The Human Pluripotent Stem Cell Registry (hPSCreg®) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 726320

Contents

1 Introduction

Genetic information on hPSC lines, including details of known disease-associated genetic variants or genetic modifications in the human pluripotent stem cell (hPSC) lines, are recorded as systematically as possible in hPSCreg in accordance with the FAIR data principles - Findable, Accessible, Interoperable, and Reusable (https://www.force11.org/group/fairgroup/fairprinciples). Disease-associated genetic variants are those which have been shown to be associated with or causative for a disease, through studies such as linkage analysis, candidate gene approach or genome-wide association studies. Within the scope of the Human Pluripotent Stem Cell Registry (hPSCreg; https://hpscreg.eu/), genetic modifications are defined to be any changes in the genomic DNA, which are not detected in the parental clone. These modifications include variants (spontaneously arising with respect to the parental clone) and engineered modifications (changes made through human intervention by genetic engineering techniques). Using this definition of genetic modifications, genetic modifications of a parental clone are recorded in the subclones of a parental line.

This guidance document is intended to assist users in entering genetic information into hPSCreg. Genetic information on a cell line includes:

- disease-associated genetic variants
- karyotypes
- genome-wide genotyping data
- HLA Types
- short tandem repeat (STR) / fingerprinting profiles
- genetic modifications

Genetic information can be entered at the donor level or the cell line level.

2 Genetic Information at the Donor Level

The donor level genetic information is entered on the Donor Information tab.

2.1 Disease-associated Genetic Variants in an hPSC Line

If there is a known disease-associated genetic variant, it should be recorded in the parental PSC line. This information is entered on the Donor Information tab, under the heading "Phenotype and Disease-related Information (Donor)". Please refer to the clinical significance terms as maintained by ClinVar for guidance (https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/).

For each disease-associated genetic variant, please enter:

- the disease name: click on enter disease, type in the disease and choose the ontology term for the disease
- free text: comments can be entered here for clarification notes
- click on + Add new genetic variant

A new box, called "Genetic variants" will open up. Please enter:

- gene symbol (press the Edit button to pick the gene symbol)
- chromosome location (cytoband location)
- nucleotide sequence variant in HGVS format*
- protein sequence variant in HGVS format*
- the zygosity of the variant (homozygous, heterozygous, hemizygous, or mosaic)
- supporting evidence for the variant of interest from database entries in ClinVar, dbSNP, dbVar, or PubMed.
- free text: place to enter details about the supporting evidence
- upload a file: supporting documentation of a variant, such as a sequencing chromatogram can be uploaded. Please do not upload any data with personally identifying information.

* The variant should be in HGVS nomenclature (see Section 7 for instructions). Alternatively, database entries in ClinVar, dbSNP, or dbVar can be used.

2.2 Karyotype

Karyotyping should be carried out and reported according to community-accepted standards, for example (ISCN 2020, ISCBI).If karyotyping has been carried out at the donor level, please enter the karyotype under the tab Donor Information -> heading "Karyotyping (Donor) -> "Donor Karyotype" in the free text box. A supporting data file, such as an image of the metaphase spread, can be uploaded as well.

2.3 Other Genotyping (Donor)

If any kind of genome-wide analysis has been performed, please enter it under the tab Donor Information -> heading "Other Genotyping (Donor)". This data may include: 1) array-based genotyping (e.g. aCGH, SNP arrays); 2) sequencing-based genotyping (e.g. exome-seq); 3) Methylation profiling (e.g. array or sequencing methods). Personally identifiable genetic and phenotypic data should be previously deposited to

a database such as the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega), which ensures secured distribution under controlled access according to the original consent agreements of the individual donors. If the data has been deposited in another database (e.g. Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo/; ArrayExpress: https://www.ebi.ac.uk/arrayexpress/), you can also enter the link to this data in the provided field.

3 Genetic Information at the Cell Line Level

Genetic information at the cell line level is entered on the Genotyping and Genetic Modification tabs.

3.1 Karyotyping (Cell Line)

Karyotyping should be carried out and reported according to community-accepted standards, for example (ISCN 2020, ISCBI).If karyotyping has been carried out on the cell line, please provide the following information in the Genotyping tab:

- passage number
- the cell line karyotype in the text box (For example, “46,XY” or “46,XX” for normal karyotypes)
- a supporting document as an upload: for example, an image of the metaphase spread

3.2 HLA Typing

If HLA typing has been carried out, please enter the HLA types using the HLA Nomenclature provided by http://hla.alleles.org, which provides up-to-date lists of alleles.

HLA Class I alleles: http://hla.alleles.org/alleles/class1.html

HLA Class II alleles: http://hla.alleles.org/alleles/class2.html

Other alleles: http://hla.alleles.org/alleles/classo.html

3.3 Genome-wide Genotyping Data

If any kind of genome-wide analysis has been performed at the cell line level, please enter it under the tab Genotyping -> heading "Other Genotyping (Cell Line)". This data may include: 1) array-based genotyping (e.g. aCGH, SNP arrays); 2) sequencing-

based genotyping (e.g. exome-seq); 3) Methylation profiling (e.g. array or sequencing methods). Personally identifiable genetic and phenotypic data should be previously deposited to a database such as the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega), which ensures secured distribution under controlled access according to the original consent agreements of the individual donors. If the data has been deposited in another database (e.g. Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo/; ArrayExpress: https://www.ebi.ac.uk/arrayexpress/), you can also enter the link to this data in the provided field.

3.4 Short Tandem Repeats/Fingerprinting

STR information is essential for authenticating the identity of a cell line. Loci can be selected from the drop-down list, or new loci can be added as necessary. Please enter information for both alleles.

3.5 Genetic Modifications

All genetic modifications pertain to subclones of a parental cell line and are entered on the Genetic Modification tab in the subclone record.

3.5.3 Genetic Modifications Related to a Disease or Phenotype Context

These kinds of modifications typically involve the introduction or correction of a disease-associated variant in a cell line. If the modification introduces a gene mutation, please name the disease associated with the gene mutation. Conversely, if the modification corrects a gene mutation, please name the disease associated with the gene mutation. Please fill in:

- the disease name: click on the box enter disease to select a disease
- free text: place for additional clarifications about the disease
- genetic modification
- type of modification (variant, transgene expression, gene knock-out, gene knock-in, isogenic modification). See Appendix, Section 11 below for details.
- gene affected
- chromosome location (cytoband location)
- free text: place for additional clarification on the genetic modification that was carried out.

- supporting file upload: for example, sequencing chromatograms of CRISPR/Cas edited sites, plasmid vector maps

If multiple gene modifications in the same disease were performed, please add all modifications by clicking on + Add new modification .

Additional gene modifications concerning a different disease can be entered by clicking on + Add new disease and entering the information about the genetic modification as described above.

3.5.4 Genetic Modifications which are not Disease-related

These kinds of modifications typically involve reporter gene constructs, which are not directly related to a disease. The steps for entering non-disease related modifications are similar to the menu for the disease-related modifications above, except that there is no disease associated with the genetic modification.

4 How to Describe a Disease Variant Using HGVS Nomenclature

4.1 Easy Cases

The HGVS nomenclature describes variations in the human genome in a unambiguous manner. Changes are described within the context of a reference sequence (i.e. genomic DNA, RNA transcript, or protein).

For example, a single nucleotide variant in KCNH2, at position c.1591C>T in the mRNA would be written in HGVS notation as: "NM_000238.3:c.1591C>T". Broken down into its components, the notation means:

- reference sequence for mRNA of KCNH2: NM_000238.3
- reference sequence is a cDNA sequence : c.
- position in the reference sequence where the change is located: 1591
- affected base at named position, in the reference sequence: C
- substitution: >
- substituted base at the named position: T

The corresponding HGVS for the protein would be "NP_000229.1:p.Arg531Trp".

A complete description of the HGVS notation can be found here: http://varnomen.hgvs.org/. For characterized disease variants, the easiest way to find

the HGVS annotation is to search for the gene symbol and cDNA or protein position of the change in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar). Please see Figures 1 and 2.

4.2 Other Cases

If your variant of interest is not in ClinVar, please try to describe it in HGVS annotation using the guidelines at: http://varnomen.hgvs.org/. You can check if your HGVS notation for your variant is correct at the Mutalyzer web tool (https://mutalyzer.nl/). A complete description of how to handle DNA variants can be found here:

http://varnomen.hgvs.org/recommendations/DNA/

Some cases that have already occurred in hPSCreg include the following topics.

4.2.3 Recommendations for reporting mutations in alleles:

https://varnomen.hgvs.org/recommendations/DNA/variant/alleles/

4.2.4 Repeated sequences in variants

http://varnomen.hgvs.org/recommendations/DNA/variant/repeated/

4.2.5 How to Report a Variant if a Genomic Reference Sequence Does Not Exist

A Locus Reference Genomic Sequence (http://www.lrg-sequence.org/) should be used, and if one is not available, it should be requested. See here for details: https://varnomen.hgvs.org/recommendations/general/

Figure 1. Searching for variant "MYH7 c.2156G>A" in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)

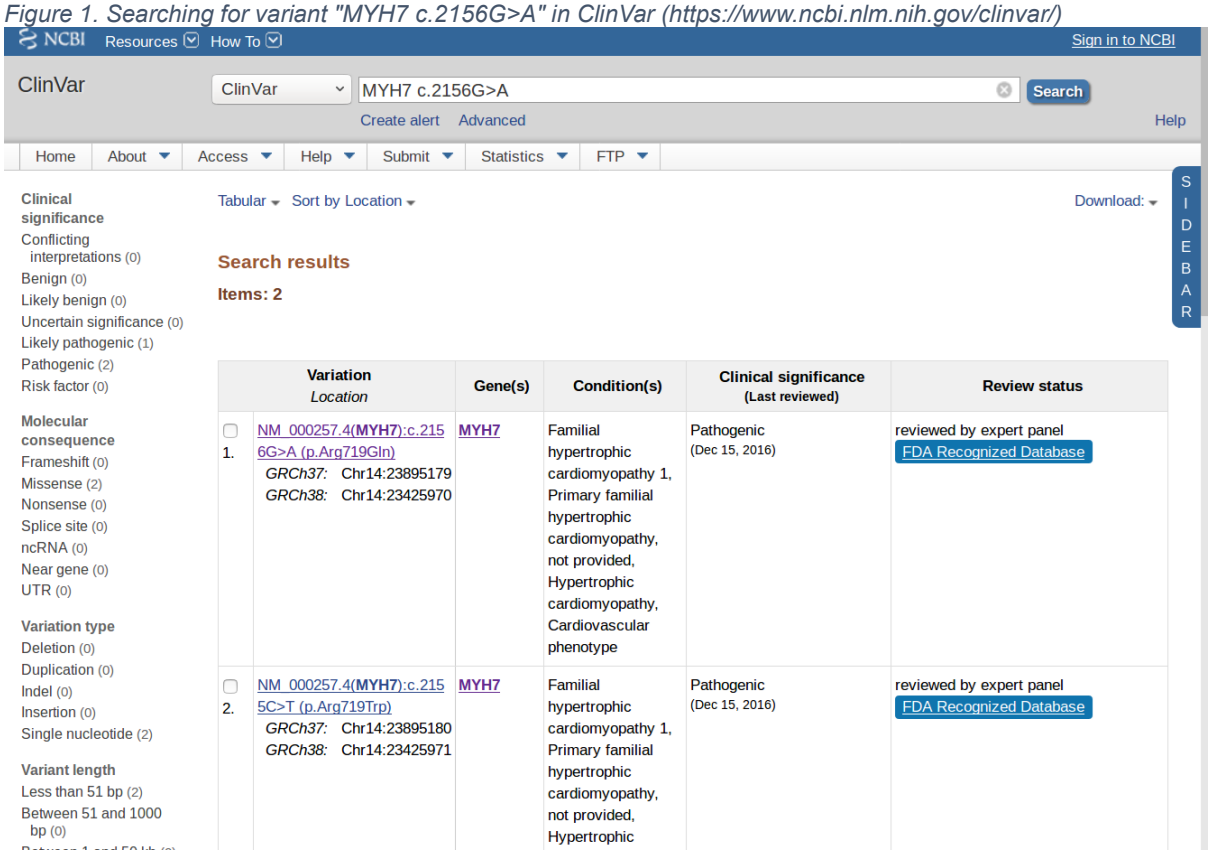
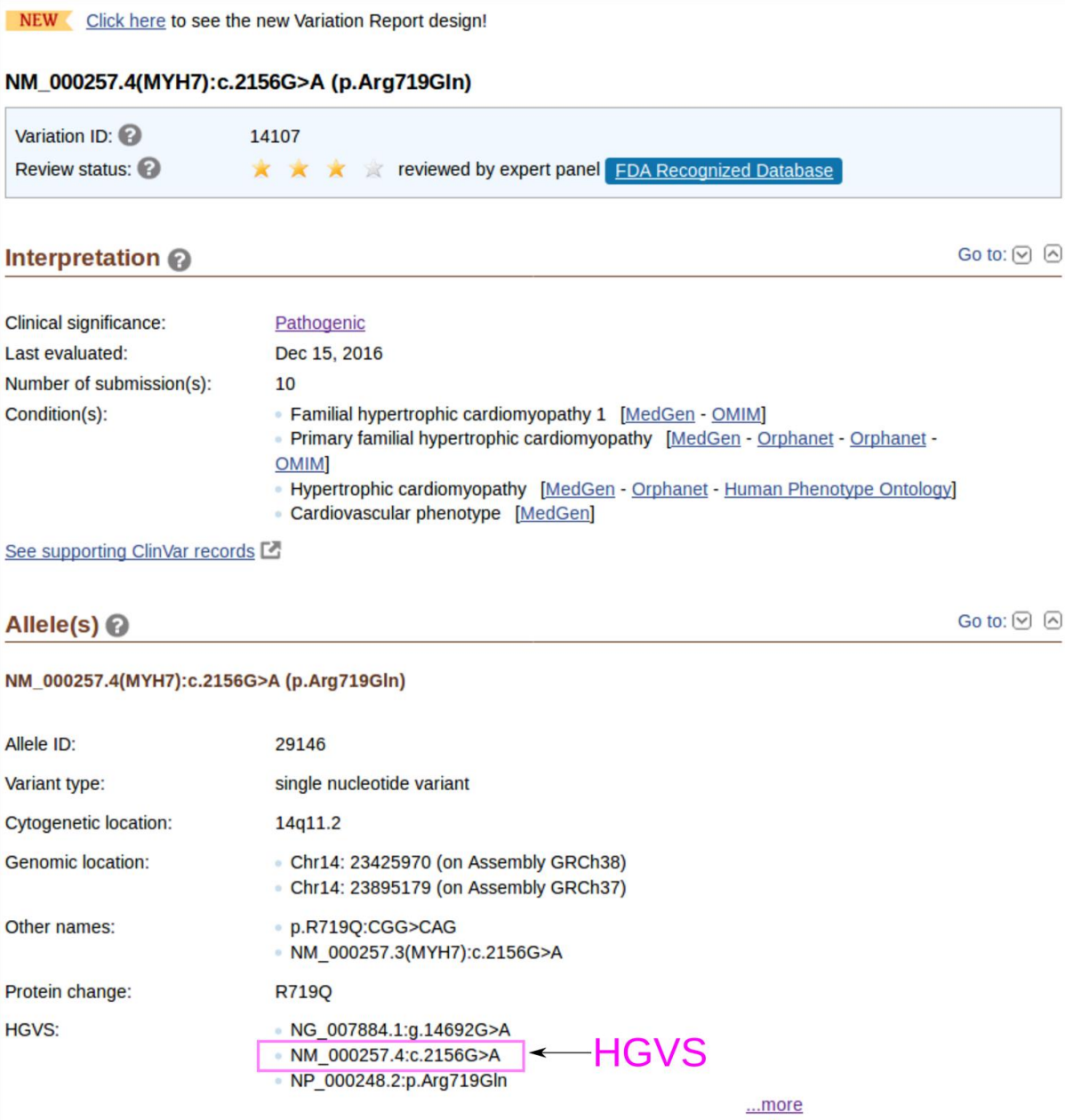


Figure 2. HGVS notation of the variant in ClinVar



5 Still need help?

If you need help in entering the genetic information, please contact hPSCreg directly using the contact form at: https://hpscreg.eu/contact

6 Appendix

6.1 Useful Resources for Annotating Human Genomic Variation

Table 1 – Databases and standards for recording human genomic variation

Name of Resource	Description	Weblink
dbSNP	database of small-scale human genome variations	https://www.ncbi.nlm.nih.gov/snp
ClinVar	resource of genomic variation and it relation to human health	https://www.ncbi.nlm.nih.gov/clinvar/
dbVar	database of human genomic structural variation	https://www.ncbi.nlm.nih.gov/dbvar/
HGVS nomenclature	international standard for reporting sequence variants	https://varnomen.hgvs.org/
Mutalyzer	web tool to check HGVS notation	https://mutalyzer.nl/

6.2 Recording the type of genetic modification

Each type of genetic modification has its own menu set-up. Explanations are provided here for each option.

6.2.3 Variant

An unintended genetic variant, which was not detected in the parental line, should be entered here. The menu layout is the same as the menu for entering a disease variation (see Section 3).

6.2.4 Transgene Expression

This refers to exogenous gene expression, for example the introduction of a GFP reporter.

- Gene (target): gene symbol of the gene which has been introduced into the genome of the cell line
- Chromosome location: cytoband location of the introduced gene, if applicable.
- Delivery Method: homologous recombination, CRISPR-associated (CRISPR/Cas) system, TALEN, Transposon, Viral, Zinc finger nuclease, Other (please specify)

- Free text: please briefly describe other details that cannot be entered in the menu system. For example, was a promoter gene construct engineered into the genome? Is the integration site known?
- Upload a file: you can upload supporting information, such as vector maps or protocols for the genetic engineering

6.2.5 Gene Knock-out

This refers to an endogenous gene that has been rendered non-functional in the cell line genome.

- Gene (target): gene symbol of the gene that has been inactivated
- Chromosome location: cytoband location of the knocked-out gene, if applicable.
- Delivery Method: homologous recombination, CRISPR-associated (CRISPR/Cas) system, TALEN, Transposon, Viral, Zinc finger nuclease, Other (please specify)
- Free text: please briefly describe other details that cannot be entered in the menu system. For example, what part of the gene was knocked-out, and how does this affect transcription or translation of a protein. Is residual functionality expected?
- Upload a file: you can upload supporting information, such as vector maps or protocols for the genetic engineering

6.2.6 Gene Knock-in

This refers to an exogenous gene that has been introduced into the cell line's genome, targeted to a specific locus.

- Gene (target): gene symbol of the gene locus where the transgene has been introduced
- Transgene: gene symbol of the exogenous gene that has been introduced
- Chromosome location: cytoband location of the genes, if applicable.
- Delivery Method: homologous recombination, CRISPR-associated (CRISPR/Cas) system, TALEN, Transposon, Viral, Zinc finger nuclease, Other (please specify)
- Free text: please briefly describe other details that cannot be entered in the menu system. For example, has a target gene been purposely disrupted by the exogenous gene? Was a promoter introduced?

- Upload a file: you can upload supporting information, such as vector maps or protocols for the genetic engineering

6.2.7 Isogenic Modification

In this case, an isogenic modification is considered to be a gene modification (usually small) that corrects a gene mutation in a parental line, or a gene modification that introduces a disease-associated mutation in a normal (non-diseased) parental cell line.

- Gene (target): gene symbol of the gene that has been modified
- Chromosome location: cytoband location of the modified gene
- Nucleotide sequence variant in HGVS format
- Protein sequence variant in HGVS format
- Is the modification homozygous or heterozygous?: homozygous, heterozygous, mosaic
- How has the target locus been modified?: a mutation has been "normalized", a mutation has been introduced into a normal locus, a mutation has been modified to another at the same locus, other (please specify)
- Free text: please briefly describe other details that cannot be entered in the menu system. For example, what method was used for gene editing? If CRISPR/Cas was used, how many clones were sequenced and how many off-target sites were found? To what extent does the intended gene edit affect the gene function (eg. introduction of a stop codon)?
- Upload a file: you can upload supporting information, such as sequencing traces of clones or details of the gene editing protocol.