Master's thesis topic: Development of a support documentation system for familial hypercholesterolemia genetic diagnosis and worldwide analysis of variant-level data

Description: Familial hypercholesterolemia (FH) is an autosomal semi-dominant genetic disorder of lipid metabolism, caused by pathogenic variants in *LDLR*, *APOB* and *PCSK9*. FH is characterized by life-long elevated cholesterol concentrations leading to an increased cardiovascular risk. Early diagnosis of FH followed by appropriate pharmacological therapies can prevent the development of premature atherosclerosis, however, FH is severely underdiagnosed worldwide.

Diagnosis of FH can be made at the clinical level, but only a genetic diagnosis can confirm the FH diagnosis. However, the identification of a genetic variant in a gene of interest does not guarantee that it is indeed the cause of FH, hence using expert-approved variant pathogenicity classification guidelines is crucial for a standardized and reliable genetic diagnosis. In 2015, the American College of Medical Genetics and Genomics (ACMG) published general interpretation recommendations with a 5-tier system where evidence considered for variant pathogenicity includes the type of variant, their frequency in population databases, functional studies, *in silico* prediction of the variant's impact, identification in unrelated individuals with the same consistent phenotype and co-segregation of variants with the disease. ClinGen is an international consortium specialized in defining clinical relevance of genes and variants that includes biocurators and baseline annotators responsible for producing variant interpretations for ClinVar, a reference public database of variant data.

FH is a common condition (~1:250) and about 8000 variants have been described in the 3 genes associated to FH. For variant interpretation, a readily available database of all published FH variants and associated data is a valuable resource for timely genetic diagnosis. Such a database was first constructed in 2016 in our lab (Cardiovascular Research Group at the Instituto Nacional de Saúde Dr. Ricardo Jorge, INSA), but with updates in variant classification and with the increase of wide-spread use of *next-generation sequencing* techniques, the amount of identified variants in FH has grown exponentially. The aim of this project is to develop and optimize new techniques to feed the FH database to support FH genetic diagnosis and to perform a worldwide analysis of variant level data.

The following tasks will be developed:

(1) implement a new system for literature annotation;
(2) capture relevant information for variant classification, from:
    a. ClinVar public database
    b. systematic literature search from 2016 through 2023;
(3) harmonize and complete HGVS correct nomenclature;
(4) perform *in silico* prediction of variant impact;
(5) search population databases for variant frequency;
(6) incorporate internal case level data from the Portuguese FH Study and other European FH studies;

(7) classify all variants according to the most recent ACMG guidelines for variant interpretation;
(8) analyse the new updated variant data database to be used for FH genetic diagnosis by:
    a. gene
    b. variant type
    c. functional effect/prediction
    d. protein domain/exon
    e. variant classification
    f. variant frequency
    g. geographical distribution

At the end of this project, it is expected that the master student will have developed skills in literature search and annotation (Pubmed and hypothes.is), using several genetic variant resources (ClinVar, ClinGen's Variant Curation Interface, gnomAD, HGVS, Mutalyzer, MaxEntScan, REVEL), management of large amounts of genetic data, and genetic variant interpretation (ACMG guidelines and FH VCEP adaptations). As part of this work the student will be integrated in an international consortium – Clinical Genome Resource (ClinGen) – as an FH biocurator/annotator. This way the student will gain valuable expertise in genetic diagnosis variant analysis.

Bibliography:

1. Alves AC, Chora JR, Bourbon M. Genomics of familial hypercholesterolaemia. *Curr Opin Lipidol*. 2019;30(2). doi:10.1097/MOL.0000000000000584
2. Chora JR, Medeiros AM, Alves AC, Bourbon M. Analysis of publicly available LDLR, APOB, and PCSK9 variants associated with familial hypercholesterolemia: Application of ACMG guidelines and implications for familial hypercholesterolemia diagnosis. *Genet Med*. 2018;20(6):591-598. doi:10.1038/gim.2017.151
3. Iacocca MA, Chora JR, Carrié A, et al. ClinVar database of global familial hypercholesterolemia-associated DNA variants. *Hum Mutat*. 2018;39(11):1631-1640. doi:10.1002/humu.23634
4. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-423. doi:10.1038/gim.2015.30
5. Chora JR, Iacocca MA, Tichý L, et al. The Clinical Genome Resource (ClinGen) Familial Hypercholesterolemia Variant Curation Expert Panel consensus guidelines for LDLR variant classification. *Genet Med*. 2022;24(2):293-306. doi:10.1016/J.GIM.2021.09.012
6. What is ClinVar? https://www.ncbi.nlm.nih.gov/clinvar/intro/.

Thesis supervisor: Mafalda Bourbon (PhD)

Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis, Unidade de Investigação e Desenvolvimento, Instituto Nacional de Saúde Dr. Ricardo Jorge.

Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa.

Mafalda.bourbon@insa.min-saude.pt

Location: Grupo de Investigação Cardiovascular (https://gicbourbonlab.owlstown.net/), Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis, Unidade de Investigação e Desenvolvimento, Instituto Nacional de Saúde Dr. Ricardo Jorge.