Inserm Cross-Cutting Program – GOLD Gen**O**mic variability in hea**L**th & **D**isease

The GOLD cross-cutting program brings together experts in the analysis of genomic data with complementary expertises in genetic epidemiology, statistical genetics, bio-informatics, functional genomics. It will build on these different expertises to tackle challenges in the analysis of genomic data to understand the role played by genes in the development and the progression of diseases. The first challenge is data collection and accessibility to both phenotypic and genotypic data on the same individuals. This program will address this issue with a special focus on longitudinal data available in the Constances cohort (Work Package 1, WP1). The second challenge is methodological. We lack powerful and innovative methods to integrate longitudinal phenotype information into genetic analyses. This issue will be addressed in WP2 that aims at developing new methods and computing programs building on the different tools and strategies already developed by the different teams. The third challenge is to assess the biological relevance of the statistical findings and the development of cost and time-effective methods for functional screening (WP3). A last challenge is to develop solutions for the integrated analysis of genetic and phenotype data and the building of a dedicated platform to interrogate how genetic variants play a role in various phenotypes. This platform will ultimately be opened to teams outside the GOLD to address their research questions (WP4).

The project is structured into a project management and dissemination work package (WP0) and four main work packages with defined key questions and deliverables.

WP0 Project Management: the goal of this work package is to ensure coordination, administrative management and promotion of the project and the coordination with the POPGEN project that will produce the sequencing data.

WP1 Phenotypes in the Constances cohort: the broad aims of this work-package are to i) identify suitable phenotypes that may be studied with the aid of data from the POPGEN project and ii) develop and utilize appropriate research strategies that will take advantage of the genetic data generated by the POPGEN project as well as other data available from the Constances cohort or from other sources.

WP2 Methodological developments for the analysis of genomic data: this WP will first focus on genetic association studies and the development of novel and powerful methods to find disease associated variants in genes and to infer causality. A second part of the WP will be devoted to the interpretation of the effect of multiple types of genetic variants, both at the level of gene expression and of protein interactions. The last part of the WP will implement best practices for data management and software development.

WP3 Functional interpretation of genetic variants, cellular model and cell-based assays development will be dedicated to: i) the development of cellular models focusing on primary neuronal cultures and cardiomyocytes (hiPSC-CM), ii) the development of high-throughput methods for the simultaneous analysis of large number of genetic variants, iii) the functional evaluation of mitochondrial variants, iv) the high-throughput identification and functional characterization of distant regulatory elements, v) the generation of a large experimental dataset to validate predictive algorithms of the effects of intronic transposable element insertions on gene expression and structure, and vi) the generation of different large datasets to evaluate the effect of punctual variants on splicing auxiliary sequences and optimize prediction algorithms.

WP4 Data integration and the GOLD platform: this WP aims at developing a framework for the smooth integration of genotypic data from thousands of individuals, their interconnexion with the Constances database, and their annotation at multiple levels.