

Innovative databases and query systems in the study of genetic disease



Alexandra Henrion-Caude

alexandra.henrion-caude@fondationimagine.org

Research scientists

Martine Le Merrer
Jean-Philippe Jaïs
Catherine Turleau

Technical staff

Patrice Gastineau
Claude Mugnier

Other members of the team

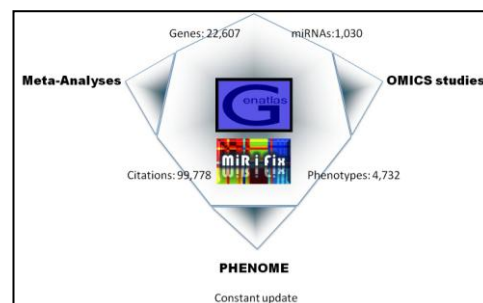
Muriel Girard
Sylvain Hanein
Marie-Liesse Chauvet

Overview

Our challenge is to provide ready access to and improved search capabilities for information on rare diseases. We are targeting both medical and scientific audiences. We have developed key components of IMAGINE's research effort by collating all the known inherited diseases, their endophenotypes, protein-coding genes and non-coding ones, medical images, animal models and therapeutic solutions. Our development should help (i) increasing the probability of a correct diagnosis, (ii) enhancing the fitting of therapeutic strategies to mutations and (iii) predicting prognoses with greater certainty.

History and background

GENATLAS is a database of genes and diseases that was created in 1986 and has been constantly updated ever since. It provides a flow of information from gene structure to expression and from mutations to function as well as phenotypic consequences when available. Data are based on the manual annotation of scientific papers and are provided by a daily literature alert. Annotations of gene-disease relationships, chromosomal cryptic/non-cryptic rearrangements



and their associated phenotypes give to GENATLAS its specificity and make it a genuine

compilation of genetic diseases. GENATLAS also develops tools to take new concepts into account: searches for candidate genes in a given interval, the prediction of messenger RNA, the development of a thesaurus for joint use by several databases (ORPHANET, CEMARA and GENATLAS). The establishment of reciprocal links with databases such as Swissprot, HGMD, IGMT, Genew, GeneCards and Orphanet has confirmed GENATLAS's international reputation. The web-enabled database receives over 250,000 visits per year.

MIRIFIX™ was developed with the general aim of providing a clinical and research portal to promote the discovery non-coding RNAs (ncRNAs) in general with a special focus on the class of microRNA genes that are potentially responsible for inherited diseases. MIRIFIX™ is unique, as it is both an integrated database with its own developments and a dynamic portal to the most up-to-date and relevant web-based tools in the field. The patented MIRIFIX™ software enables the user to query microRNAs at each level of action (as either a candidate or a modifier) through a single interface. It allows the user to explore (i) the location of ncRNAs with respect to annotation of diseases and sets of protein-coding genes, and (ii) the predicted targeting of the gene of interest by microRNAs (SNPs, 3'UTR and/or exons). It also provides the adequate tools for assessing a given sequence as a target for microRNA.

Major projects

Our system is being developed to serve simultaneously as:

- a phenome database. To this end, we study the most appropriate arborescence for defining phenotypes and endophenotypes. Appropriate systems are devised to aggregate clinical images (X-rays, brain MRI, path lab and histological images), in order to define reference entities.
- an OMICS interpretation tool. In close collaboration with the Imagine Bioinformatics Platform, we can help to interpret transcriptomics, miRnomics, epigenomics and next-generation sequencing data.
- a source for meta-analyses and pan-genomic studies. Through our research, we translate the information that is accumulating so rapidly in bioinformatics databases into insights relevant to human disease and biology in general.

Our developments benefit strongly from the interaction with the 19 rare genetic disease reference centers at Necker Children's Hospital. Our projects are based on a novel architecture, with the ontologies used for annotation to enable flow-through information on genes, phenotypes and references.

Developments

<http://www.genatlas.org>

<http://www.mirifix.com> (access on request, via the MIRIFIX™ contact person)

Best publications

1. Bandiera S. et al. Nuclear outsourcing of RNA interference components to human mitochondria. *PLoS One.* (2011). 6(6):e20746.
2. Warman ML. et al. Nosology and classification of genetic skeletal disorders: 2010 revision. *Am J Med Genet A.* (2011). 155A(5):943-68.
3. Jardin F. et al. Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment: a GELA study. *Blood.* (2010). 116(7):1092-104.
4. Bandiera S. et al. microRNAs in diseases: from candidate to modifier genes. *Clin Genet.* (2010). 77(4):306-13.
5. Girard M. et al. miR-122, a paradigm for the role of microRNAs in the liver. *J Hepatol.* (2008). 48(4):648-56.