

Plus d'information sur le référencement croisé des bases de données :

[OMIM](#) est la base de données de gènes et de phénotypes mendéliens créée par le Professeur McKusick et maintenue par le NCBI.

[Genatlas](#) contient des informations pertinentes concernant la cartographie des gènes et les maladies génétiques. La base de données a été créée en 1986 par le Professeur Jean Frézal et se situe à l'Université Descartes de Paris. L'information est recueillie à partir de la littérature. Les données sont mises à jour quotidiennement et disponibles immédiatement en ligne. Un gène ou maladie sont inclus dans Genatlas si leur position est cautionnée par un lod-score significatif (>3) ou une cartographie physique pertinente. Genatlas permet une navigation facile entre les bases de données de Gene et de Phenotype.

[HGNC](#) est un comité fondé conjointement par le « National Human Genome Research Institute » (NHGRI)- US - et le « Wellcome Trust » (UK). Il opère sous les auspices de HUGO. Il est en charge de l'approbation des noms et des symboles des gènes. Tous les symboles approuvés sont stockés dans la base de données du HGNC. Chaque symbole est unique et chaque gène reçoit un seul et unique symbole approuvé. A ce jour, plus de 25 000 symboles ont été homologués ; la grande majorité le sont pour des gènes codant pour des protéines, mais aussi pour des pseudogènes, des ARN non codants et des caractères phénotypiques et génomiques.

[SwissProt](#) est une base de connaissances de protéines annotée à la main, établie en 1986 et entretenue depuis 2003 par le Consortium UniProt, une collaboration entre le « Swiss Institute of Bioinformatics » (SIB) et le « Department of Bioinformatics and Structural Biology » de l'Université de Genève, l' « European Bioinformatics Institute » (EBI) et le « Georgetown University Medical Center's Protein Information Resource » (PIR).

La base de connaissances [UniProt](#) consiste en des entrées de séquence. Les entrées de séquence sont composées de différents types de lignes, chacun ayant son propre format. À des fins de standardisation, le format de la base UniProt suit d'aussi près que possible celle de l' « EMBL Nucleotide Sequence Database ». Les entrées de séquence sont annotées avec les informations suivantes : fonction de la protéine, modification (s) post traductionnelle, domaines et sites ; structure secondaire, structure quaternaire, similitudes avec d'autres protéines, maladie(s) associée(s) au (x) déficience(s) de la protéine, conflits de séquence, variants, etc.