

# Introduction à la Bioinformatique

**Rabah Lebsir**

Université de guelma

Contact: [lebsir.rabah@univ-guelma.dz](mailto:lebsir.rabah@univ-guelma.dz) / [r.lebsir@gmail.com](mailto:r.lebsir@gmail.com)



# Plan

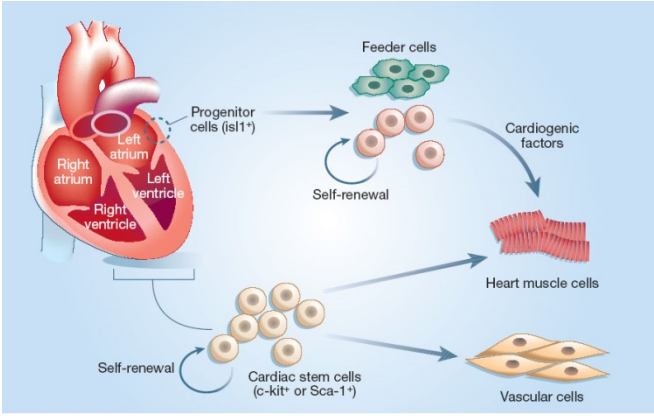
- Quelques notions de base
  - L'ADN
  - Le chromosome
  - Le Gene
  - Génome
  - Le Projet génome humain (HGP)
  - Défis de Recherche en Génomique et en Biologie
  - Dogme de la biologie moléculaire
- Les Sciences « Omiques »
- C'est quoi la Bioinformatique
- Apparition de la Bioinformatique
- Quelques grandes institutions
- Applications de la Bioinformatique
  - Analyse des séquences
  - Comparaison de séquences
  - Comparaison d'une séquence d'ADN avec celles des bases de données
- variabilité génétique
  - Polymorphismes nucléiques
  - SNP
  - Le projet HapMap (1)
  - Le Projet 1000 Genomes
  - Pourquoi étudier la variabilité génétique
  - Techniques pour étudier la variabilité génétique

# Quelques notions de base

- La cellule est l'entité de base de l'organisme
- Il existent plusieurs types de cellules.

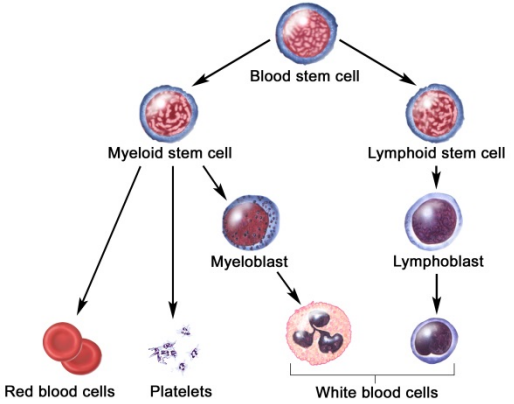


Peau

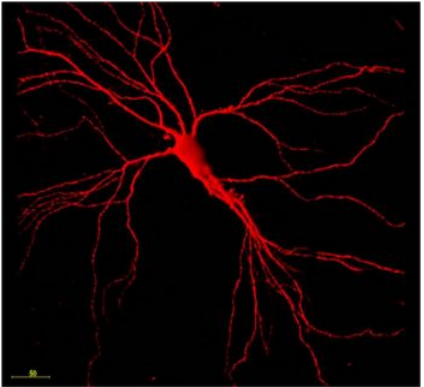


Cœur

(source: Nature 433, 585-587)



Sang



Cellule Nerveuse

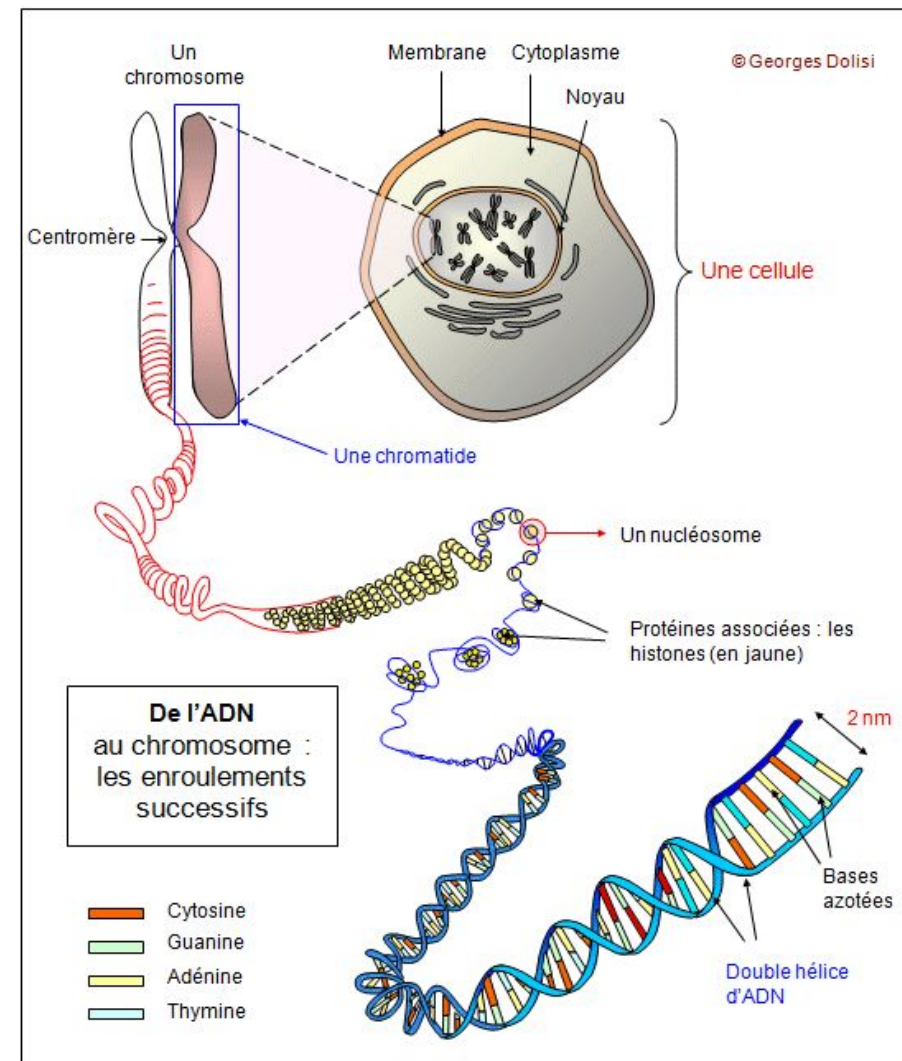
Source:  
<https://sites.google.com/site/aphysionado/home/cellsnv/cellnervx>

- Ces cellules contiennent l'information Héréditaire

# L'ADN - Acide Désoxyribonucléique

- Support matériel de l'hérédité
- Composé de quatre bases (ou nucléotides)  
    **A**: Adénine; **C**: Cytosine; **G**: Guanine; **T**: Thymine
- Deux brins antiparallèles: A  $\begin{cases} \text{A} \\ \text{T} \end{cases}$  | C  $\begin{cases} \text{C} \\ \text{G} \end{cases}$
- Structure en double hélice

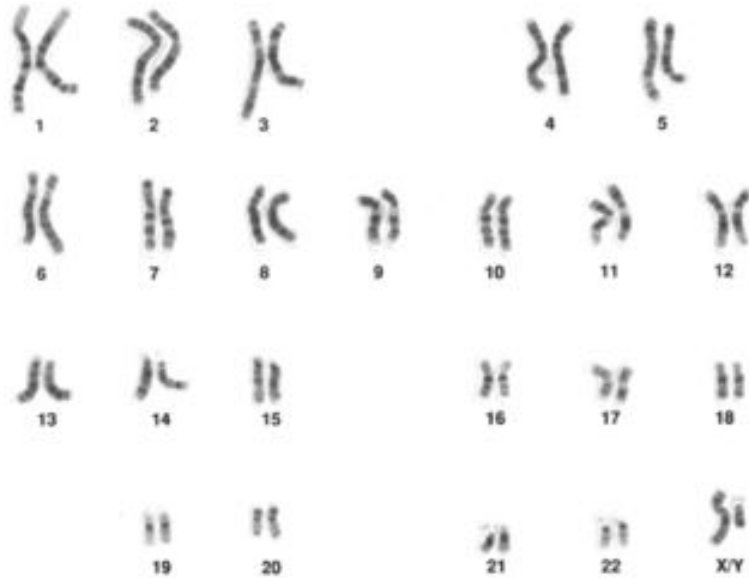
AGTCACGACGTTGTAAAACGACGGGATCATATGACAA  
GATGTGTATCCACCTTAACTTAATGATTTTTACCAAAA  
TCATTAGGGGATTCATCAGCGTTAAACCCGCCACCAG  
ATGGGCGTTAAACGAGTATCCCGGCAGCAGGGGATCA  
TTTTGCGCTTCAGCCATACTTTTCATACTCCCACC





# Le chromosome

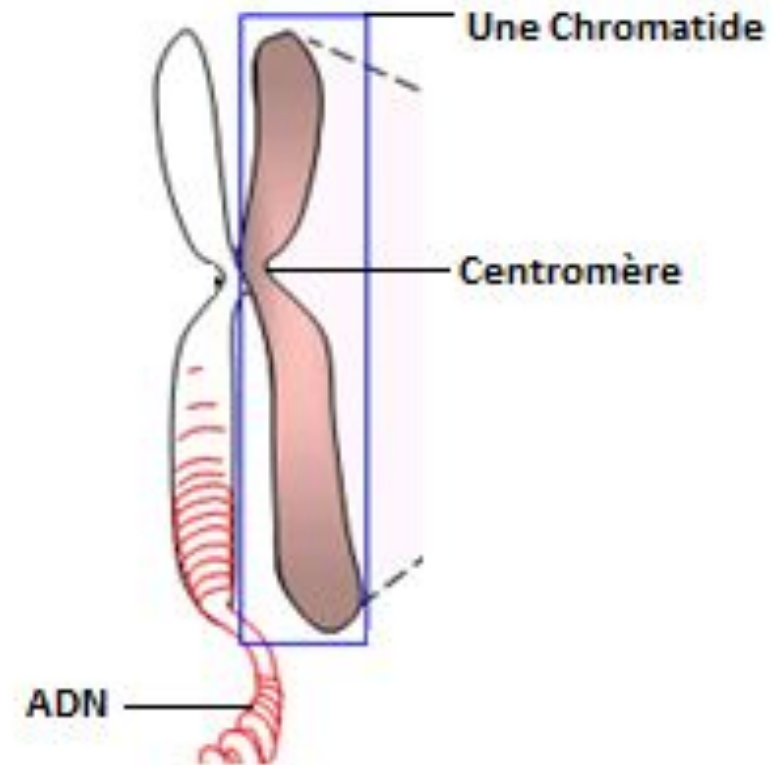
- Le chromosome représente le support physique de l'ADN
- les chromosomes sont composés de deux chromatides identiques
- Chez l'humain, il existe 23 paires de chromosome
  - 22 Chromosomes homologues (autosomes)
  - 1 Chromosome Sexuels (XX ou XY)



Chromosomes chez l'humain

Source:

[http://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9tique\\_humaine](http://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9tique_humaine)



# Le Gène

- portion codante du génome
- Contient toute la recette d'assemblage d'une protéine
- Selon les individus, il peut exister plusieurs versions d'un même gène, appelées **allèles**.
- La localisation d'un gène est appelée « **Locus** »



# Le Gène (suite)

- Un gène à un nom officiel unique attribué par « HUGO Gene Nomenclature Committee »
- Il a une position sur le chromosome
- Il a une fonction

## APOE apolipoprotein E [ *Homo sapiens* (human) ]

Gene ID: 348, updated on 23-Feb-2014

### Summary

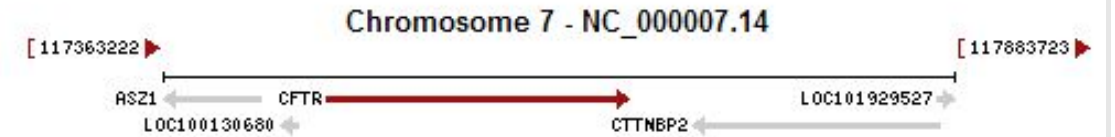
<b>Official Symbol</b>	APOE <small>provided by HGNC</small>
<b>Official Full Name</b>	apolipoprotein E <small>provided by HGNC</small>
<b>Primary source</b>	<a href="#">HGNC:613</a>
<b>See related</b>	<a href="#">Ensembl:ENSG00000130203</a> ; <a href="#">HPRD:00135</a> ; <a href="#">MIM:107741</a> ; <a href="#">Vega:OTTHUMG00000128901</a>
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<a href="#">Homo sapiens</a>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
<b>Also known as</b>	AD2; LPG; LDLCQ5
<b>Summary</b>	Chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. Apolipoprotein E, a main apoprotein of the chylomicron, binds to a specific receptor on liver cells and peripheral cells. ApoE is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. The APOE gene is mapped to chromosome 19 in a cluster with APOC1 and APOC2. Defects in apolipoprotein E result in familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), in which increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants. [provided by RefSeq, Jul 2008]

Base de données de gènes:  
NCBI Entrez, Vega, Ensembl

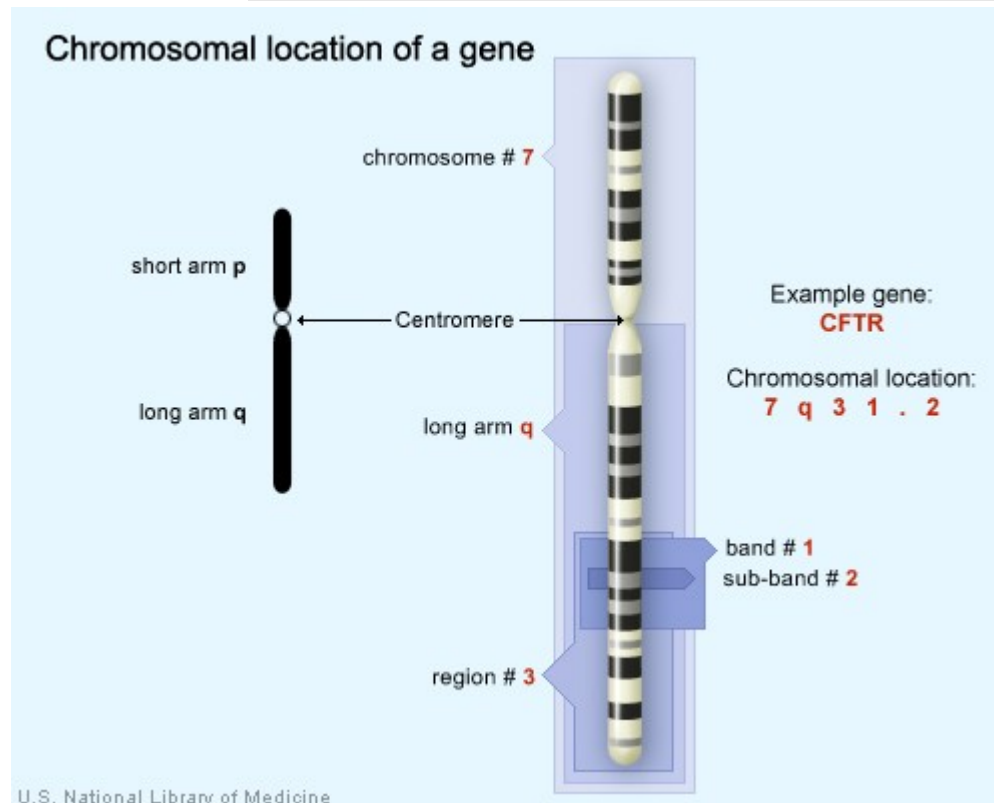
# Gène (suite)

Ex. Gene CFTR

**Location:** 7q31.2  
**Sequence:** Chromosome: 7; NC\_000007.14 (117470772..117668665)



Résultat dans la base de donnée GenBank de NCBI



Explication



# Génome

- Tous les gènes d'un organisme
- Il contient toutes les séquences codantes (gènes) et non codantes.
- La taille du génome se mesure en nombre de nucléotides :  
pb (*paire de bases*), kpb, mpb...

# Génome (Suite)

Taille des génomes d'espèces modèles

↕	Organisme	↕ Taille du génome (Mpb) ↕	↕ Nombre de gènes protéiques estimés ↕
Virus	Virus de la grippe	0,013	
	Bactériophage λ	0,05	
	Bactériophage T4	0,165	
	Mimivirus	1,2	1 260
Bactéries	<i>Mycoplasma pneumoniae</i>	0,816	689
	<i>Pelagibacter ubique</i>	1,3	1 354
	<i>Haemophilus</i>	1,8	1 657
	<i>Staphylococcus aureus</i>	2,8	2 619
	<i>Bacillus subtilis</i>	4,2	4 106
	<i>Escherichia coli</i>	4,64	4 243
Archaea	<i>Nanoarchaeum equitans</i>	0,49	536
	<i>Pyrococcus abyssi</i>	1,77	1 898
	<i>Sulfolobus solfataricus</i>	3	2 977
Eucaryotes	<i>Encephalitozoon cuniculi</i>	2,9	1 996
	<i>Saccharomyces cerevisiae</i> (levure)	12	5 863
	<i>Plasmodium falciparum</i>	21,8	5 314
	<i>Caenorhabditis elegans</i> (nématode)	100	22 628
	<i>Drosophila melanogaster</i> (mouche)	118	16 548
	<i>Arabidopsis thaliana</i> (plante)	119	27 379 <sup>9</sup>
	<i>Populus trichocarpa</i> (peuplier)	485	45 500
	<i>Zea mais</i> (maïs)	5 000	54 606 <sup>10</sup>
	<i>Mus musculus</i> (souris)	3 400	30 000
	<i>Homo sapiens</i> (homme)	3 400	26 517
	<i>Polychaos dubium</i> (amibe)	675 000	

Source: [http://fr.wikipedia.org/wiki/Taille\\_du\\_g%C3%A9nome](http://fr.wikipedia.org/wiki/Taille_du_g%C3%A9nome)

# Le Projet génome humain (HGP)

- Le projet HGP est entrepris en 1990 dont la mission était d'établir le séquençage complet de l'ADN du génome humain.
- Fin du séquençage en 2000

## **Buts:**

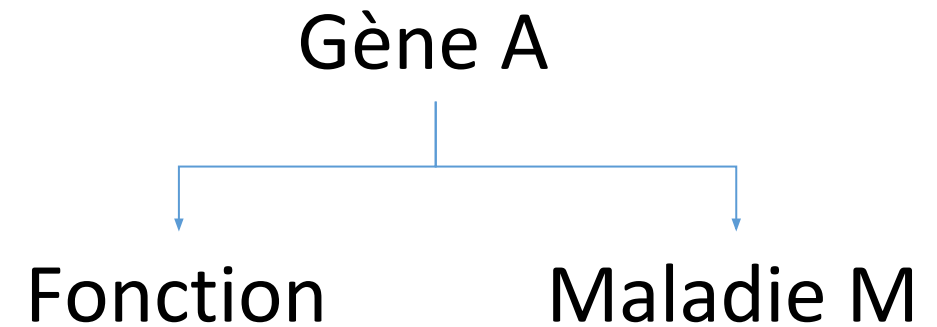
- Séquencer l'ensemble du génome humain avec un taux d'erreur minimal
- Identifier tous les gènes dans cette grande quantité de données
- développer des méthodes plus rapides et efficaces pour le séquençage de l'ADN et l'analyse des séquences
- Publier les séquences déchiffrées immédiatement sur Internet (Accès libre).



GenBank

# Défis de Recherche en Génomique et en Biologie

- Quelle est la fonction de ce gène ?
- Comment le gène influence une maladie?

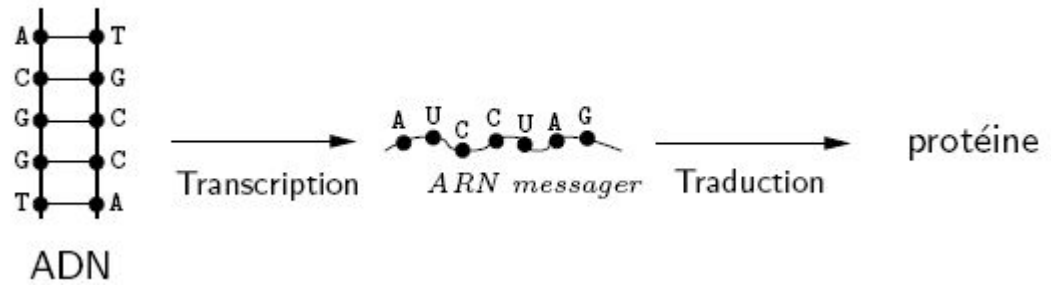


La séquence du génome ne suffit pas

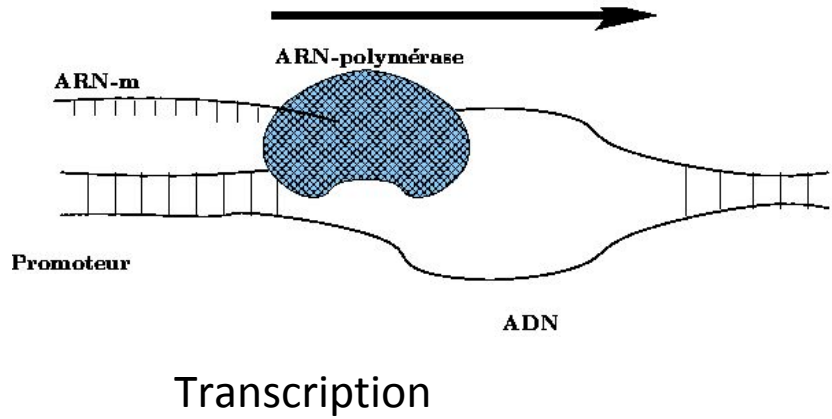
☐ On a besoin d'étudier les produits des gènes (ARN, protéine,..)

# Dogme de la biologie moléculaire

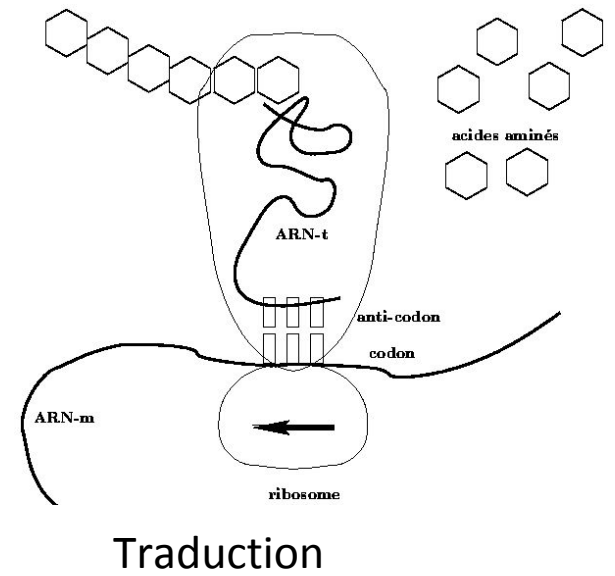
- De l'ADN à la Protéine



Génomique                      Transcriptomique                      Protéomique

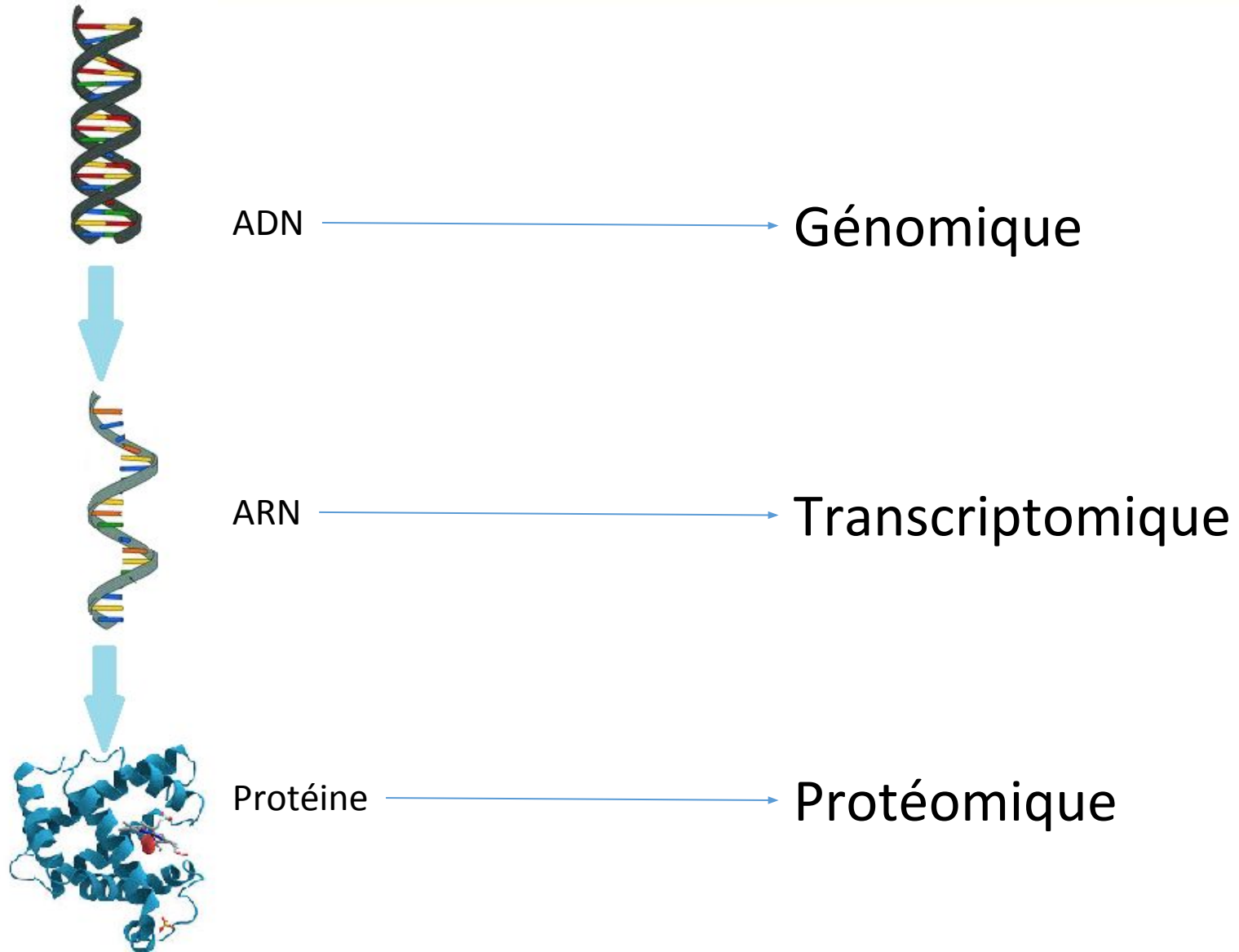


?

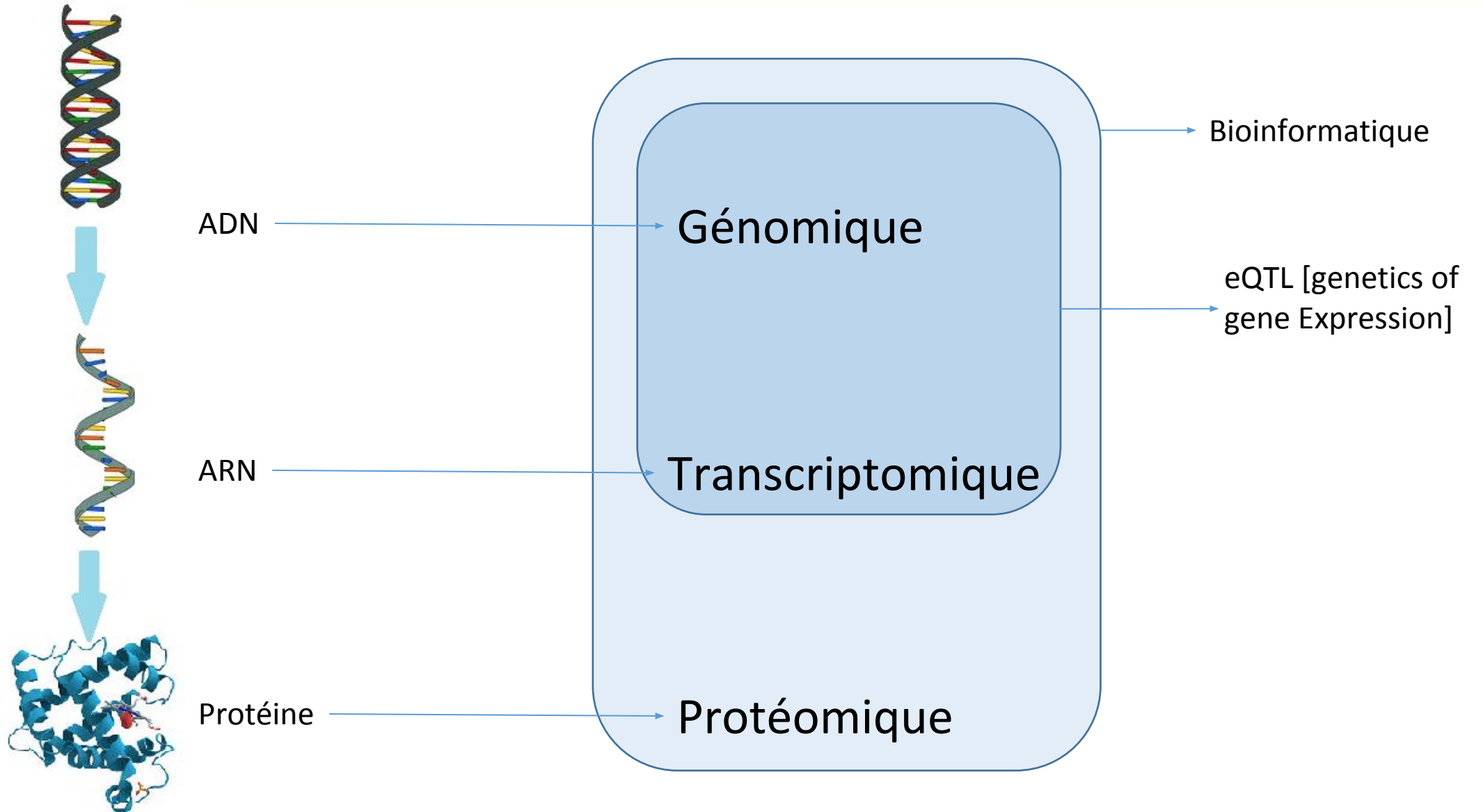




# Les Sciences « Omiques »

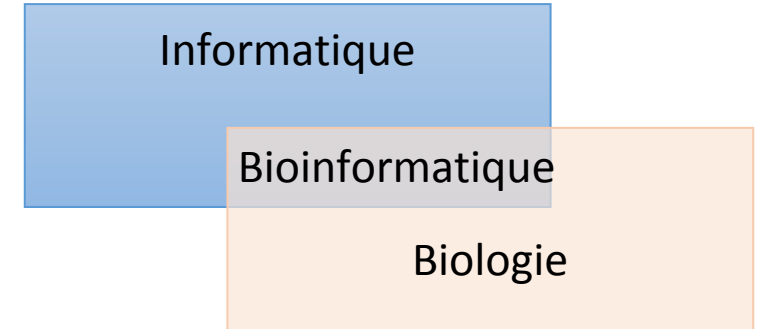


# Les Sciences « Omiques »

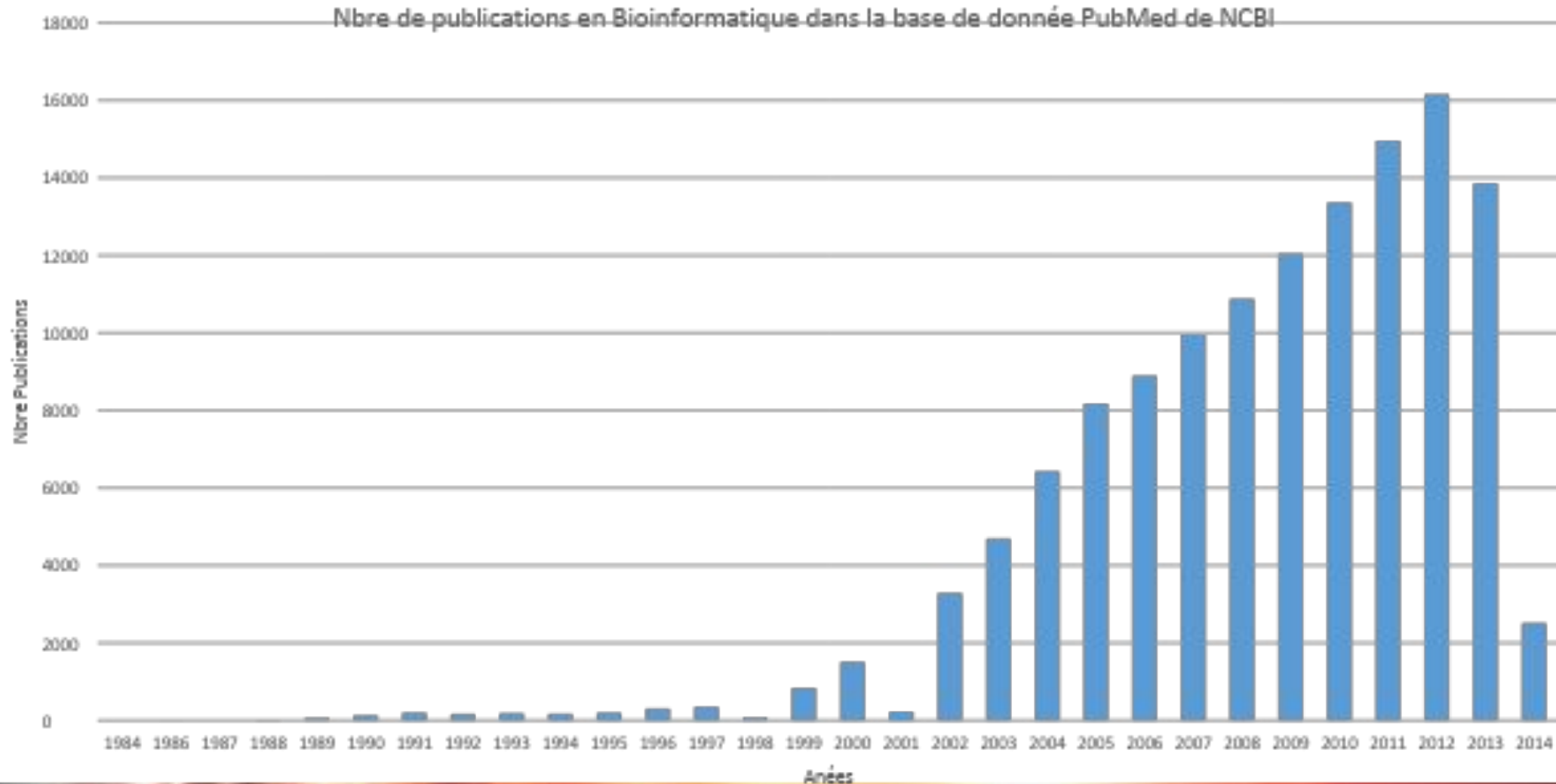


# C'est quoi la Bioinformatique

- Application des techniques computationnelles pour gérer, manipuler et analyser des données biologiques.
- Buts:
  - Organiser les données pour que les chercheurs puissent accéder aux informations existantes ainsi que publier d'autres (**Tirer des informations Biologiques à partir de données brutes**).
  - Développer des outils d'analyse des données.
  - Utiliser ces outils pour interpréter les résultats.



# Apparition de la Bioinformatique



# Quelques grandes institutions

- **NCBI:** National Center for Biotechnology Information (USA)

<http://www.ncbi.nlm.nih.gov/>

- **EBI:** European Bioinformatics Institute

<http://www.ebi.ac.uk/>

- **DDBJ:** DNA Data Bank of Japan

<http://www.ddbj.nig.ac.jp/>

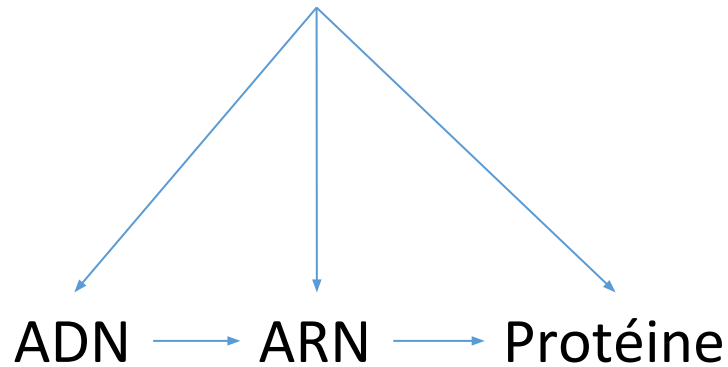
- **SIB:** Swiss Institute of Bioinformatics

<http://www.isb-sib.ch/>



# Applications de la Bioinformatique

## analyse des séquences



## Recherche de similarité

```
TCT AGA GAC GCT CTC CCA AAC ACT GAA
GCC AGT GGA CCA ACA CAC TCC AAG GAA ATT
CCG GCA CTC ACC GCA GTG GAA ACT GGG GCC
ACA AAT CCA CTA GTC CCT TCT GAT ACA GTG
CAA ACC AGA CAT GTT GTA CAA CAT AGG TCA
AGG TCA GAG TCT AGC ATA GAG TCT TTC TTC
GCG CGG GGT GCA TGC GTG ACC ATT ATG ACC
GTG GAT AAC CCA GCT TCC ACC ACG AAT AAG
CAT AAG CTA TTT GCA GTG TGG AAG ATC ACT
TAT AAA GAT ACT GTC CAG TTA CGG AGG AAA
TTG GAG TTC TTC ACC TAT TCT
```

## Génomique comparative

```
TATTTACCATATCAGATTCACATTCAGTCCTCAGCAAAATGAAGGGCTCCATTTTCACTCTGTTTTATT
CTCTGTCCTATTTGCCATCTCAGAAGTCCGGAGCAAGGAGTCTGTGAGACTCTGTGGCTAGAATACATA
CGGACAGTCATCTATATCTGTGCTAGCTCCAGGTGGAGAAGGCATCAGGAGGGGATCCCTCAAGCTCAGC
AAGCTGAGACAGGAACTCCTCCAGCTCCACATAAACGTGAGTTTTCTGAGGAAAATCCAGCGCAAAA
CCTTCCGAAGGTGGATGCCTCAGGGGAAGCCGCTTTGGGGTGGACAGATGCCCACTGAAGAGCTTTGG
AAGTCAAAGAAGCATTTCAGTGATGTCAAGACAAGATTTACAACTTTGTGTTGCACTGATGGCTGTTCCA
TGACTGATTTGAGTGCTCTTTGCTAAGACAAGAGCAAAATACCCAATGGGTGGCAGAGCTTTATCACATGT
TTAATTACAGTGTTTTACTGCCTGGTAGAACACTAATATTTGTTTATAAAATGATGGCTTTTGGGTAGG
CAAACTTCTTTTCTAAAGGTATAGCTGAGCGGTTGAAACCACAGTGATCTCTATTTTCTCCCTTTGCC
AAGGTTAATGAACGTGTTCTTTTCAAATTCATAATGCTTTGAAATTTCAAATGCTGGCAAAAATGCAA
TAAAAATGCTATAAACCA
```

Humain

```
TATTTACCATATCAGATTCACATTCAGTCCTCAGCAAAATGAAGGGCTCCATTTTCACTCTGTTTTATT
CTCTGTCCTATTTGCCATCTCAGAAGTCCGGAGCAAGGAGTCTGTGAGACTCTGTGGCTAGAATACATA
CGGACAGTCATCTATATCTGTGCTAGCTCCAGGTGGAGAAGGCATCAGGAGGGGATCCCTCAAGCTCAGC
AAGCTGAGACAGGAACTCCTCCAGCTCCACATAAACGTGAGTTTTCTGAGGAAAATCCAGCGCAAAA
CCTTCCGAAGGTGGATGCCTCAGGGGAAGCCGCTTTGGGGTGGACAGATGCCCACTGAAGAGCTTTGG
AAGTCAAAGAAGCATTTCAGTGATGTCAAGACAAGATTTACAACTTTGTGTTGCACTGATGGCTGTTCCA
TGACTGATTTGAGTGCTCTTTGCTAAGACAAGAGCAAAATACCCAATGGGTGGCAGAGCTTTATCACATGT
TTAATTACAGTGTTTTACTGCCTGGTAGAACACTAATATTTGTTTATAAAATGATGGCTTTTGGGTAGG
CAAACTTCTTTTCTAAAGGTATAGCTGAGCGGTTGAAACCACAGTGATCTCTATTTTCTCCCTTTGCC
AAGGTTAATGAACGTGTTCTTTTCAAATTCATAATGCTTTGAAATTTCAAATGCTGGCAAAAATGCAA
TAAAAATGCTATAAACCA
```

Souris

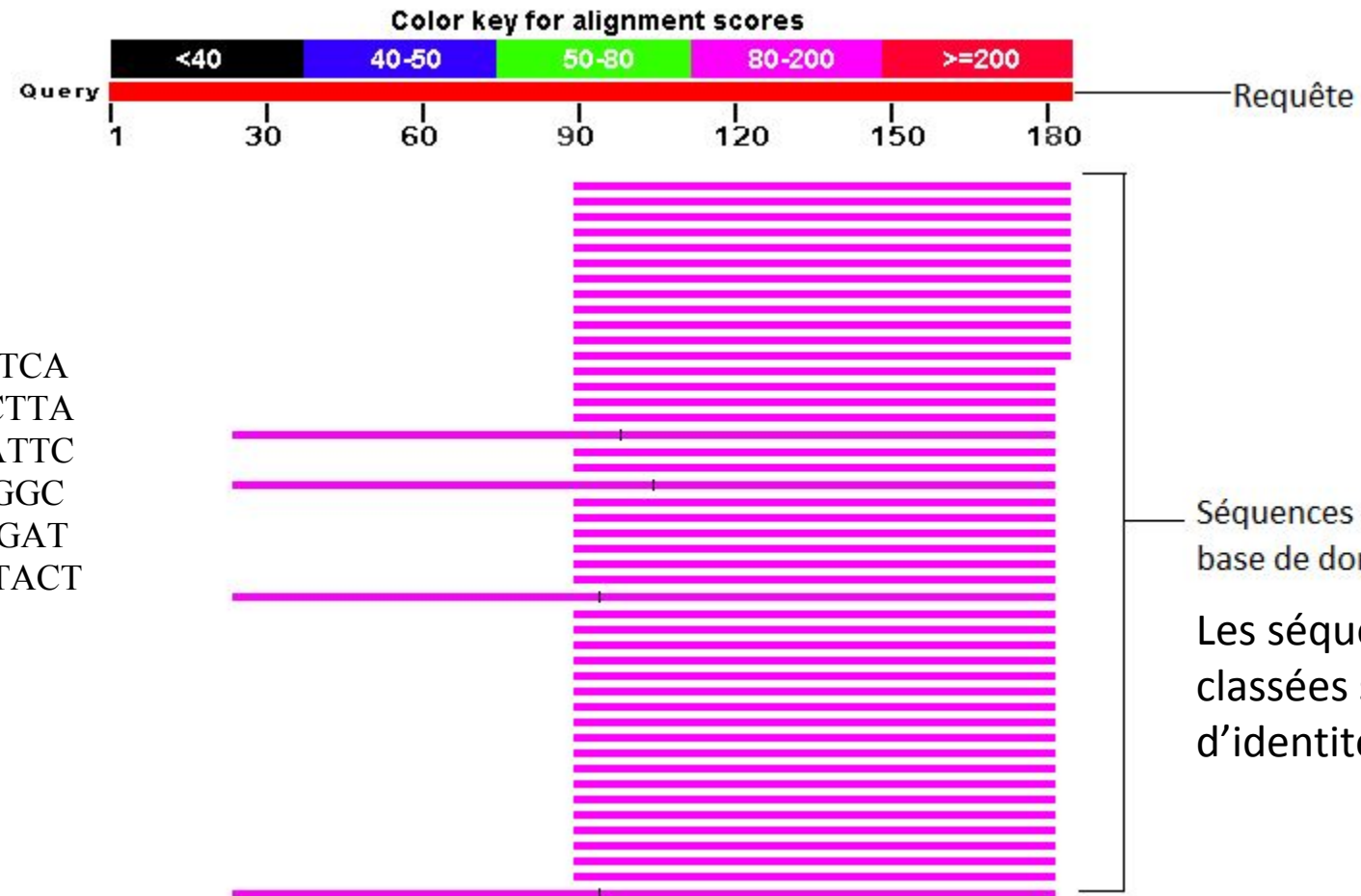
# Analyse des séquences

- Buts:
  - Comparer des séquences afin de trouver des similitudes et déduire si elles sont liées (homologues)
  - Identifier les caractéristiques de la séquence (Ex. sites actifs, introns, Exons...),
  - Identifier les différences et les variations de séquence (Ex. SNP).

# Comparaison de séquences

- Alignement: deux séquences sont comparées pour mesurer le taux d'homologie de séquence.
  - Alignement global : alignement de deux séquences sur toute leur longueur
  - Alignement multiple: alignement global appliqué à plusieurs séquences à la fois
  - Alignement local: alignement de deux séquences seulement sur une portion de leur longueur

# Comparaison d'une séquence d'ADN avec celles des bases de données



## Séquence à Comparer:

```
AGTCACGACGTTGTAAAACGACGGGATCA  
TATGACAAGATGTGTATCCACCTTAACTTA  
ATGATTTTTACCAAATCATTAGGGGATTC  
ATCAGCGTTAAACCCGCCACCAGATGGGC  
GTAAACGAGTATCCCGGCAGCAGGGGAT  
CATTTTGCCTTCAGCCATACTTTTCATACT  
CCCACC
```

# Exemple: Résultat d'alignement avec Blast

Edit and Resubmit Save Search Strategies > Formatting options > Download

**Nucleotide Sequence (184 letters)**

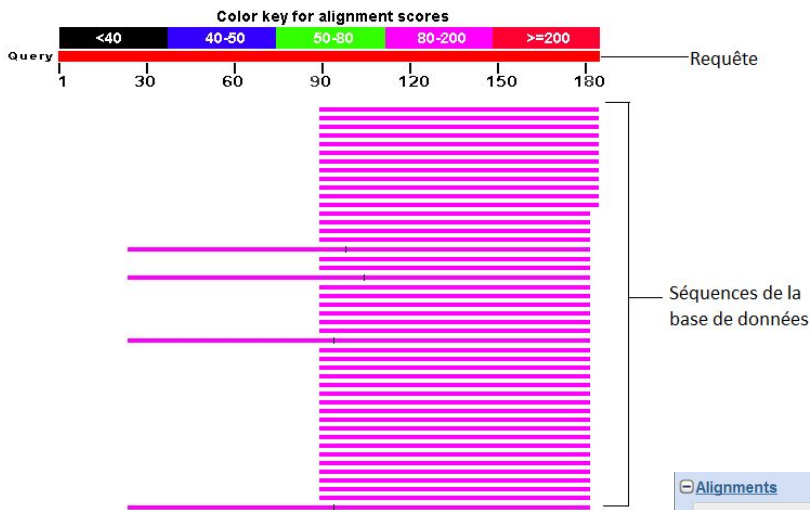
RID [H5XUJKB0014](#) (Expires on 03-03 16:18 pm)

Query ID |c|1597  
Description None  
Molecule type nucleic acid  
Query Length 184

Database Name nr  
Description Nucleotide collection (nt)  
Program BLASTN 2.2.29+ > Citation

Other reports: > Search Summary [Taxonomy reports] [Distance tree of results]

1-Récapitulatif de la requête



2-Représentation graphique du résultat

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Escherichia coli C321.deltaA, complete sequence</a>	171	171	51%	3e-39	99%	<a href="#">CP006698.1</a>
<input type="checkbox"/>	<a href="#">Escherichia coli str. K-12 substr. MG1655, complete genome</a>	171	171	51%	3e-39	99%	<a href="#">U00096.3</a>
<input type="checkbox"/>	<a href="#">Escherichia coli str. K-12 substr. MDS42 DNA, complete genome</a>	171	171	51%	3e-39	99%	<a href="#">AF012308.1</a>
<input type="checkbox"/>	<a href="#">Escherichia coli UMNK38, complete genome</a>	171	171	51%	3e-39	99%	<a href="#">CP002729.1</a>
<input type="checkbox"/>	<a href="#">Escherichia coli DH1 (ME8569) DNA, complete genome</a>	171	171	51%	3e-39	99%	<a href="#">AF012030.1</a>
<input type="checkbox"/>	<a href="#">Cloning vector pGNS-RAC, complete sequence</a>	171	171	51%	3e-39	99%	<a href="#">HQ245711.1</a>

3- Résumé du résultat

Alignments

Escherichia coli C321.deltaA, complete sequence  
Sequence ID: [gb|CP006698.1](#) Length: 4643553 Number of Matches: 1

Range 1: 70200 to 70294 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
171 bits(92)	3e-39	94/95(99%)	0/95(0%)	Plus/Minus

Features: [DNA-binding transcriptional dual regulator](#)

Query 90 ATCAGCGTTAAACCCGCCACAGATGGGCGTTAAACGAGTATCCCAGCAGGGGATCA 149  
Sbjct 70294 ATCGGCGTTAAACCCGCCACAGATGGGCGTTAAACGAGTATCCCAGCAGGGGATCA 70235

Query 150 TTTTGGCTTCAGCCATACITTTTCTACTCCACC 184  
Sbjct 70234 TTTTGGCTTCAGCCATACITTTTCTACTCCACC 70200

Escherichia coli str. K-12 substr. MG1655, complete genome  
Sequence ID: [gb|U00096.3](#) Length: 4641652 Number of Matches: 1

Range 1: 70260 to 70462 GenBank Graphics

4- Les alignements



# variabilité génétique

Individu 1



Génome 1

Individu 2



Génome ?



- 99,9% de la séquence d'ADN (pour les 2 individus) est identique
- Le reste représente **les variations génétiques**

# Polymorphismes nucléiques

- **RFLP** : Restriction fragment length polymorphism
- **Microsatellites** : SSR (simple sequence repeats) ; STR (short tandem repeats)
- **CNV** : (*copy number variation*)
- **SNP** : Single-nucleotide polymorphism

# SNP

- deux chromosomes différents sur un segment donné par une seule paire de bases
- 90% des différences sont dues aux SNPs

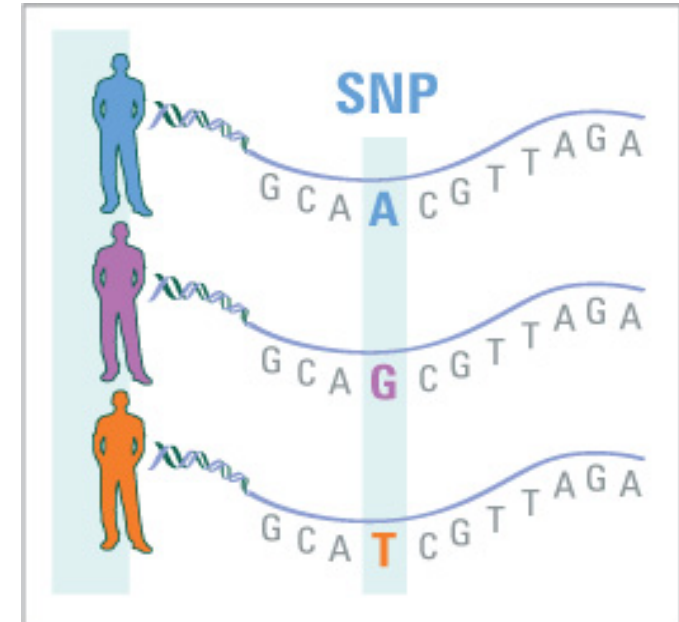
Rq: On peut avoir plusieurs SNPs dans un même gène

Ex. Gene A

ACGT ...ATC.....AGCT

ACTT ...ACC.....AGTT

SNP 1   SNP 2   SNP 4



Source:

<http://www.broadinstitute.org/education/glossary/snp>

# SNP (Suite)

SNP= single-nucleotide polymorphism

Un SNP a un identifiant unique (rs ID, e.g. rs12354)

Il existent des bases de données dédiées aux SNPs

- **dbSNP** (National Cancer Bioinformatics Institute)
- **Allele FREquency Database** (Yale University)
- **HapMap** (Projet international)

rs12354 [Homo sapiens]

1.

AACCAATTTATCACCCATGACTGAC [G/T] GAGTCTGAGAAAAGGCAGAAGAAGG

Chromosome: 12:6500105

Gene: LTBR (GeneView)

Functional Consequence: utr variant 3 prime

Validated: by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap,by submitter

Global MAF: T=0.1878/409

HGVS: NC\_000012.11:g.6500105T>G, NG\_033039.1:g.20572T>G, NM\_001270987.1:c.\*2T>G, NM\_002342.2:c.\*2T>G, NT\_009759.16:g.6440105T>G

NCBI

dbSNP  
Short Genetic Variations

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for [ ] Go

Have a question about dbSNP? Try searching the SNP FAQ Archive!  
[ ] Go

ANNOUNCEMENT

10/31/2013: RELEASE: NCBI dbSNP Build 139

Component Availability Dates:

Component	Date Available
dbSNP Web Query	October 31, 2013
FTP Data	October 31, 2013

GENERAL

HUMAN VARIATION

Search, Annotate, Submit

Annotate and Submit Batch Data

Search by IDs on All Assemblies

Note: rs# and ss# must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)



# Le projet HapMap (1)

- Projet international dont l'objectif est de cartographier les variations génétiques (débuté en 29 Octobre 2002)

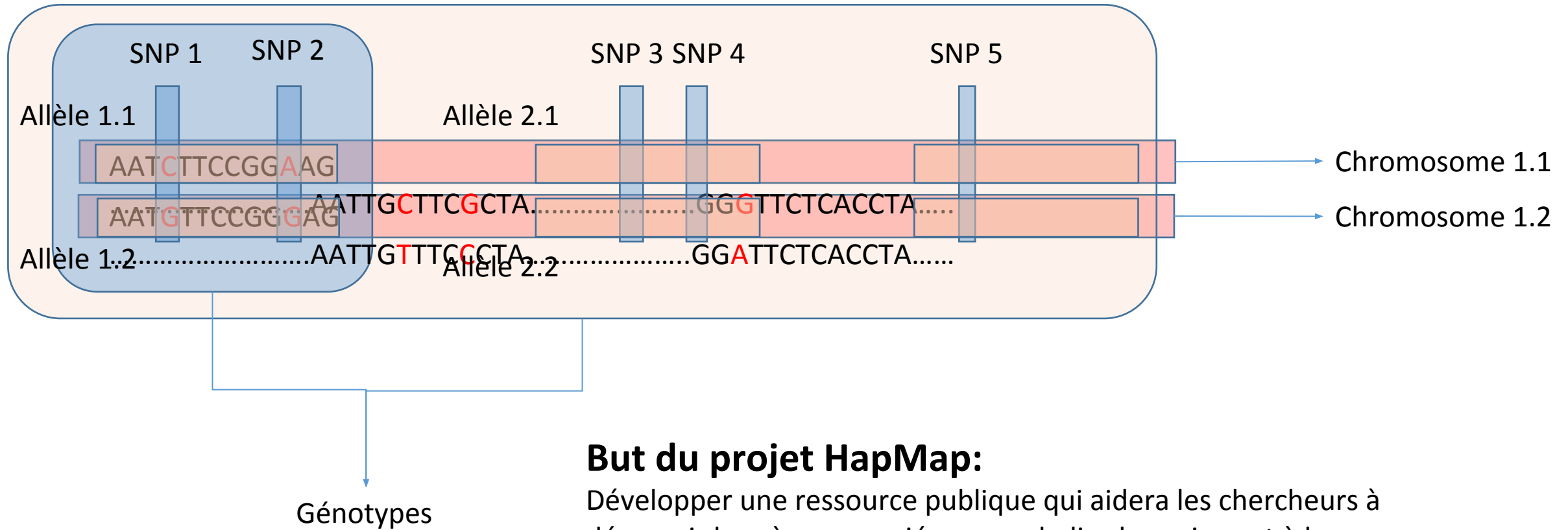


中文 | [English](#) | Français | 日本語 | Yoruba

The International HapMap Project is a partnership of scientists and funding agencies from [Canada, China, Japan, Nigeria, the United Kingdom and the United States](#) to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "[About the International HapMap Project](#)" for more information.

Project Information	News
<a href="#">About the Project</a> <a href="#">HapMap Publications</a> <a href="#">HapMap Tutorial</a> <a href="#">HapMap Mailing List</a> <a href="#">HapMap Project Participants</a>	<ul style="list-style-type: none"><li>• 2013-06-14: <b>HapMap data conversion tool</b> There are several inquires for a conversion tool to convert HapMap data into the VCF format. Please take a look of <a href="#">The Genome Analysis Toolkit</a> (by Broad Institute).</li><li>• 2012-12-06: <b>Downtime for hardware maintenance</b> From December 15 - 16, Hapmap site will be taken offline for an internal hardware maintenance. Sorry for the inconvenience.</li><li>• 2011-06-13: <b>HapMap help desk announcement</b> There was a problem with the HapMap help desk system. In the past several weeks, emails sent to <a href="mailto:hapmap-help@ncbi.nlm.nih.gov">hapmap-help@ncbi.nlm.nih.gov</a> did not reach the help desk, and thus user requests were not addressed. Please resend your email request if you sent emails to the HapMap help desk in the past several weeks. Sorry for the inconvenience.</li></ul>
<b>Project Data</b> <a href="#">HapMap Genome Browser release #28 ( Phases 1, 2 &amp; 3 - merged genotypes &amp; frequencies )</a> <a href="#">HapMap3 Genome Browser release #3 ( Phase 3 - genotypes &amp; frequencies )</a> <a href="#">HapMap Genome Browser release #27 ( Phase 1, 2 &amp; 3</a>	

# Le projet HapMap (2)

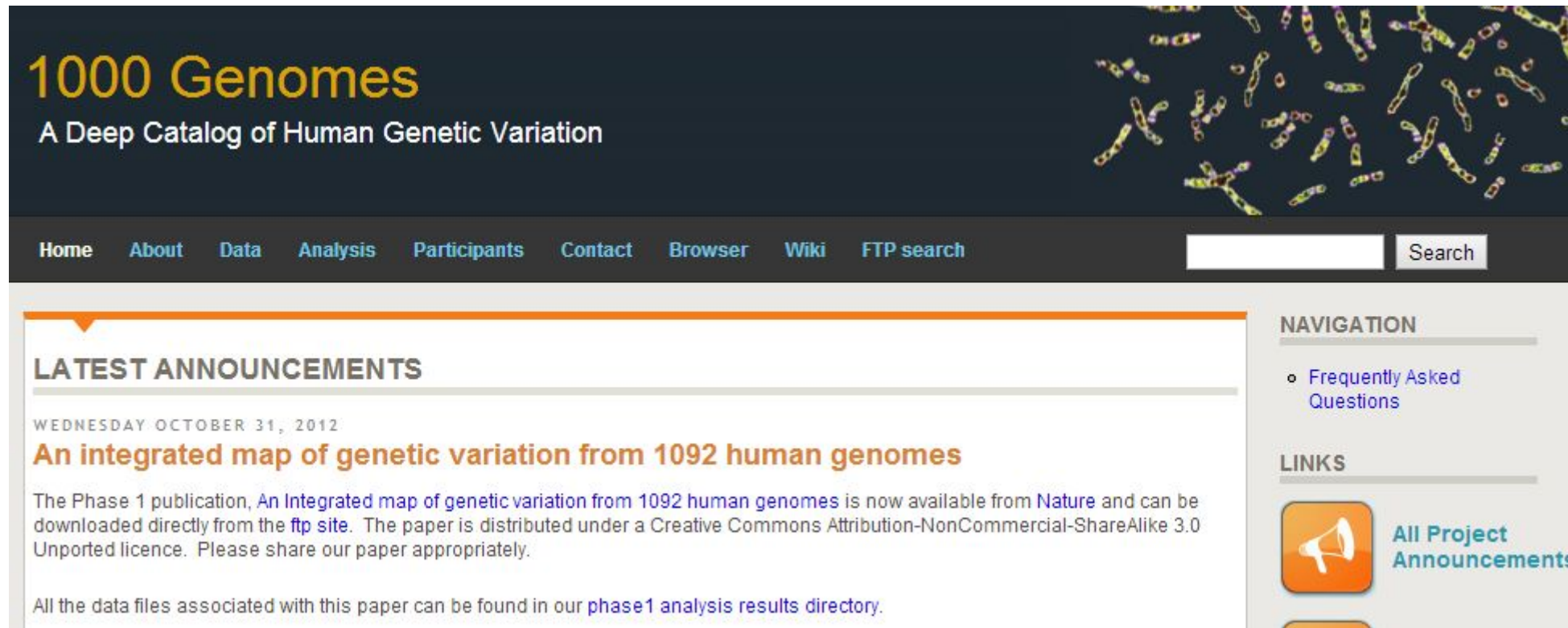


## But du projet HapMap:

Développer une ressource publique qui aidera les chercheurs à découvrir les gènes associés aux maladies humaines et à la réponse aux médicaments



# Le Projet 1000 Genomes

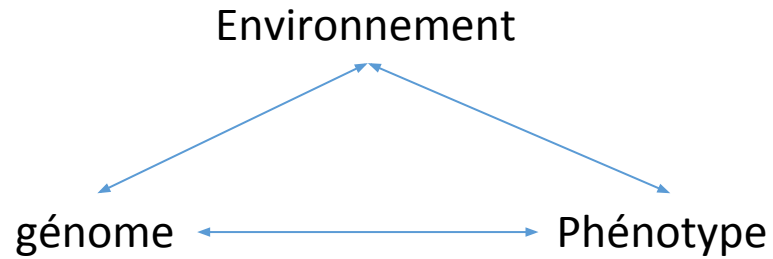


The screenshot shows the homepage of the 1000 Genomes Project. At the top, the title "1000 Genomes" is displayed in large yellow font, with the subtitle "A Deep Catalog of Human Genetic Variation" below it. The background features a stylized representation of human chromosomes. A navigation menu includes links for Home, About, Data, Analysis, Participants, Contact, Browser, Wiki, and FTP search. A search bar is located on the right side of the menu. Below the navigation, a section titled "LATEST ANNOUNCEMENTS" features a date "WEDNESDAY OCTOBER 31, 2012" and a headline "An integrated map of genetic variation from 1092 human genomes". The text below the headline states that the Phase 1 publication is available from Nature and can be downloaded from the ftp site. A sidebar on the right contains a "NAVIGATION" section with a link to "Frequently Asked Questions" and a "LINKS" section with a button for "All Project Announcements".

## Buts:

1. Création d'un catalogue complet détaillé de la diversité génétique humaine pour étudier la relation Variation Génétique &gt; Maladie
2. Amélioration de la séquence humaine de référence
3. Création d'une base de données complète (Séquence d'ADN, Variations génétiques...)

# Pourquoi étudier la variabilité génétique



## Buts:

1. Étudier la corrélation entre le génome et le phénotype
2. Étudier l'influence de l'environnement sur la relation génome-phénotype

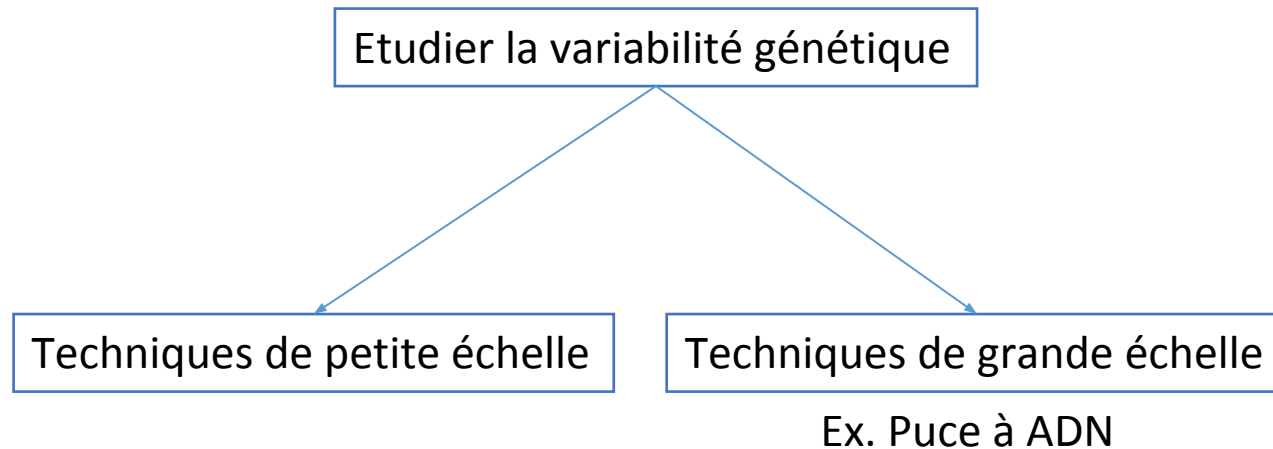
SNP ID	Indiv 1	indiv 2	....	indiv k
rs1235	CC	CT		CT
rs234	AG	AA		GG
rs1221	AT	AA		AA
....				
SNP n	....	...		...

Données génétiques

individu	age	Sex	maladie
Indiv 1	65	H	1
indiv 2	79	F	0
...			
indiv k			

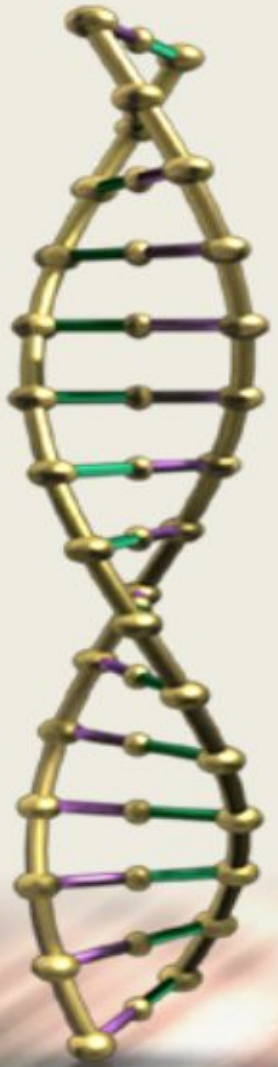
Données démographique

# Techniques pour étudier la variabilité génétique



# Références

- **What is bioinformatics?** An introduction and overview, N.M.Luscombe, Department of Molecular Biophysics and Biochemistry Yale University New Haven, USA
- [www.ncbi.nlm.nih.gov/genbank/statistics](http://www.ncbi.nlm.nih.gov/genbank/statistics)
- **The International HapMap Project**, The International HapMap Consortium
- <http://hapmap.ncbi.nlm.nih.gov/whatishapmap.html>
- **Bases de données et outils bioinformatiques utiles en génétique**, Collège National des Enseignants et praticiens de Génétique Médicale, c.Beroud



Merci