

**GENOPOLE**

VIVRE L'INNOVATION

**AG TERATEC**

**LE DÉFI NUMÉRIQUE DE LA MÉDECINE GÉNOMIQUE**

**PAR JEAN-MARC GROGNET**

**LE 10 OCTOBRE 2019**



## BIOCLUSTER : « ECONOMIE DU SAVOIR »

Concentration d'acteurs  
(recherche, enseignement  
supérieur, services,  
entreprises, hôpital) inter-  
reliés dans un domaine  
particulier,  
sur un territoire géographique.

Des savoir-faire  
complémentaires qui peuvent  
conférer un avantage compétitif à  
chaque acteur,  
une fois atteinte  
une masse critique.

La croissance économique générée se  
propage aux autres activités locales,  
nationales et internationales notamment  
dans les services et la sous-traitance.



# GENOPOLE FÉDÈRE UN ÉCOSYSTÈME EXCEPTIONNEL EN BIOTECHNOLOGIE ET BIOTHÉRAPIE... EN 2018



715 M € levés par les entreprises  
(en 14 ans)



2 413 emplois directs



3 grandes écoles

Partenariat avec  
l'Université d'Evry Paris-Saclay

avec près de 500 étudiants en  
biologie et santé par an

1 hôpital  
(1 000 lits)



16 laboratoires  
académiques

avec la présence d'acteurs  
majeurs



96 entreprises  
labellisées



1 200 chercheurs



29 plateformes  
technologiques de pointe

financées ou co-financées par  
Genopole et mutualisées au  
bénéfice des laboratoires et  
entreprises labellisées



GENOPOLE

# 4 CAMPUS, 2 NOUVEAUX À VENIR, 5,5 HECTARES EN PLUS



4,5 ha  
(2020)

1,5 ha  
15 000 sq  
mètres  
(2021)







1995  
1998 > 2017  
2018

  
GENOPOLE





?

gènes



©Grazia

©Stylbio



?

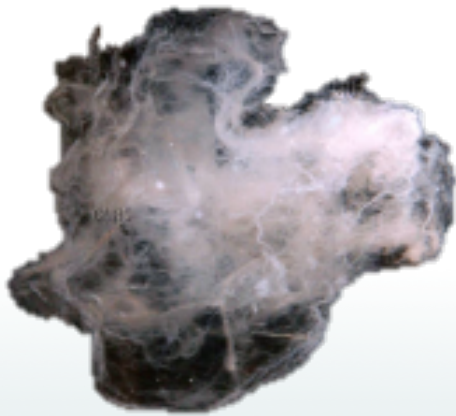
gènes



©Futura science



# L'ADN, LES GÉNOMES : C'EST QUOI ?

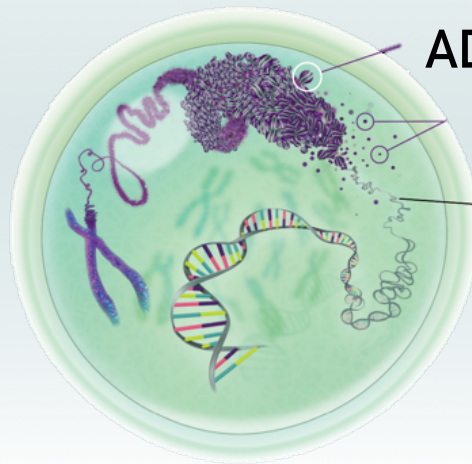


## L'ADN

une longue chaîne visible lorsqu'on l'extrait des cellules  
(pelote d'ADN).



L'ADN humain d'une cellule : 2m de long.



ADN enroulé

Protéines associées à l'ADN dans les chromosomes

Brin d'ADN

L'ADN total d'un organisme = son génome



GENOPOLE



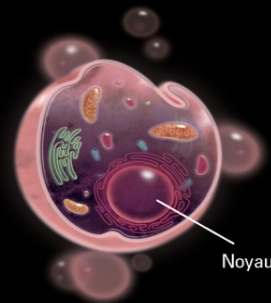
# DES CELLULES ET DES GÈNES

**Les cellules** : des formes et fonctions différentes selon les organes

Mais : **la même organisation générale**

70 000 milliards de cellules dans le corps humain !

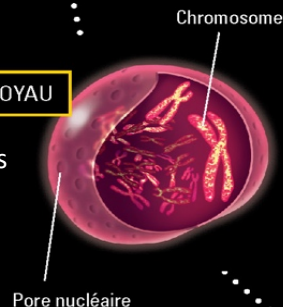
LA CELLULE, BRIQUE  
ÉLÉMENTAIRE DU VIVANT



Noyau

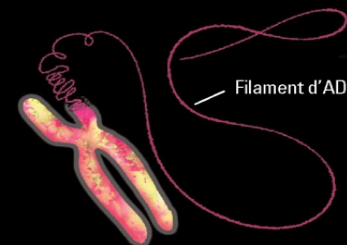
LE NOYAU

Le noyau renferme les chromosomes



Pore nucléaire

LE CHROMOSOME



Filament d'ADN

**Le message génétique**  
est contenu dans la succession  
des 4 éléments A T G C

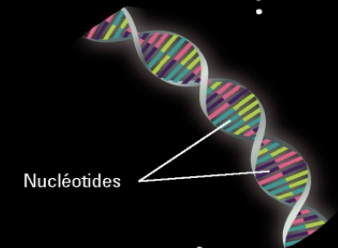
Au total pour les 23  
chromosomes humains :  
**3 milliards** de paires de bases !

L'ADN porte les informations qui  
déterminent les **caractères**  
**héréditaires** des individus (couleur des  
yeux, de la peau, des cheveux, taille...)

LES NUCLÉOTIDES



L'ADN



Nucléotides

**La quasi-totalité des cellules d'un  
organismes possèdent le même  
ADN**

**Ex de la cellule humaine :**  
paires de chromosomes

46 chromosomes au total (23 de  
la mère, 23 du père)





GENOPOLE



### STRUCTURE 3D DE LA PROTÉINE



La conformation dans  
l'espace des protéines  
est essentielle à leur  
fonction.

### Les protéines : indispensables à la vie

- Activation des réactions
- Composants de structure des cellules
- Communication entre cellules (hormones)
- Transport de l'oxygène (hémoglobine)
- ...

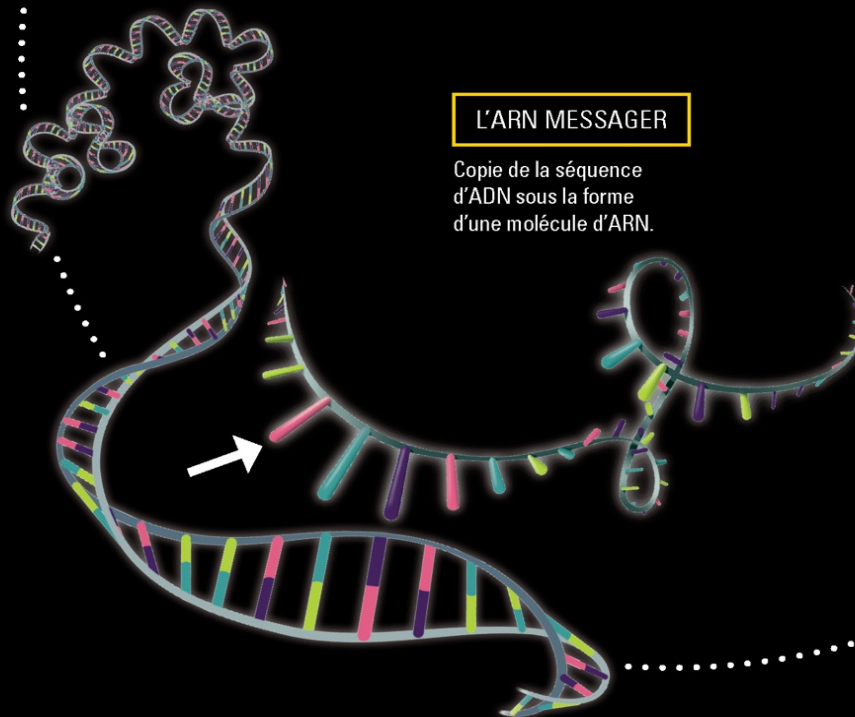
### LA PROTÉINE

Un enchaînement spécifique  
d'acides aminés



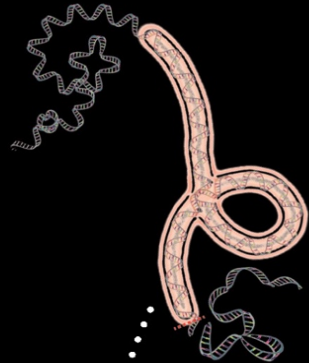
### L'ARN MESSAGER

Copie de la séquence  
d'ADN sous la forme  
d'une molécule d'ARN.



### LE GÈNE

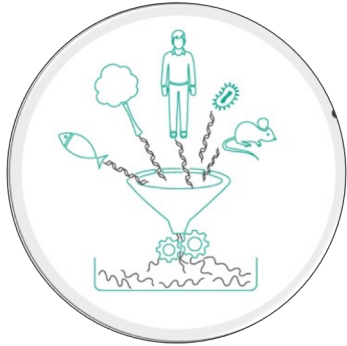
Une séquence  
spécifique  
de nucléotides.



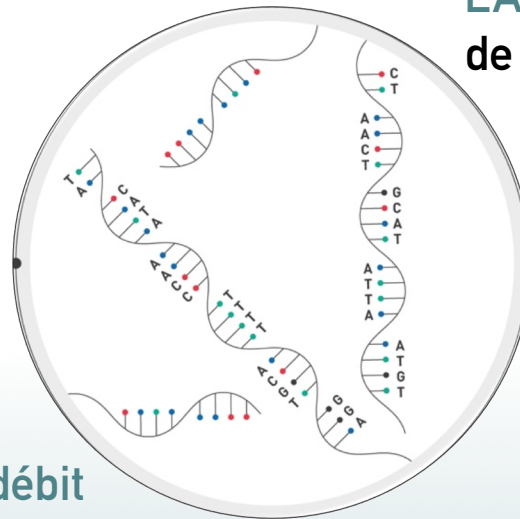
### Certaines mutations

Fabrication d'une protéine  
anormale, dysfonctionnelle.  
C'est le cas notamment des  
maladies génétiques

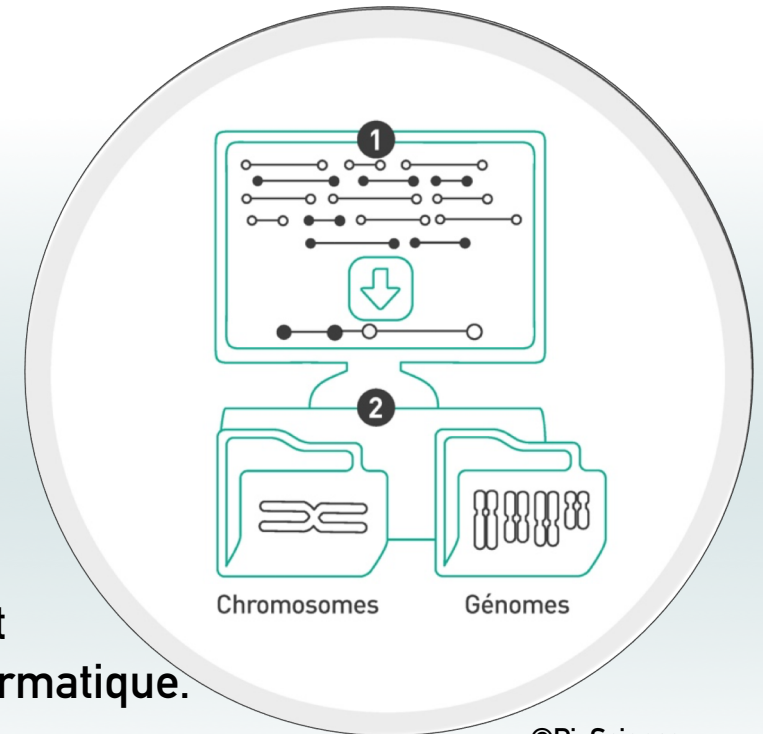
# SÉQUENCER L'ADN : LIRE LE MESSAGE GÉNÉTIQUE



Des séquenceurs à haut débit exécutant des protocoles complexes lisent la suite de lettres de chaque fragment.



L'ADN est isolé puis découpé en une multitude de fragments monobrin.



Les fragments d'ADN lus sont réassemblés par analyse informatique.

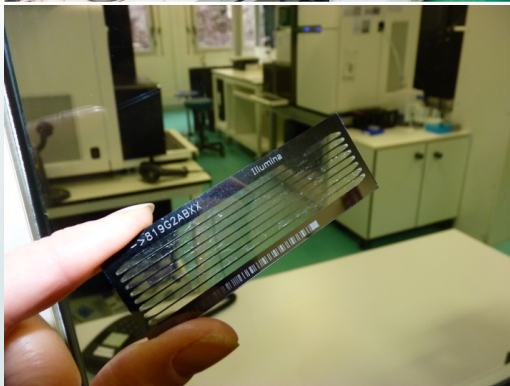
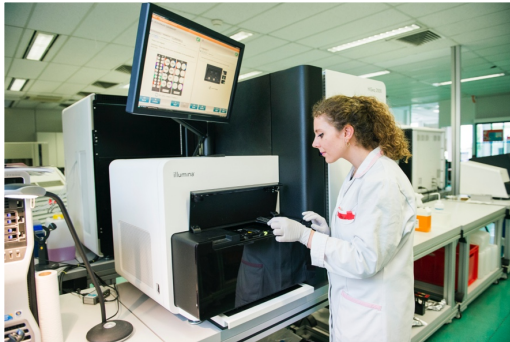
L'ordinateur reconstitue les génomes et les stocke dans de grandes bases de données.





La séquence du chromosome 14 a été révélée en 2001 par Genoscope.

Le génome humain complet (séquençage de tous les chromosomes par un consortium international) fut achevé en 2003.



### Un extrait du chromosome 14 :

...AGCATTCTGAGAAATTACTTTGTGATGTGTGCATTTCACAAAGAGTTGAACCTTTCTTTTGGTTGAGCAGTTTGAAGCACTCTTTTGTAGAATCTGTAA  
GTGGATATTTGGAGTTCTTTGAGGCTTATGGTGGTAAAGGAAATATCTTCACAAAACTACACAGAACCTTCTGAAATACCTCTTTGTGATGCTTGCATT  
CATCTCACATAGTTGAACCATCTTTTATTGAGCAGTTTGAACAATCTCCTTGTAGAATGTGCAAGTGGATATTTGGAATGCTTTGATGAGTATGGTGGGA  
AAATGAAAAATCTTCACATAAAAACTAGACAGAATTACTCTGAGAAACTCTTTGTGATGTGCACATTTCACAAATTTGAAAAATTTCTTTGATTGAGC  
AGTTTTGAAACGCTCTTTTCTAGAATCTGCCAGTGGTATTTGGAGTGCTTTGAGTCTATGGTGGAGAAGGAAATATCTTCACATAAAAACTAGAGAGAA  
GCATTCTGAGAAACTCTTTCTGATGTGTGCATACATCTCACGGAGTTGAAACTTTCTATTGATTAGCATTTTTATACACTTTTTGTAGGATCTGCAAGTTGCT  
ATTTGGAGCCCTTTGGGGCCAATGGTGGAAAAGTACTATCTTCTCATAAAAACTAGACAGAAGCATTCTGAGAAACTCTTTGGTGTGATGTGTGCATTTCAC  
CAGTAGTTGAACCTTTCTTTGATTGAGCAGTTTGAACGCTCTTTTCTGTAATCTGCAAGTGCATATTTAGAGTGCTTTGAGGCACGTGGTGGAAAAGGA  
AATATCTTCACATAAACTAGACAGAAGCATTCTGAGAAATGCTTTGTGATGTGTGCCATTCACTTCACAGAGTTGAAACTTTCTTTTCATTGAGCAGTTTGT  
AAACACTCTTTTATAGAATATGCAAGTGGATATTTGGAGCGTTTTGGAGAGAATGGTGGAAATGGAATATCTTCATATAAAAACTACGGAGAAGCATTCTG  
GAGAAACGGCTTTGTTATGTGTGACTTCAGCTCACACAGTTGAACCTTTCTTTGATTGAGCATTTTGATTCCTCTTTTGTAGAATCTGCAAGTGGATATTT  
GGAGAGCTTTAGGCCTACGGTGGAAAAGGAAATATCTTCACATAAAAACTACAAAAAGCATTCTGAGAAACTCTTTCTGATGTGTGCATACAACTCACA  
GAGTCGAAAACCTTTCTTTGATTGTGTCAGTTTGAACACTCTTTTGTAGAATCTGCAAGTGGATATTCAGAGGGCTTTGTGGAGTATAGTGGAAAAGGAAAT  
AACTTTGGATAAAAGCTAGACAGCAGAATTCTGAGAAACTCTTTGTGATGTGTGCATTCAACGTACAGAGTTGAACCTTTCTTTAGATTGGCAGTTTGTAA  
ACACTACTTTTGAATATCTGCAAGTGGATATTTGGTGACCTATGGTGGAAAGGCAATATCTTCACATAAAAACTAGACCAAGGCATTCTGA  
GAATCTCTTTGTGATGTGTGCATTCTTCACAGATTCAACTTTCTTTGATTGAGCAGTTTGGAAACAGTATTTTCTACAATCTGCAAGGGGATCTCT  
TAGCCGATTTAGGCCTATGGTGAATAGGAAATATCTTCACATAAAAAATAACAGAAACTTTCTGAGAAACTCTTTGGGATGTGTGTTTCATCTCACAGA  
GATGAAACTTTCTTTGATTGAGCAATTTGGAACCTCTCTTTTGTAGGATCTGCAATGGATATTTAGAGTGCTTTGAGGCCTGTGGTGGAAAAGGAAATAT  
CTCCACATAAACTAGACAGAAGCATTCTGGGAACATCTTTGTGATGTGTGCATTTCACAGAGTTGAACCTTTCTTTGATTGAGTATTTGGAAACA  
GTCTTTGATAGTATCTGCAGAGAGATATTTGTGAGCAATTTGAGGACTTTGGTGAGAAAAGAAATATCTTCATATAAAACCTAGTCAGAGCATTCTGAGAA  
ACTTTCTTTGTGATGTGTGCATTCTGACAGAGTTGAAACTTTGTTTGTGATTGAGCAGTTTGGAAACAGTCCTTTTGTAGGATCTGCAAGGGGATATTTCTGA  
GCCATTGAGACCTATGGTGAAGAAGAAATATCTTCACTTAAAACTAGACATAAGCATTCTGAGAAACTCTTATGATGTGTGCTTTCATCTCACAGGTT  
TGAACCTTTCTTTGATTGAGCAGTTTGGAAACAGTGTTTTGTAGAATCTGCAAGGATATTTGAGCGCTTTGACGCCTATGGTGGAAAAGGACATATCTTC  
ACATGAAATC...

# 3 x 10<sup>9</sup> INFORMATIONS : LES LIVRES DE LA VIE !







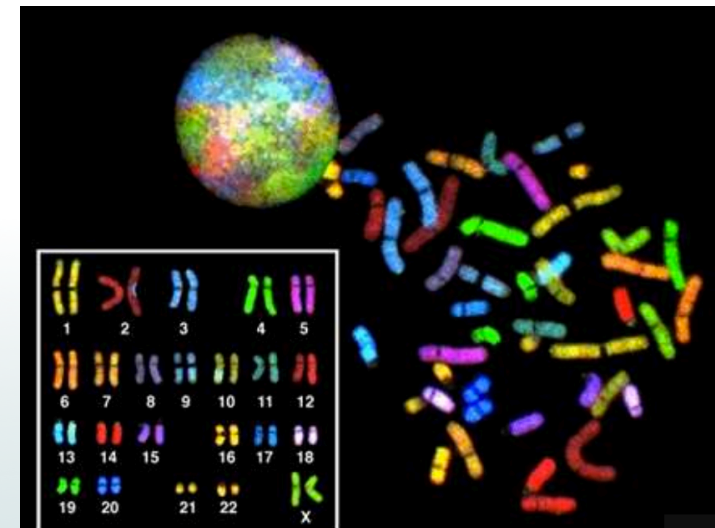
# LIRE, MAIS AUSSI...

## COMPRENDRE LE MESSAGE GÉNÉTIQUE

- Rechercher la succession de « mots » : **les gènes**
- Exemple du génome humain

L'homme qui valait 3 milliards... d'informations

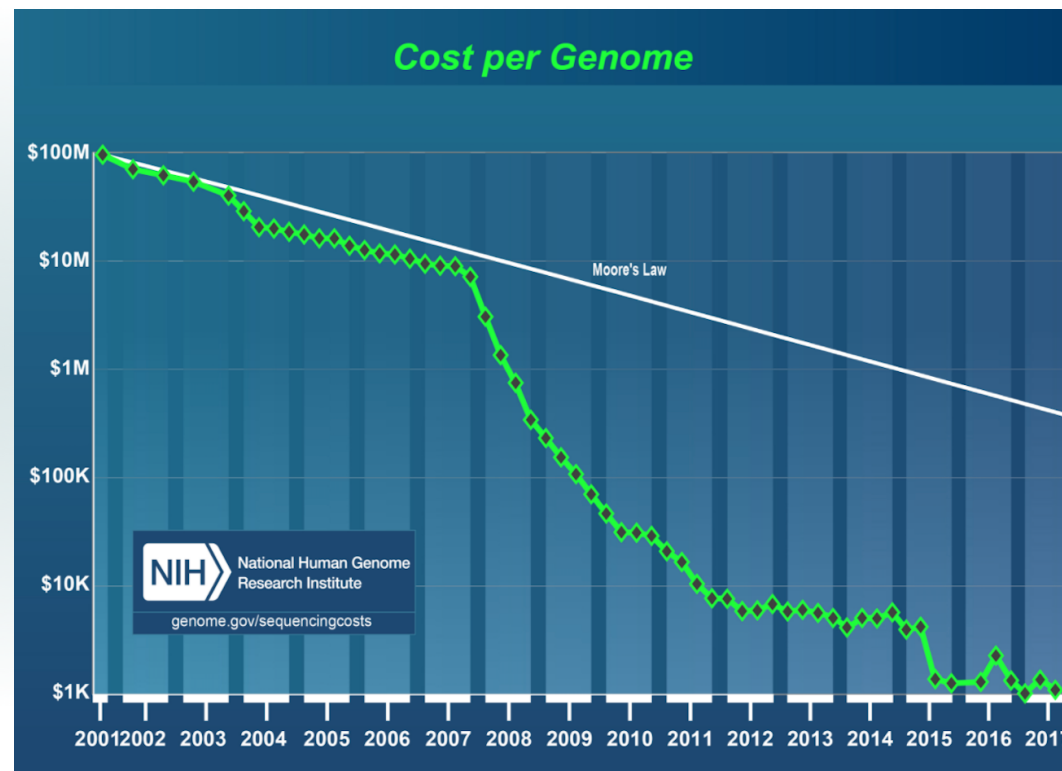
- ✓ 23 paires de chromosomes
- ✓ 3 milliards de paires de bases
- ✓ 22 000 gènes



© NIH, DP



Avec les technologies moléculaires, la robotique, l'informatique :  
la biologie vit une révolution méthodologique !  
Nous entrons dans l'ère du big data et du séquençage pour tous





# SÉQUENCER LES GÉNOMES : POUR QUOI FAIRE ?

- Explorer la diversité du monde vivant
  - Mieux connaître les espèces vivantes = **génomique**
  - Découvrir des écosystèmes complexes formés de toute une communauté d'organismes grâce à leur ADN = **métagénomique**
- Comprendre les mécanismes du vivant :  
Déterminer la fonction des gènes repérés dans les génomes = **génomique fonctionnelle**
- L'humain et sa santé = **génomique médicale**



# INTERPRÉTER LES SÉQUENCES D'ADN : CHERCHER DES AIGUILLES...

La génomique annonce le retour des grands explorateurs



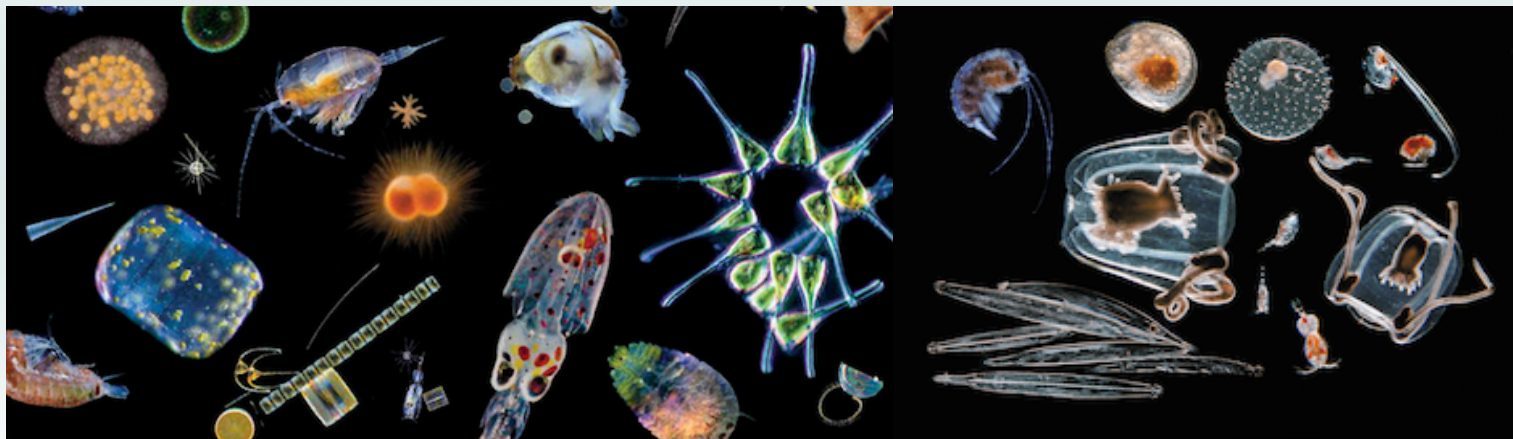
## TARA OCEANS

10-100 milliards d'organismes/L d'eau de mer

35 000 échantillons de plancton

117 millions de gènes identifiés

dont plus de la moitié jusqu'alors inconnus



© Science, 2015 - Nature Communications, 2018



# LE GÉNOME HUMAIN : PROMESSES ET ENJEUX

Temps/coût de séquençage  $\searrow$   
 mais... temps/coût d'analyse  $\nearrow$

Hétérogénéité des données  
 (gènes, ADN non codant, répétitions...)



©usine-digitale



Il faut :



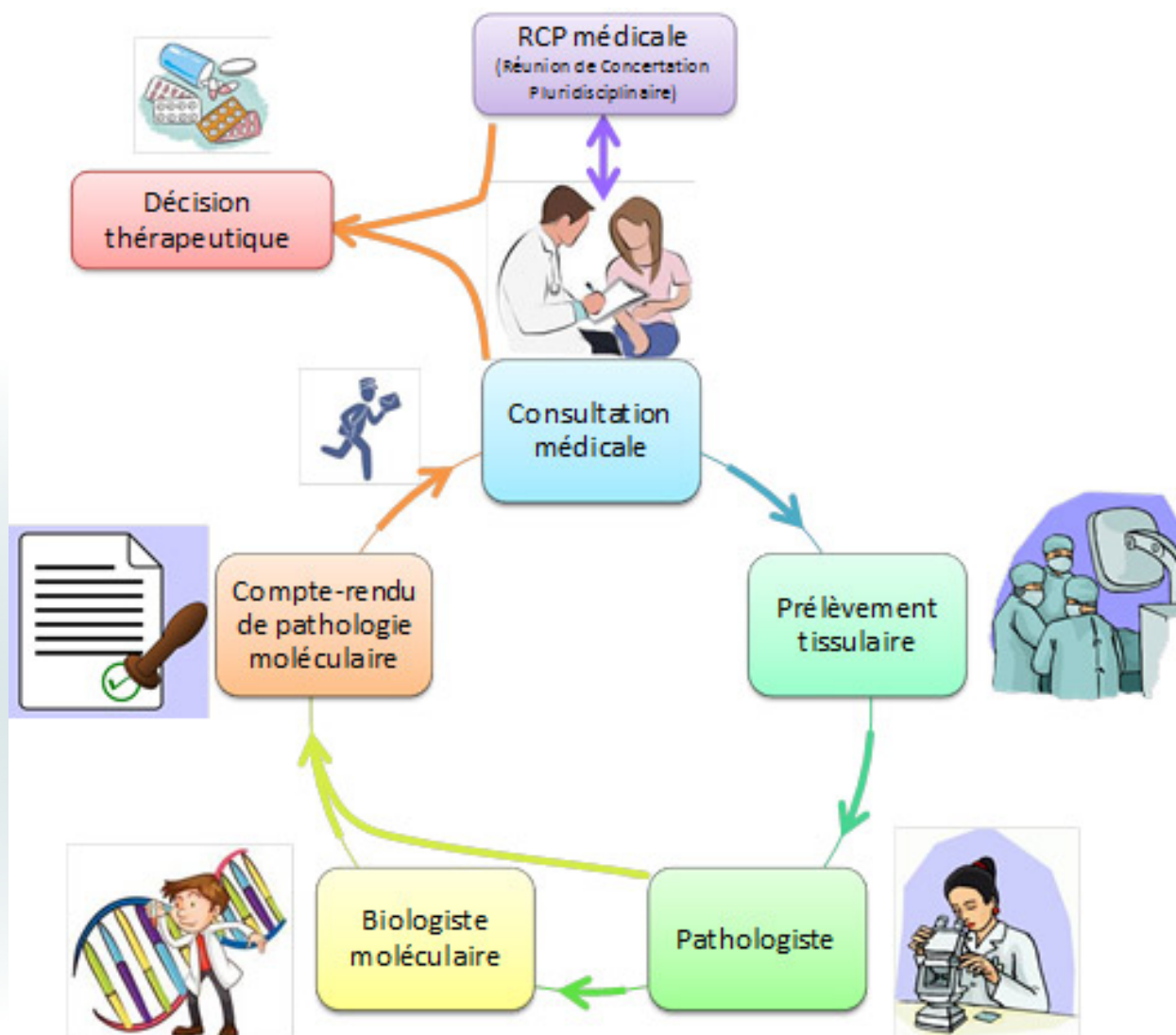
Algorithmes (math, stat) et  
 outils bio-informatiques

Interdisciplinarité croissante

Ils doivent pouvoir être exploités par les biologistes, voire les médecins



# DIAGNOSTIC MOLÉCULAIRE

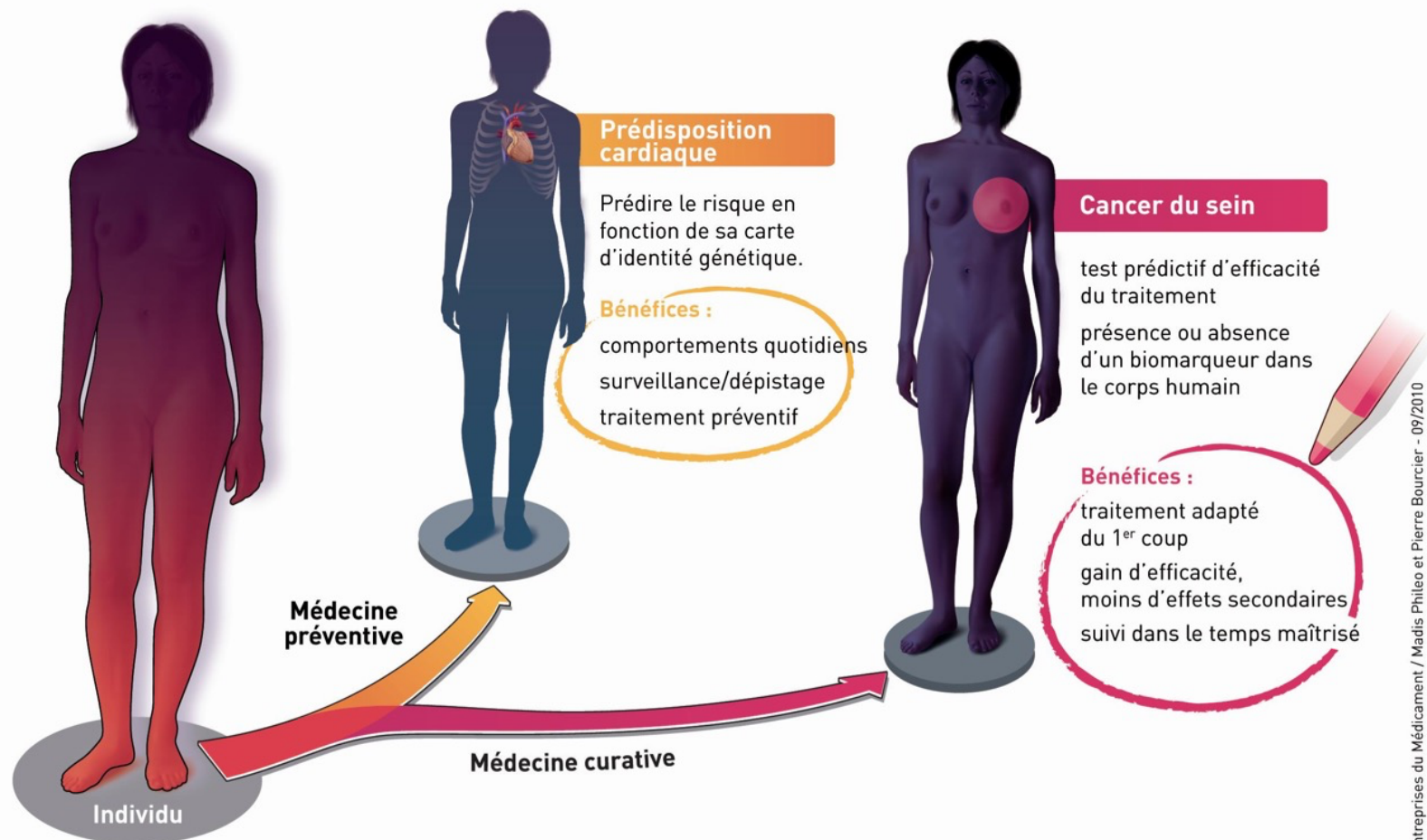


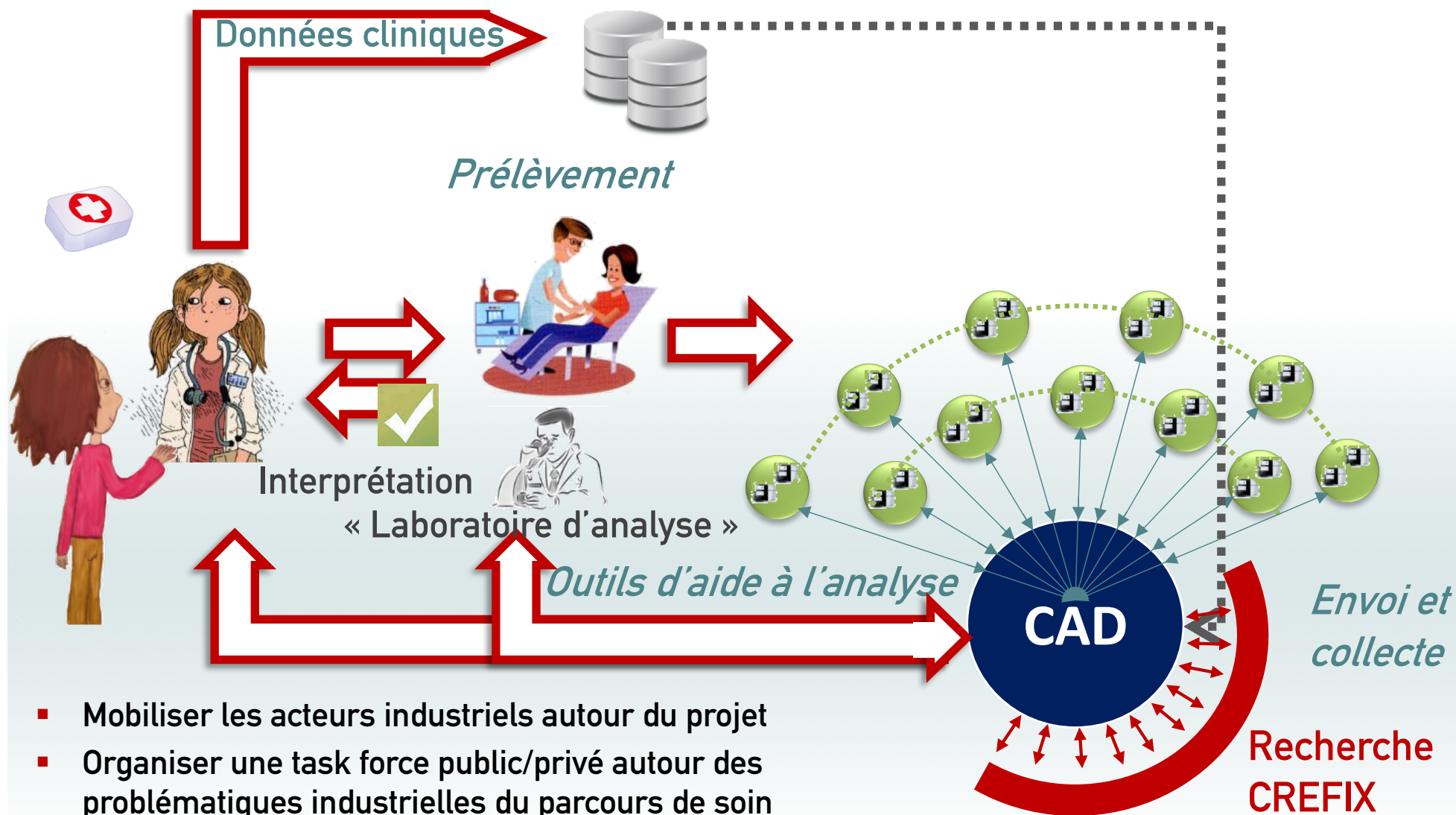
©CHU Tours



# LES DEUX BRANCHES DE LA MÉDECINE GÉNOMIQUE

## VERS UNE MÉDECINE PERSONNALISÉE





- Mobiliser les acteurs industriels autour du projet
- Organiser une task force public/privé autour des problématiques industrielles du parcours de soin
- Assurer un suivi à l'échelle internationale du champ de la médecine génomique
- Mettre en œuvre un programme dédié aux aspects médico-économiques



## LES SÉQUENCES GÉNOMIQUES : DES DONNÉES PARTICULIÈRES



- **Petit nombre d'individus par échantillons versus grand nombre de descripteurs.**  
Environ une dizaine d'échantillons (très inférieur aux autres Big data!) mais...  
des milliers ou millions de descripteurs (code génétique, transcrits.... )
- **Distinguer données brutes et données réellement exploitées.**  
Données brutes : de l'ordre de la dizaine de giga  
données « nettoyées » pour le statisticien : parfois seulement de l'ordre du méga.

Une illustration : Etude de SNP chez l'homme

Après traitement des données : tableau de 400 individus et 10 millions de colonnes !

S'il fallait le recopier, avec des cases de 1cm x 1cm,  
cela constituerait **un tableau de 4 m de haut et 10 km de large**

*Christophe Ambroise – mathématicien à Genopole – LaMME (Laboratoire de mathématiques et modélisation d'Evry)*





## PETITE ANALYSE COMPARATIVE...

| <u>Data Phase</u>   | <u>Astronomy</u>                                   | <u>Twitter</u>              | <u>YouTube</u>                                       | <u>Genomics</u>                                                  |
|---------------------|----------------------------------------------------|-----------------------------|------------------------------------------------------|------------------------------------------------------------------|
| <b>Acquisition</b>  | 25 zetta-bytes/year                                | 0.5–15 billion tweets/year  | 500–900 million hours/year                           | 1 zetta-bases/year                                               |
| <b>Storage</b>      | 1 EB/year                                          | 1–17 PB/year                | 1–2 EB/year                                          | 2–40 EB/year                                                     |
| <b>Analysis</b>     | In situ data reduction                             | Topic and sentiment mining  | Limited requirements                                 | Heterogeneous data and analysis                                  |
|                     | Real-time processing                               | Metadata analysis           |                                                      | Variant calling, ~2 trillion central processing unit (CPU) hours |
|                     | Massive volumes                                    |                             |                                                      | All-pairs genome alignments, ~10,000 trillion CPU hours          |
| <b>Distribution</b> | Dedicated lines from antennae to server (600 TB/s) | Small units of distribution | Major component of modern user's bandwidth (10 MB/s) | Many small (10 MB/s) and fewer massive (10 TB/s) data movement   |

doi:10.1371/journal.pbio.1002195.t001

Stephens *et al.*, PloS Biology, 2015

Les données génomiques humaines : **des informations à partager** pour progresser dans la connaissance des maladies, des mécanismes...

Mais aussi **des données** confidentielles à ne pas mettre entre toutes les mains.

- A qui appartiennent les données ?
- Qui les conservera ?
- Qui les interprétera ?
- Qui les ré-analysera régulièrement ?

Elles contiennent des informations qu'elles ne révéleront que lorsque les méthodes d'analyse auront progressé ! Elles renseignent sur nos ascendants, nos fratries, nos descendants.

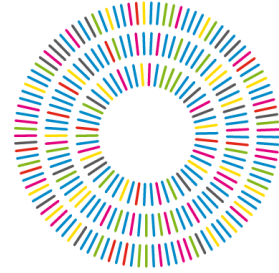


Des champs de recherche restent à développer afin que les données de la génomique puisse être exploitées pour la **médecine de demain**

- De nouveaux systèmes d'intelligence artificielle pouvant traiter les données de génomiques et autres données –omics et médicales **de manière intégré et compréhensible** doivent être développés ;
- De nouveaux outils **sécurisant les données et leur partage** devront être développés ;
- Afin d'augmenter le volume de données pouvant être traitées, les **bases de données** devront devenir **interopérable**.







# GENOPOLE

VIVRE L'INNOVATION



**NOUS VOUS REMERCIONS DE VOTRE  
ATTENTION**