

La bio-informatique : définition, applications, limitations

Hervé Seitz

10 novembre 2022

Ce diaporama est accessible à :

<http://www.igh.cnrs.fr/equip/Seitz/INRS2022.pdf>

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Définition

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses informatiques de données biologiques.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Définition

La
bio-informatique

Analyses informatiques de données biologiques.

Automatisation des tâches répétitives :

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses informatiques de données biologiques.

Automatisation des tâches répétitives :

- Gain de temps (spécialement pour les jeux de données « omiques »).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses informatiques de données biologiques.

Automatisation des tâches répétitives :

- ▶ Gain de temps (spécialement pour les jeux de données « omiques »).
- ▶ Limite le risque d'erreur.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses informatiques de données biologiques.

Automatisation des tâches répétitives :

- ▶ Gain de temps (spécialement pour les jeux de données « omiques »).
- ▶ Limite le risque d'erreur.
- ▶ Traçabilité des analyses, partage, reproductibilité.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses informatiques de données biologiques.

Automatisation des tâches répétitives :

- ▶ Gain de temps (spécialement pour les jeux de données « omiques »).
- ▶ Limite le risque d'erreur.
- ▶ Traçabilité des analyses, partage, reproductibilité.

Sur des problèmes spécifiques : analyse en intelligence artificielle (multiplier les hypothèses testées).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit ; protéomique, métabolomique, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- ▶ Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit ; protéomique, métabolomique, ...).
- ▶ Comparaison de séquences (identification d'homologues, génomique comparative, identification d'espèces dans un ADN environnemental, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Définition

La
bio-informatique

Définition

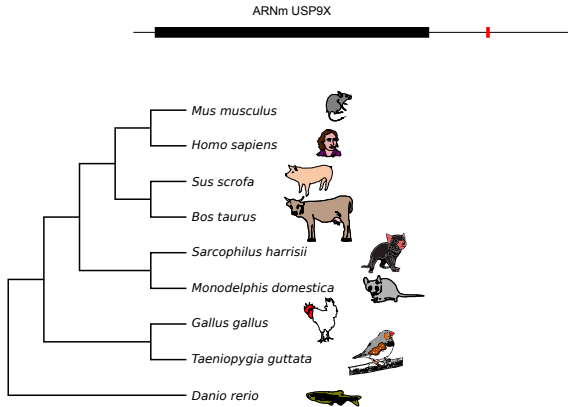
Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion



Définition

La
bio-informatique

Définition

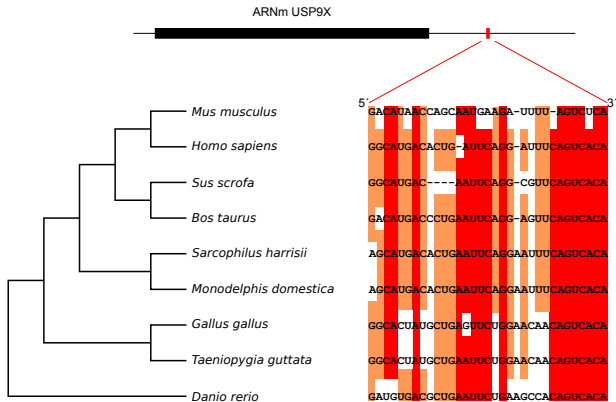
Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion



 : conserved in 8 or 9 species out of 9  : conserved in 6 or 7 species out of 9  : conserved in less than 6 species out of 9

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- ▶ Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit; protéomique, métabolomique, ...).
- ▶ Comparaison de séquences (identification d'homologues, génomique comparative, identification d'espèces dans un ADN environnemental, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- ▶ Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit; protéomique, métabolomique, ...).
- ▶ Comparaison de séquences (identification d'homologues, génomique comparative, identification d'espèces dans un ADN environnemental, ...).
- ▶ Analyses statistiques (tests statistiques, modèles linéaires, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- ▶ Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit; protéomique, métabolomique, ...).
- ▶ Comparaison de séquences (identification d'homologues, génomique comparative, identification d'espèces dans un ADN environnemental, ...).
- ▶ Analyses statistiques (tests statistiques, modèles linéaires, ...).
- ▶ Apprentissage automatique ou réseaux de neurones artificiels (classifications, reconnaissance de motifs, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Main d'œuvre : un biologiste sachant programmer, ou un informaticien connaissant la biologie (**rare !**).

Quelques champs d'application :

- ▶ Expériences à haut débit en biologie moléculaire (années 2000 : *microarrays*; depuis ≈ 15 ans : séquençage à haut débit; protéomique, métabolomique, ...).
- ▶ Comparaison de séquences (identification d'homologues, génomique comparative, identification d'espèces dans un ADN environnemental, ...).
- ▶ Analyses statistiques (tests statistiques, modèles linéaires, ...).
- ▶ Apprentissage automatique ou réseaux de neurones artificiels (classifications, reconnaissance de motifs, ...).
- ▶ Cryptographie : vérification de l'intégrité de fichiers de données.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Définition

Ah ! non ! c'est un peu court, jeune homme !
On pouvait dire Oh ! Dieu ! bien des choses en somme
En variant le ton, par exemple, tenez :
Agressif : « Moi, monsieur, si j'avais un tel nez,
Il faudrait sur-le-champ que je me l'amputasse ! »
Amical : « Mais il doit tremper dans votre tasse
Pour boire, faites-vous fabriquer un hanap ! »
Descriptif : « C'est un roc ! c'est un pic ! c'est un cap !
Que dis-je, c'est un cap ? C'est une péninsule ! »
Curieux : « De quoi sert cette oblongue capsule ?
D'écritoire, monsieur, ou de boîte à ciseaux ? »
Gracieux : « Aimez-vous à ce point les oiseaux
Que paternellement vous vous préoccupâtes
De tendre ce perchoir à leurs petites pattes ? »
Truculent : « Ça, monsieur, lorsque vous pétunez,
La vapeur du tabac vous sort-elle du nez
Sans qu'un voisin ne crie au feu de cheminée ? »
Prévenant : « Gardez-vous, votre tête entraînée
Par ce poids, de tomber en avant sur le sol ! »
Tendre : « Faites-lui faire un petit parasol
De peur que sa couleur au soleil ne se fane ! »
Pédant : « L'animal seul, monsieur, qu'Aristophane
Appelle Hippocampéléphantocamélos
Dut avoir sous le front tant de chair sur tant d'os ! »
Cavalier : « Quoi, l'ami, ce croc est à la mode ?
Pour pendre son chapeau, c'est vraiment très commode ! »
Emphatique : « Aucun vent ne peut, nez magistral,
T'enrhumer tout entier, excepté le mistral ! »
Dramatique : « C'est la Mer Rouge quand il saigne ! »
Admiratif : « Pour un parfumeur, quelle enseigne ! »
Lyrique : « Est-ce une conque, êtes-vous un triton ? »

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

md5sum : 3190eeb9f4c4d6b7982b5c4d6726b7bf

Définition

Ah ! non ! c'est un peu court, jeune homme !
On pouvait dire Oh ! Dieu ! bien des choses en somme
En variant le ton, par exemple, tenez :
Agressif : « Moi, monsieur, si j'avais un tel nez,
Il faudrait sur-le-champ que je me l'amputasse ! »
Amical : « Mais il doit tremper dans votre tasse
Pour boire, faites-vous fabriquer un hanap ! »
Descriptif : « C'est un roc ! c'est un pic ! c'est un cap !
Que dis-je, c'est un cap ? C'est une péninsule ! »
Curieux : « De quoi sert cette oblongue capsule ?
D'écritoire, monsieur, ou de boîte à ciseaux ? »
Gracieux : « Aimez-vous à ce point les oiseaux
Que paternellement vous vous préoccupez
De tendre ce perchoir à leurs petites pattes ? »
Truculent : « Ça, monsieur, lorsque vous pétenez,
La vapeur du tabac vous sort-elle du nez
Sans qu'un voisin ne crie au feu de cheminée ? »
Prévenant : « Gardez-vous, votre tête entraînée
Par ce poids, de tomber en avant sur le sol ! »
Tendre : « Faites-lui faire un petit parasol
De peur que sa couleur au soleil ne se fane ! »
Pédant : « L'animal seul, monsieur, qu'Aristophane
Appelle Hippocampéléphantocamélos
Dut avoir sous le frond tant de chair sur tant d'os ! »
Cavalier : « Quoi, l'ami, ce croc est à la mode ?
Pour pendre son chapeau, c'est vraiment très commode ! »
Emphatique : « Aucun vent ne peut, nez magistral,
T'enrhumer tout entier, excepté le mistral ! »
Dramatique : « C'est la Mer Rouge quand il saigne ! »
Admiratif : « Pour un parfumeur, quelle enseigne ! »
Lyrique : « Est-ce une conque, êtes-vous un triton ? »

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

md5sum : ece1ae779c457926347639be7d6484c0

Outils de comparaison de séquence

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.
Définir « similaire » !

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données.

Ancrage

...AAGTGCTAGCTGTAC...
| | | | | | | |
...AGGCGCTAGCCCTAC...

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données.



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.
Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données.



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.
Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données.


...NKQMEAEFRKQEEVLRKFR...
:: :| | | | | | |:::| |
...SRDLEMEHRKQEEVLKRFR...

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données (spécialement pour les protéines).

← **Ancrage** →

...NKQMEAEFRKQEEVLRKFR...

:: :| | | | | | |:::| |

...SRDLEMEHRKQEEVLKRFR...

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données (spécialement pour les protéines).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données (spécialement pour les protéines).

Alignements multiples (**Clustal**, **muscle**, **T-Coffee**, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

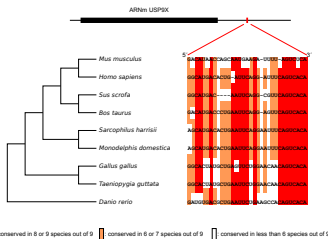
La
bio-informatique

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données (spécialement pour les protéines).

Alignements multiples (**Clustal**, **muscle**, **T-Coffee**, ...).



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : trouver des similarités entre séquences.

Définir « similaire » !

Programme **BLAST** : très utilisé pour la recherche d'homologues dans des bases de données (spécialement pour les protéines).

Alignements multiples (**Clustal**, **muscle**, **T-Coffee**, ...).

HMMer : basé sur des modèles de Markov cachés (→ poids différentiel des positions individuelles).

Outils de comparaison de séquence

La
bio-informatique

Définition

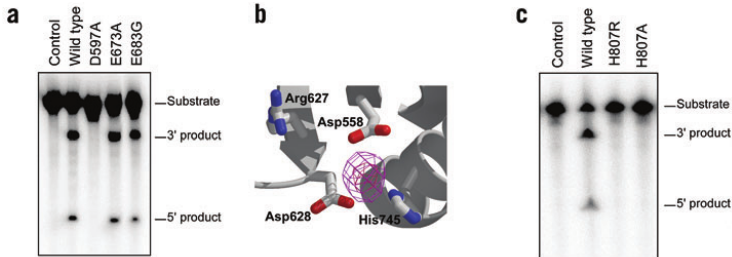
Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion



Tiré de [Rivas et al., 2005](#).

Un cas particulier (n°1) : le *mapping* de résultats de séquençage à haut débit (NGS : *next generation sequencing* ; séquençage génomique, transcriptomique, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Un cas particulier (n°1) : le *mapping* de résultats de séquençage à haut débit (NGS : *next generation sequencing* ; séquençage génomique, transcriptomique, ...).

Le format fastq :

```
@A00924:344:HFNCYDRX2:1:2101:1615:1000 1:N:0:GACAGCAG
CAATTGCTGTAGGCTTTGGACTGGAATTCTCGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCGCGTATGCCGCTTCTCTGCTTGAAAAATGGGGGGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00924:344:HFNCYDRX2:1:2101:1886:1000 1:N:0:GACAGCAG
CCATTGAGGTAGTAGGTTGTATGGTTACACTGGAATTCTCGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCTCGTATGCCGCTTCTCTGCTTGAAAAAT
+
FFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00924:344:HFNCYDRX2:1:2101:1958:1000 1:N:0:GACAGCAG
CTACTCACAGTGAACCGTCTCTTTTCACTGGAATTCTAGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCTCGTATGCCGCTTCTGCTTGAAAAAT
+
FFFFFFFFFFFFF:F:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,
F,FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,
FFFFFFFF:,,
```

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Un cas particulier (n°1) : le *mapping* de résultats de séquençage à haut débit (NGS : *next generation sequencing* ; séquençage génomique, transcriptomique, ...).

Le format fastq :

```
@A00924:344:HFNCYDRX2:1:2101:1615:1000 1:N:0:GACAGCAG
CAATTGCTGTAGGCTTTGGACTGGAATTCTCGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCGCGTATGCCGCTTCTGCTTGAAAAATGGGGGGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,,FF::,F:FF:FFFF:FFF:FFF::F:FFF:F
@A00924:344:HFNCYDRX2:1:2101:1886:1000 1:N:0:GACAGCAG
CCATTGAGGTAGTAGGTTGTATGGTTACACTGGAATTCTCGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCTCGTATGCCGCTTCTGCTTGAAAAAT
+
FFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FF,,
@A00924:344:HFNCYDRX2:1:2101:1958:1000 1:N:0:GACAGCAG
CTACTCACAGTGAACCGGTCCTTTTCACTGGAATTCTAGGGTGCCAAGGAAGTCCAGTCACGACAGCAGATCTCGTATGCCGCTTCTGCTTGAAAAAT
+
FFFFFFFFFFF:F:FFFFFFFFFFFFFFFFFFFFFFFFF:F,F,FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFFFFFFF,FFFFFF:,,
```

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Un cas particulier (n°1) : le *mapping* de résultats de séquençage à haut débit (NGS : *next generation sequencing* ; séquençage génomique, transcriptomique, ...).

Comparer des dizaines de millions de séquences à un génome animal ($\approx 10^7 - 10^9$ paires de bases) en quelques minutes, et en tenant compte des scores de qualité : [bowtie2](#), [hisat2](#), ...

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Un cas particulier (n°2) : la protéomique.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Outils de comparaison de séquence

La
bio-informatique

Un cas particulier ($n^{\circ}2$) : la protéomique.

Définition

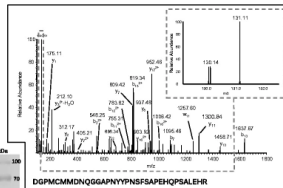
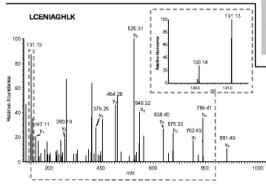
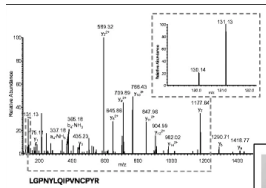
Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion



Analyses statistiques en biologie

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : entre deux mesures, il y aura inévitablement des différences. Sont-elles dues au hasard d'échantillonnage, ou à une vraie différence ?

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : entre deux mesures, il y aura inévitablement des différences. Sont-elles dues au hasard d'échantillonnage, ou à une vraie différence ?

Les méthodes usuelles (t-test, test de Wilcoxon, ...) sont peu adaptées aux gros jeux de données (problème du test des hypothèses multiples).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : entre deux mesures, il y aura inévitablement des différences. Sont-elles dues au hasard d'échantillonnage, ou à une vraie différence ?

Les méthodes usuelles (t-test, test de Wilcoxon, ...) sont peu adaptées aux gros jeux de données (problème du test des hypothèses multiples).

→ Développement de méthodes statistiques spécifiques (corrections de p -value ; modèles linéaires ; ...).

Analyses statistiques en biologie

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion



Search:

Home

Install

Help

Developers

About

About *Bioconductor*

The mission of the *Bioconductor* project is to develop, support, and disseminate free open source software that facilitates rigorous and reproducible analysis of data from current and emerging biological assays. We are dedicated to building a diverse, collaborative, and welcoming community of developers and data scientists.

Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, and an active user community. *Bioconductor* is also available as [Docker](#) images.

News

- *Bioconductor* [BIOC 3.16](#) Released.
- *Bioconductor* [Community Blog](#)
- *Bioconductor* [browseable code base](#) now available.

Install »

- Discover [2183 software packages](#) available in *Bioconductor* release 3.16.

Get started with *Bioconductor*

- [Install *Bioconductor*](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

Learn »

Master *Bioconductor* tools

- [Courses](#)
- [Education and Training](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)
- *Bioconductor* [Community Blog](#)

Use »

Create bioinformatic solutions with *Bioconductor*

- [Software](#), [Annotation](#), and [Experiment](#) packages
- *Bioconductor* Books: [csawBook](#), [OSCA](#), and [SingleRBook](#)
- [Docker](#)
- Latest [release announcement](#)

Develop »

Contribute to *Bioconductor*

- [Developer resources](#)
- [Use Bioc 'devel'](#)
- ['Devel' packages](#)
- [Package guidelines](#)
- [New package submission](#)
- [Git source control](#)
- [Git Credentials App](#)

Analyses en intelligence artificielle

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Apprentissage automatique ("machine learning") :



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

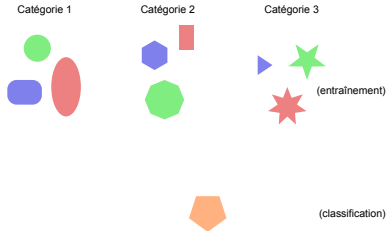
Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Apprentissage automatique ("machine learning") :



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

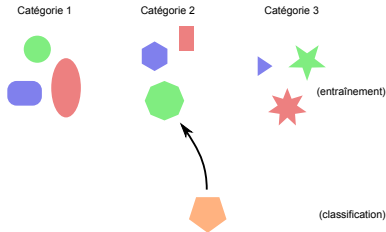
Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Apprentissage automatique ("*machine learning*") :



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Définition

Outils de
comparaison de
séquence

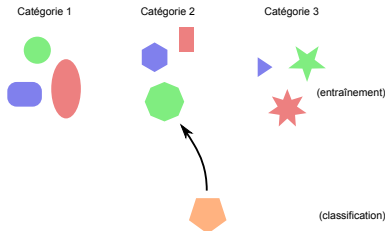
Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

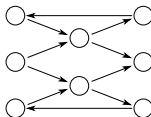
Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Apprentissage automatique ("machine learning") :



Réseau de neurones artificiel ("artificial neuron network") :



Analyses en intelligence artificielle

La
bio-informatique

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Analyses en intelligence artificielle

La
bio-informatique

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

—> Exploration d'un grand nombre de variables explicatives potentielles.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

—> Exploration d'un grand nombre de variables explicatives potentielles.

Danger : le problème de l'*over-fitting*.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

→ Exploration d'un grand nombre de variables explicatives potentielles.

Danger : le problème de l'*over-fitting*. → surtout utiles pour de très gros jeux de données (« *big data* »: génomique, grosses cohortes, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

→ Exploration d'un grand nombre de variables explicatives potentielles.

Danger : le problème de l'*over-fitting*. → surtout utiles pour de très gros jeux de données (« *big data* »: génomique, grosses cohortes, ...).

Exemples d'application :

- Problèmes de classification (→ médecine personnalisée, agronomie, ...).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Principe : émuler des processus cognitifs (apprentissage, convergence de l'état d'un réseau de neurones) par un programme informatique.

→ Exploration d'un grand nombre de variables explicatives potentielles.

Danger : le problème de l'*over-fitting*. → surtout utiles pour de très gros jeux de données (« *big data* »: génomique, grosses cohortes, ...).

Exemples d'application :

- ▶ Problèmes de classification (→ médecine personnalisée, agronomie, ...).
- ▶ Génération de séquences génomiques artificielles (pour mise au point de méthodes analytiques sans compromission du secret médical).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Un exemple de biomarqueurs : ARN circulants

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Un exemple de biomarqueurs : ARN circulants

La
bio-informatique

Principe : faciles à quantifier (NGS) ; certains sont très stables (microARN).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

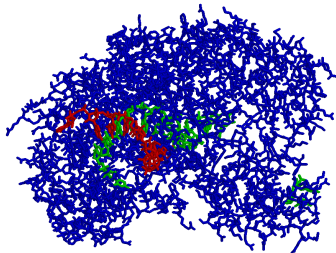
Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Un exemple de biomarqueurs : ARN circulants

La
bio-informatique

Principe : faciles à quantifier (NGS) ; certains sont très stables (microARN).



PDB n°4W5T, décrit par [Schirle et al., 2014](#).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Un exemple de biomarqueurs : ARN circulants

La
bio-informatique

Principe : faciles à quantifier (NGS) ; certains sont très stables (microARN).

Nombreux efforts d'identification de biomarqueurs parmi les microARN circulants (séquençage de fluides biologiques après exposition à des agents toxiques).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Depuis une vingtaine d'années : une révolution dans la pratique de la biologie moléculaire (l'« ère post-génomique »).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Depuis une vingtaine d'années : une révolution dans la pratique de la biologie moléculaire (l'« ère post-génomique »).

- Objectivation de la notion de « preuve ».

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Depuis une vingtaine d'années : une révolution dans la pratique de la biologie moléculaire (l'« ère post-génomique »).

- ▶ Objectivation de la notion de « preuve ».
- ▶ Expansion de la taille des jeux de données.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Depuis une vingtaine d'années : une révolution dans la pratique de la biologie moléculaire (l'« ère post-génomique »).

- ▶ Objectivation de la notion de « preuve ».
- ▶ Expansion de la taille des jeux de données.

→ L'automatisation des analyses est devenue indispensable.

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Depuis une vingtaine d'années : une révolution dans la pratique de la biologie moléculaire (l'« ère post-génomique »).

- ▶ Objectivation de la notion de « preuve ».
- ▶ Expansion de la taille des jeux de données.

→ L'automatisation des analyses est devenue indispensable.

Un bénéfice supplémentaire : traçabilité et reproductibilité.

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Partage :

- des données (ex. pour le NGS : [SRA](#)) ;

An official website of the United States government [Learn how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

SRA [Help](#)

[Advanced](#)

SRA - Now available on the cloud

Sequence Read Archive (SRA) data, available through multiple cloud providers and NCBI servers, is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

Getting Started	Tools and Software	Related Resources
Documentation	Download SRA Toolkit	Submission Portal
How to submit	SRA Toolkit Documentation	dbGaP Home
How to search and download	SRA-BLAST	BioRxiv
How to use SRA in the cloud	SRA Run Browser	BioSamples
Submit to SRA	SRA Run Selector	

FOLLOW NCBI

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

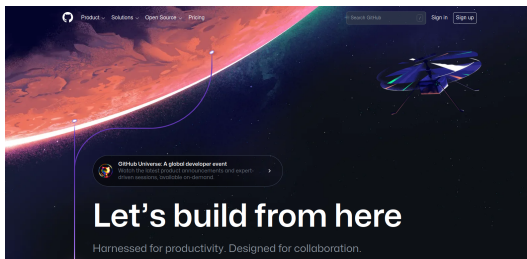
Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Partage :

- ▶ des données (ex. pour le NGS : [SRA](#));
- ▶ des scripts (ex. : [GitHub](#)).



Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Partage :

- ▶ des données (ex. pour le NGS : [SRA](#));
- ▶ des scripts (ex. : [GitHub](#)).

—> Reproductibilité des analyses à l'octet près, pour un effort modique.

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion

Conclusion : les spécificités de la bio-informatique

La
bio-informatique

Partage :

- ▶ des données (ex. pour le NGS : [SRA](#));
- ▶ des scripts (ex. : [GitHub](#)).

→ Reproductibilité des analyses à l'octet près, pour un effort modique.

Des limitations :

- ▶ disponibilité de la main d'œuvre ;
- ▶ adaptation des comportements ;
- ▶ des besoins spécifiques en termes d'infrastructures (bases de données, serveurs de calcul, sécurisation des données médicales).

Définition

Outils de
comparaison de
séquence

Analyses
statistiques en
biologie

Analyses en
intelligence
artificielle

Un exemple de
biomarqueurs :
ARN circulants

Conclusion