

Projet Long: Investigation bio-informatique du syndrome ICF par analyse de génomique

Sujet: Team of Claire Francastel
(contact: Costas Bouyioukos - costas.bouyioukos@univ-paris-diderot.fr)
CNRS UMR7216 Epigenetics and Cell Fate – University Paris Diderot

Summary

1) Introduction	2
2) Matériels et Méthodes	2
Installations	2
Jeu de données	3
Principe du programme	3
3) Résultats	4
4) Conclusion	4
5) Références	4

1) Introduction

La méthylation de l'ADN est l'une des modifications épigénétiques les mieux étudiées chez les vertébrés. Il est essentiel pour le développement embryonnaire normal et les modèles de méthylation perturbés de l'ADN sont les caractéristiques de nombreuses maladies humaines.

Le syndrome ICF (immunodéficience combinée (I), instabilité de l'hétérochromatine paracentromérique (C) et dysmorphie faciale (F)) est une maladie récessive rare. Seul 50 patients sont recensés à travers le monde (Ehrlich, 2006), aujourd'hui la tendance est stable avec moins de 100 patients atteints. Ce syndrome est caractérisé par une immunodéficience, malgré la présence de lymphocytes B, et par des réarrangements caractéristiques au voisinage des centromères des chromosomes 1, 16 et 9.

Les mutations d'ADN méthyltransférases, comme le gène DNMT3B, sont les premières causes de la maladie. Mais 20 % des patients ne sont pas expliqués par les mutations du gène DNMT3B ou le doigt de zinc ZBTB24. Ces patients possèdent des profils de méthylation différents avec des mutations sur des gènes de cycle de division cellulaire associés 7 (CDCA7) et à l'hélicase HELLS. Le syndrome ICF est caractérisé par des mutations dans DNMT3B, ZBTB24, CDCA7 ou HELLS. En 2015, l'équipe de Thijssen PE ont mis en évidence l'hétérogénéité génétique du syndrome ICF.

Plus récemment, des mutations ont été trouvées dans des facteurs ayant un rôle pratiquement inconnu. Ces nouveaux "facteurs ICF" représentent de potentiels candidats à considérer dans les voies de la méthylation de l'ADN, du développement normal et du maintien de l'intégrité du génome. Cette découverte soulève d'importantes questions qui doivent encore être abordées en ce qui concerne leur fonction, mais aussi où et comment ces facteurs sont ciblés sur le génome.

Le but de ce projet est d'identifier une signature de séquence dans les régions de méthylation correspond peut-être au site de fixation de facteur ICF.

2) Matériels et Méthodes

Installations

L'utilisation du programme nécessite l'installation du logiciel MEME, il est téléchargeable sur le site http://meme-suite.org/doc/download.html?man_type=web. Pour ce projet la version meme.4.12.0 a été utilisée. Après avoir téléchargé MEME, exécuter les commandes d'installation suivante:

```
tar xzf meme_4.12.0.tar.gz
cd meme_4.12.0
./configure --prefix=$HOME/meme --with-url=http://meme-suite.org --enable-build-libxml2
--enable-build-libxslt --with-db=../motif_databases --with-gs=/usr/bin/gs
make
make test
make install
export PATH=$HOME/meme/bin:$PATH
```

Jeu de données

Le jeu de données fourni a été obtenu par puces à ADN. Les sondes utilisées recouvrent et sont centrées sur des CpG d'intérêt. Il s'agit de séquences différentiellement méthylées du gène ZBTB de patient atteint de ICF, par rapport à des patients sains. Les séquences de la puce ont une taille fixe de 100 pb.

Principe du programme

Le principe du programme est d'utiliser le logiciel MEME pour trouver des motifs connus ou non, et ensuite vérifier la présence ou non de CpG, ainsi que la présence de chevauchement de motifs pouvant constituer des modules de transcriptions (fig.1).

MEME permet de découvrir de nouveaux motifs dans des collections de séquences nucléotidiques ou protéiques non alignées, et d'effectuer une grande variété d'autres analyses basées sur des motifs.

Voici une commande modèle pour lancer le programme:

```
./source/projet_ICF.py -fa data/genes_zbtb.fasta -n 40 -w 8 -db
bin/motif_databases/HUMAN/HOCOMOCov9.meme -o results_testdna3
```

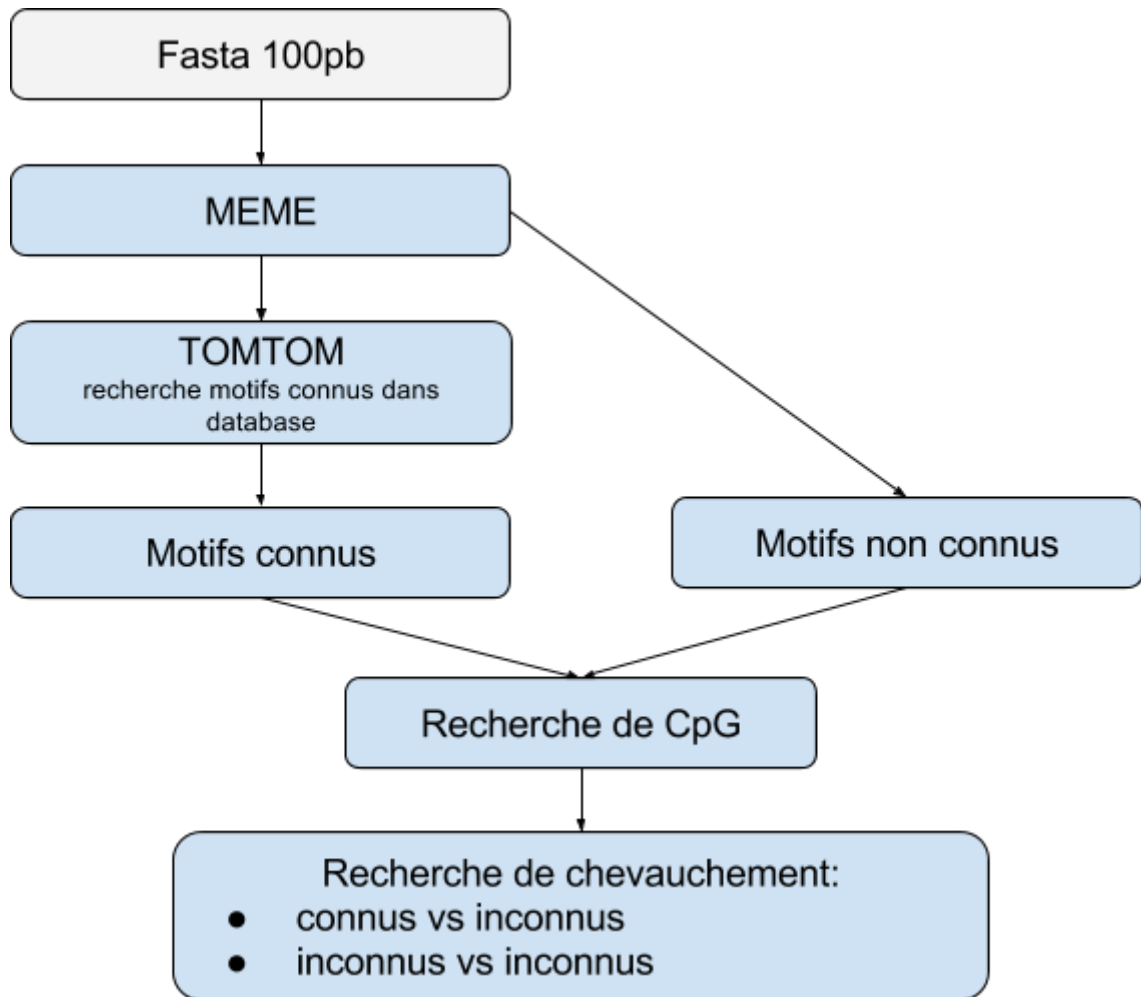


Figure 1 : workflow du programme.

3) Résultats

On obtient des fichiers avec des motifs connus et un autre avec de motifs inconnus.

Le fichier avec les motifs connus contient: le motif, l'identifiant de la séquence avec ses coordonnées chromosomiques et le nom de référence du motifs connus.

4) Conclusion

Le programme n'est pas terminé. Il reste à vérifier la présence de CpG dans les motifs trouver, et voir s'ils forment des modules de transcription s'il y a chevauchement des coordonnées chromosomique des motifs trouvés.

5) Références

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, 37(SUPPL. 2). <https://doi.org/10.1093/nar/gkp335>

Miyata, K., Miyata, T., Nakabayashi, K., Okamura, K., Naito, M., Kawai, T., ... Asahara, H. (2015). DNA methylation analysis of human myoblasts during in vitro myogenic differentiation: De novo methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation. *Human Molecular Genetics*, 24(2), 410–423. <https://doi.org/10.1093/hmg/ddu457>

Thijssen, P. E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., ... Sasaki, H. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms8870>