



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΒΙΟΪΑΤΡΙΚΗ

ΕΡΓΑΣΙΑ 3

Υλόποιήθηκε για το μάθημα “ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ”

ΦΟΙΤΗΤΗΣ

[ΛΑΖΑΡΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ-ΠΑΝΑΓΙΩΤΗΣ, 1639]

ΚΑΘΗΓΗΤΡΙΑ

[ΑΡΤΕΜΙΣ ΧΑΤΖΗΓΕΩΡΓΙΟΥ]

ΛΑΜΙΑ, 2021-2022

Περίληψη

Το κείμενο αυτό, αφορά στις θεωρητικές ερωτήσεις της τρίτης εργασίας του μαθήματος "ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ".

1 Μέρος Α

Κατά τη μετατροπή ενός Position Probability Matrix (PPM) σε έναν Position Weight Matrix (PWM) οι τιμές αρχικά διαιρούνται με μία μεταβλητή b .

1. Στην πιο απλή περίπτωση η μεταβλητή b είναι ίση με 0.25 και για τις τέσσερις νουκλεοτιδικές βάσεις (A, T, C, G). Ποια παραδοχή κάνουμε όταν χρησιμοποιούμε την τιμή 0.25;
2. Αν οι υπό μελέτη αλληλουχίες βρίσκονται σε γονιδιωματική περιοχή όπου το "G-C content" είναι ίσο με 40% ποια/ες τιμή/ές για τη μεταβλητή b θα χρησιμοποιούσατε;

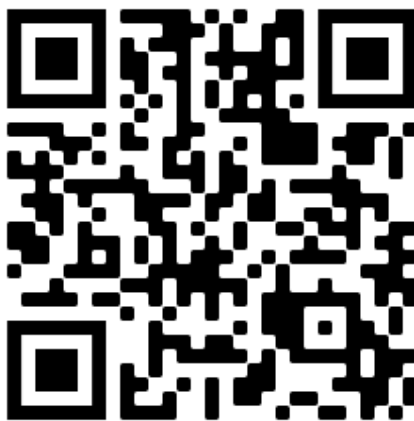
Απαντήσεις

1. Όταν χρησιμοποιούμε την τιμή 0.25 τότε κάνουμε την παραδοχή ότι για κάθε θέση του πίνακα οι πιθανότητες εμφάνισης των τεσσάρων νουκλεοτιδίων (A, T, G, C) είναι ίσες (δηλαδή η πιθανότητα εμφάνισης κάθε νουκλεοτιδίου είναι $1/4$ ή 0.25 ή 25 %).
2. Αν οι υπό μελέτη αλληλουχίες βρίσκονται σε γονιδιωματική περιοχή όπου το "G-C content" είναι ίσο με 40% τότε γ στην περίπτωση των G και C για την μεταβλητή b θα χρησιμοποιούσαμε την τιμή 0.4 (δηλαδή 40%). Στην περίπτωση των A και T για την μεταβλητή b , θα χρησιμοποιούσαμε την τιμή 0.6 (δηλαδή 60%) το οποίο προκύπτει εάν αφαιρέσουμε από το 1 την τιμή που χρησιμοποιούμε για τα G και C (δηλαδή $1 - 0.4$).

2 Μέρος Β και Μέρος Γ

Ο κώδικας της εργασίας και τα ζητούμενα αποτελέσματα είναι διαθέσιμα στο github (project03):

https://github.com/kostaslazaros/compbio_projects



Σχήμα 1: Qr code για το github repository

3 Bonus ερωτήσεις

1. Ποιο σύστημα ταξινόμησης χρησιμοποιεί η JASPAR για να ταξινομήσει τους μεταγραφικούς παράγοντες; Να αναφέρετε με ποιο κριτήριο ταξινομούνται οι μεταγραφικοί παράγοντες στις

διάφορες ομάδες, σε ποια δημοσίευση (link/authors) παρουσιάζεται αυτό το σύστημα ταξινόμησης και σε ποιο site μπορείτε να αναζητήσετε τις διάφορες ταξινομικές ομάδες και τα μέλη τους.

2. Τι σημαίνει η ονομασία “Other C4 zinc finger-type factors”, που είναι το όνομα της Κλάσης του GATA1;
3. Σε ποια Υπερκλάση (Superclass) ανήκει ο μεταγραφικός παράγοντας GATA1;

Απαντήσεις

1. Η Jaspar για να ταξινομήσει τους μεταγραφικούς παράγοντες χρησιμοποιεί ένα σύστημα ταξινόμησης που ονομάζεται TFClass. Το σύστημα TFClass, ταξινομεί ευκαρυωτικούς μεταγραφικούς παράγοντες με βάση τις περιοχές πρόσδεσης τους στο DNA. Η δημοσίευση στην οποία παρουσιάζεται αυτό το σύστημα ταξινόμησης έχει γραφτεί από τους:

- Edgar Wingender
- Torsten Schoeps
- Martin Haubrock
- Mathias Krull
- Jürgen Dönitz

και έχει τίτλο: “TFClass: expanding the classification of human transcription factors to their mammalian orthologs”.

Δίνεται ο σύνδεσμος για το paper: <https://academic.oup.com/nar/article/46/D1/D343/4566021>.

Περισσότερες πληροφορίες για τις ομάδες του συγκεκριμένου συστήματος ταξινόμησης είναι διαθέσιμες στην εξής ιστοσελίδα: <http://tfclass.bioinf.med.uni-goettingen.de/>.

Πληροφορίες σχετικά με το σχήμα (Schema) του TFClass είναι διαθέσιμες στην εξής ιστοσελίδα: http://www.edgar-wingender.de/TFClass_schema.html.

2. Η ονομασία “Other C4 zinc finger-type factors”, αφορά σε μεταγραφικούς παράγοντες οι οποίοι φέρουν μοτίβο δακτύλου τύπου C4. Κάθε δάκτυλο αποτελείται από 4 κατάλοιπα κυστεΐνης τα οποία μάλλον προσελκύουν ένα ιόν ψευδαργύρου. Οι περισσότεροι μεταγραφικοί παράγοντες GATA σε ζώα φέρουν δύο δακτύλους ψευδαργύρου.
3. Ο μεταγραφικός παράγοντας GATA1 ανήκει στην κλάση “Other C4 zinc finger-type factors”. Αυτή η κλάση είναι μέλος της υπερκλάσης “Zinc-coordinating DNA-binding domains”. Με λίγα λόγια ο GATA1 ανήκει στην προαναφερθείσα υπερκλάση (“Zinc-coordinating DNA-binding”).