



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΪΑΤΡΙΚΩΝ ΠΡΟΣΟΜΟΙΩΣΕΩΝ ΚΑΙ
ΑΠΕΙΚΟΝΙΣΤΙΚΗΣ (BIOSIM)

6η Εργαστηριακή Αναφορά στο μάθημα
**“ΠΡΟΣΟΜΟΙΩΣΗ ΦΥΣΙΟΛΟΓΙΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ”** του 9ου Εξαμήνου

του Εμμανουήλ Αναστάσιου Σερλή, Α.Μ. 03118125

3. Γονίδια και μεταβολικές ασθένειες

3.4 Πρακτικό μέρος

Ζητούμενο 3.1:

Γονίδιο INS:

Το γονίδιο της ανθρώπινης ινσουλίνης περιέχει 3 εξόνια. Το εξόνιο 2 κωδικοποιεί το πεπτιδίο σήματος, την αλυσίδα B και μέρος του C-πεπτιδίου, ενώ το εξόνιο 3 κωδικοποιεί το υπόλοιπο του πεπτιδίου C και της αλυσίδας A. Ο αρουραίος, το ποντίκι και τουλάχιστον 3 είδη ψαριών έχουν 2 γονίδια ινσουλίνης. Το μοναδικό γονίδιο ανθρώπινης ινσουλίνης αντιστοιχεί στο γονίδιο II του αρουραίου, με το καθένα να έχει 2 εσώνια στις αντίστοιχες θέσεις. Οι Deltour et al. (1993) έδειξαν ότι στο έμβρυο ποντικού τα 2 γονίδια προϊνσουλίνης ρυθμίζονται μερικώς ανεξάρτητα. Η ύπαρξη ενός μόνο γονιδίου ινσουλίνης στον άνθρωπο υποστηρίζεται από τα ευρήματα σε ασθενείς με μεταλλάξεις. Η μεγαλύτερη διαφοροποίηση μεταξύ των ειδών είναι στο C-πεπτιδίο. Τα επιμέρους τμήματα σύνδεσης του υποδοχέα έχουν διατηρηθεί σε μεγάλο βαθμό. Ορισμένες από αυτές τις θέσεις εμπλέκονται με δραστηριότητα παρόμοια με την ινσουλίνη, άλλες με δραστηριότητα αυξητικού παράγοντα και άλλες και με τις δύο.

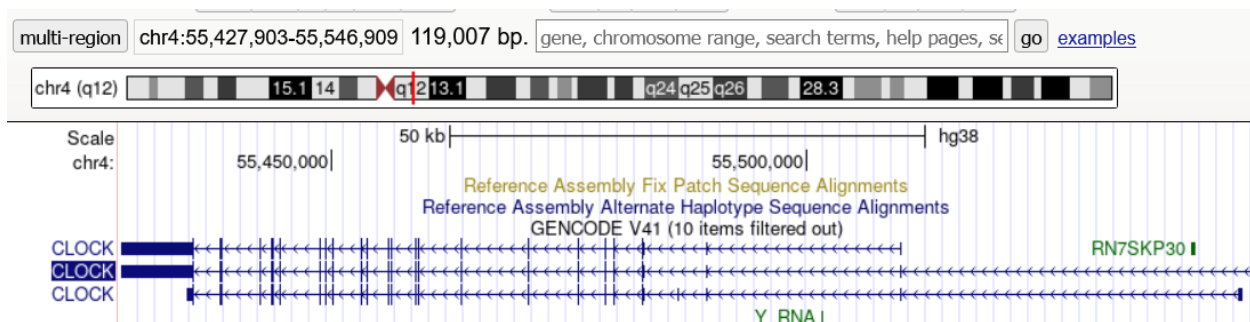
Για να προσδιοριστεί ποιοι γενετικοί παράγοντες προδιαθέτουν στους παχύσαρκους ασθενείς σε δυσλειτουργία των βήτα-κυττάρων του παγκρέατος και πιθανώς σε διαβήτη τύπου II, οι Le Stunff et al. (2000) μελέτησαν πολυμορφισμούς μονονουκλεοτιδίου στην περιοχή του γονιδίου INS σε 615 παχύσαρκα παιδιά. Βρήκαν ότι στην πρώιμη φάση της παχυσαρκίας, τα αλληλόμορφα του τόπου INS VNTR συσχετίστηκαν με διαφορετικές επιδράσεις του σωματικού λίπους στην έκκριση ινσουλίνης. **Νεαροί παχύσαρκοι ασθενείς που ήταν ομόζυγοι για τα αλληλόμορφα INS VNTR γονίδια κατηγορίας I εκκρίνουν περισσότερη ινσουλίνη από εκείνους με άλλους γονότυπους.**

Γονίδιο CLOCK:

Το γονίδιο CLOCK κωδικοποιεί έναν βασικό μεταγραφικό παράγοντα έλικα-βρόχου-έλικα (bHLH)-PAS που είναι απαραίτητος για την ρύθμιση του κirkάδιου ρυθμό στα περισσότερα είδη θηλαστικών. Οι Reppert και Weaver (1997) σχολίασαν ότι οι μελέτες για τον κirkάδιο ρυθμό θα έδιναν νέα γνώση για τον έλεγχο του ύπνου και τις διαταραχές του στους ανθρώπους, τις επιπτώσεις της εργασίας σε βάρδιες καθώς και του jet lag και τις διαταραχές ύπνου που εμφανίζονται με την προχωρημένη ηλικία. Εξέτασαν το θέμα των κirkάδιων ρυθμών και παρουσίασαν 3 υποθετικά μοντέλα για το πώς το Clock gene θα μπορούσε να λειτουργήσει ως στοιχείο κirkάδιου ρολογιού. Τόνισαν ότι το 1997 ήταν η αργυρή επέτειος της ανακάλυψης ότι οι υπερχιασματικοί πυρήνες έχουν λειτουργία κirkάδιου ρολογιού.

Ζητούμενο 3.2:

Έπειτα από αναζήτηση στην βάση genome.ucsc.edu, βρέθηκε ότι το γονίδιο CLOCK βρίσκεται στο 4^ο χρωμόσωμα, στην περιοχή q12 και έχει μήκος 119.007 βάσεις (από την βάση 55.327.903 ως και την βάση 55.546.909 του χρωμοσώματος). Τα εν λόγω αποτελέσματα φαίνονται και στην κάτωθι εικόνα που αποτελεί το αποτέλεσμα της σχετικής αναζήτησης:



Ζητούμενο 3.3:

Η ζητούμενη αλληλουχία DNA του γονιδίου INS στον άνθρωπο:

```
>chromosome:GRCh38:11:2159179:2161821:-1
GTCTGAGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTGTGGG
GGGGACAGGGGTCTGGGGACAGGGGTCCGGGGACAGGGGTGTGGGGACAGGGGTGTGGG
GACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGAC
AGGGGTGTGGGGATAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGG
GTCTGGGGACAGCAGCGCAAAGAGCCCCGCCCTGCAGCCTCCAGCTCTCCTGGTCTAATG
TGGAAAGTGGCCCAGGTGAGGGCTTTGCTCTCCTGGAGACATTTGCCCCCAGCTGTGAGC
AGGGACAGGTCTGGCCACCGGGCCCCCTGGTTAAGACTCTAATGACCCGCTGGTCTGAGG
AAGAGGTGCTGACGACCAAGGAGATCTTCCCACAGACCCAGCACCAGGGAAATGGTCCGG
AAATTGCAGCCTCAGCCCCAGCCATCTGCCGACCCCCCACCAGGCCCTAATGGGGC
AGGCGGCAGGGGTTGAGAGGTAGGGGAGATGGGCTCTGAGACTATAAAGCCAGCGGGGC
CCAGCAGCCCTCAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGGTCTGT
TCCAAGGGCCTTTGCGTCAGGTGGGCTCAGGATTCCAGGGTGGCTGGACCCAGGCCCA
GCTCTGCAGCAGGGAGGACGTGGCTGGGCTCGTGAAGCATGTGGGGGTGAGCCAGGGGC
CCCAAGGCAGGGCACCTGGCCTTCAGCCTGCCTCAGCCCTGCCTGTCTCCAGATCACTG
TCCTTTCTGCCATGGCCCTGTGGATGCGCCTCCTGCCCTGTCTGGCGCTGCTGGCCCTCTG
GGGACCTGACCCAGCCGACGCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGG
AGCTCTCTACCTAGTGTGCGGGGAACGAGGCTTCTTCTACACACCCAAGACCCGCCGGGA
GGCAGAGGACCTGCAGGTTGAGCCAACTGCCATTGCTGCCCTGGCCGCCCCCAGCCAC
CCCCGTCTCCTGGCGCTCCCAACCCAGCATGGGCAGAAGGGGGCAGGAGGCTGCCACCCAG
CAGGGGGTCAGGTGCACCTTTTAAAGAAAGATTCTCTTGGTCACGTCTCTAAAGTGACC
AGCTCCCTGTGGCCAGTCAGAATCTCAGCCTGAGGACGGTGTGGCTTCGGCAGCCCCG
AGATACATCAGAGGGTGGGCACGCTCCTCCCTCCACTCGCCCCCTCAAACAAATGCCCCG
AGCCCATTTCTCCACCCTCATTTGATGACCGCAGATTCAAGTGTTTTGTAAAGTAAAGTC
CTGGGTGACCTGGGGTCACAGGGTGCCCCACGCTGCCTGCCTCTGGGCGAACACCCCATC
ACGCCCCGAGGAGGGCGTGGCTGCCTGCCTGAGTGGGCCAGACCCCTGTGCGCCAGGCCTC
ACGGCAGCTCCATAGTCAGGAGATGGGGAAGATGCTGGGGACAGGCCCTGGGGAGAAGTA
CTGGGATCACCTGTTTCAGGCTCCCACTGTGACGCTGCCCCGGGGCGGGGGAAGGAGGTGG
GACATGTGGGCGTTGGGGCCTGTAGGTCCACACCCAGTGTGGGTGACCCTCCCTCTAACC
TGGGTCCAGCCCGGCTGGAGATGGGTGGGAGTGCAGACCTAGGGCTGGCGGGCAGGCGGGC
ACTGTGTCTCCCTGACTGTGTCTCCTGTGTCCCTCTGCCTCGCCGCTGTTCCGGAACCT
GCTCTGCGCGGCACGTCTTGGCAGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAG
GCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAAT
GCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAGACGCAGCCCG
CAGGCAGCCCCACACCCGCCGCTCCTGCACCGAGAGAGATGGAATAAAGCCCTTGAACC
AGCCCTGCTGTGCCGTCTGTGTGTCTTGGGGGCCCTGGGCCAAGCCCCACTTCCCGGCAC
TGTTGTGAGCCCTCCAGCTCTCTCCACGCTCTCTGGGTGCCACAGGTGCCAACGCCG
GCCAGGCCCAGCATGCAGTGGCTCTCCCCAAGCGGCCATGCCGTGTCGGCTGCCTGCTGC
CCCCACCCTGTGGCTCAGGGTCCAGTATGGGAGCTGCGGGGGTCTCTGAGGGGCCAGGGG
TGTTGGGGCCACTGAGAAGTGACTTCTTGTTCAGTAGCTCTGGACTCTTGGAGTCCCCAG
AGACCTTGTTCAGGAAAGGGAATGAGAACATTCCAGCAATTTTCCCCCACCAGCCCTC
CCAGGTTCTATTTTATAGATTATTTCTGATGGAGTCCCTGTGGAGGGAGGAGGCTGGGCT
GAGGGAGGGGGTCTGCAGGGCGGGGGGCTGGGAAGGTGGGGAGAGGCTGCCGAGAGCCA
CCCCGTATCCCCAGCTCTGGGCAGCCCCGGGACAGTCACACACCCCTGGCCTCGCGGCCCA
AGCTGGCAGCCGTCTGCAGCCACAGCTTATGCCAGCCAGGTCCAGCCAGACACCTGAGG
GAC
```

Η ζητούμενη αλληλουχία του ορθόλογου γονιδίου INS στον χιμπατζή

```
>chromosome:Pan_tro_3.0:11:2161377:2176144:-1
GGGGACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGTCTGGGGACAGGGGTCTGGG
GGACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGAACAGGGGTCTGGG
GACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGGA
CAGGGGTCTGGGGATAGGGGTCTGGGGACAGGGTCTGGGGACAGGGGTCTGGGGACAG
```

CGGTGCAAAGAGCCCCGCCCTGCAGCCTCCAGCTCTCCTGGTCTAATGTGGAAAGTGGCC
CAGGTGACGGCTTTTGTCTCTCCTGGAGACATTTGCCCCAGCTGTGAGCAGGGACAGGTCT
GGCCACCGGGCCCCCTGGTTAAGACTCTAATGACCCGCTGGCCCTAAGGAAGAGGTGCTGA
CGACCAAGGAGATCTTCCACAGACCCAGCACCAGGGAAATGGTCCGGAAATTGCAGCCT
CAGCCCCCAGCCATCTGCCGACCCCCCACCACAGGCCCTAATGGGCCAGGCGGCAGGGG
TTGACAGGTAGGGGAGATGGGCTCTGAGACTATAAAGCCAGCGGGGGGCCAGCAGCCCTC
AGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGGTCTGTTCCAAGGGCCTT
TGCGTCAGGTGGGCTCAGGGTTCCAGGGTGGCTGGACCCCAGGCCCCAGCTCTGCAGCAG
GGAGGACGTGGCTGGGCTCTTGAAGCATGTGGGGGTGAGCCCAGGGGGCCCCAAGGCAGGG
CACCTGGCCTTCAGCGGCCCTCAGCCCTGCCTGTCTCCCAG**ATCACTGTCTTTCTGCCAT**
GGCCCTGTGGATGCGCCTCTGCCCTGCTGGTGTCTGCTGGCCCTCTGGGGACCTGACCC
AGCCTCGGCCCTTTGTGAACCAACACCTGTGCGGCTCCACCTGGTGGAAGCTCTCTACCT
AGTGTGCGGGGAACGAGGCTTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCT
GCAGGGTGAGCCAACCGCCCCGTTGCTGCCCCCTGGCCACCCCCAGCCACCCCCTGCTCCTG
GCGCTCCCACCCAGCATGGGCAGAAGGGGGCAGGAGGCTGCCACCCAGCAGGGGGTCAGG
TGCACTTTTTTAAAAAAGAAATGAAGTTCTCTTGGTCACATCCTAAAAGTGACCAGCTCCC
TGTGGCCCAGTCAGAATCTCAGCCTGAGGACGGTGTGGCTTCGGCAGCCCCGAGATACA
TCAGAGGGTGGGCACGCTCCTCCCTCCACTCGCCCCCTCAAACAAATGCCCCACAGCCCAT
TTCTCCACCCTCATTTTGATGACCGCAGATTCAAGTGTTTTGTTAAGTAAAGTCTTGGGTG
ACCTGGGGTACAGGGTGCCCCACGCTGCCTGCCTCTGGGCGAACACCCCATCACGCCCT
GAGGAGGGCGTGGCTGCCTCCCTGAGTGGGCCAGACCCCTGTGCGCAGGCCTCACGGCAG
CTCCATAGTCAGGAGATGGGGAAGATGCTGGGGACAGGCCCTGGGGAGAAGTACTGGGGC
CACCTGTTTCAGGCTCCCGCTGTGACACCGCCCCGGGGCGGGGGAAGGAGGTAGGACATGT
GGGCGTTGGGGCCTGTAGGTCCACACCCAGTGTGGGTGACCCTCCCTCTAACCTGGGTCC
AGCCCGGCTGGAGATGGGTGGGAGTGCACCTAGGGCTGGTGGGCAGGCGGGCACTGTCT
CTCCCTGACTGTGTCTCTCTGTGTCCCTCTGCCTCGCCGCTGTTCCGGAACCTGCTCTGC
GCGGCACGCCCTGGCAG**TGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCT**
GACGCCCTTGGCCCTGGAGGGTCCCTGCAGAAGCGTGGTATCGTGGAACAATGCTGTAC
CAGACTCTGCTCCCTCTACGACTGGAGACTACTGCAACTAGATGGAATAAAGCCCTTG
AACCAGCCCTGCTGTGCCGTCTGTGTCTTGGGGGCCCTGGGCCAAGCCCCACTTCCCG
GCACTGTTGTGAGCCCCCTCCAGCTCTCTCCATGCTCTCTGGGTGCCACAGGTGCCAAT
GCCGGCCAGGCCCAGCATGCAGTGGCTCTCCCCAAAGCGGCCATGCCTGTGCGCTGCCTG
CTACCCCCACCCTGTGGCTCAGGGTCCAGTATGGGAGCTGCGGGGGTCTCTCAGGGGCCA
GGGGTGGTGCAGCCACTGAGAAATGACTTCTTGTTCAGTAGCTCTGGACTCTTGGAGTCC
CCAGAGACCTTGTTTCAGGAAAGGGAATGAGAACATTCCAGCAATTTTCCCCCACCTAGC
CCTCCCAGGTTCTATTTTTAGAGTTATTTCTGATGGAGTCCCTGTGGAGGGAGGAGGCTG
GGCTGAGGGAGGGGGTCTGCAGGGCAGGGGGCTGGGAAGGTGGGGAGAGGCTGCCGAGA
GCCACCCGCTATCCCCAGCTCTGGGCAGCCCCGGGACAGTCACACACCCTGGCCTCGCGG
CCCAAGCTGGCAGCCGTCTGCAGCCACAGCTTATGCCAGCCCAGGTCCAGCCAGACACCT
GAGGGACCCACTGGTGCCTTGGAGGAAGCAGGAGAGGTGAGATGGCACCATGAGCTGGGG
CAGGTGCAGGGACCGTGGCAGCACCGGGCAGGGCCTCAGAACCCATGCCTTGGGCAACCT
GGCCATAAGGCCCTGAGGATTGCAGCCCAGGAGAAGCAGGGAACCGCCAGGGCCACAAGG
GCAGAGACCAGGGCCAGGGTCCCCCTGCAGCCCCTTAGCCCACTGCCTCCCAGTAAGCAG
GGCTGCTTGGCTGGCTTCCTTGGCTACAGACCTGCTGCTCACCCAGAAGGGGCCACGGGC
CCTGGTGACAAGGTCGTTGTGGCTCCAGGTCCCTTGGGGGTCTTGACACAGAGCCTCTTCT
TCCGCACCCCTGAGGACAGGGTGGCTCCGCTGGGCACCCAGCCTAGTGGGCAGACGAGAAC
CTAGGGGCTGCCTGGGCCTACTGTGGCCTGGGAGGTGAGCGGGTGACCCTAGCTACCCTG
TGGCTGGGCCAGTCTGCCTGCCACCCAGGCCAAACCAATCTGCACCTTTCTGAGAGCAC
CACCCAGGGCTGGGCTGGGGATGGCTGGGCCTGGGGCTGGCATGGGCTGTGGCTGCAGAC
CACTGCCAGCTTGGGCCTCGAGGCCAGGAGCTCACCCCTCCAGCTGGGGACCTGGCCATT
GGGGCAGCCCTGTTTCTGAAGCTCTGAGCTCACCCCTTCCCCATGACCACATCAGCCCCC
TTCCACCCAAAGATGTACAGCCCCCAGCTAGCCCCGCCTCCAGAGTGGGGGCCAAGGCT
GGGCAGGCGGGTGGACGGCCGGACACTGGCCCCGGAAGAGGAGGGAGGCGGTGGCTGGGA
TCTGCAGCAGCCGTCCATGGGAACACCCAGCCGGCCCCACTCGCACGGGTAGAGACAGGG
GCGCCCTGCTGGAGCTGAGGTATGTGAGCTCGCGCGGGGCTGGGCCAAAGCGGGGCCCGG
TGGGCCGGCTGGGAGGCTGCCACCAAGTCAGCCATCGGCCAAGCTGTTGCCCTGGCTGAC
CCTGATGGCCAACAAGCCGCAGGGAGTGATGGGCAGAGGCCCGTTCTGGGAGGGGAGGG

TCAGTGCTCTGTGGGGGACCATGTGTTGGAGTGGAGGGCAGCAGGAGGAGCCCTTTGGTG
TCCAGGGACTCCTGGAGCTGCCCCAGCCTTCCAGGACTTGCAGGGCAGCTGGCACTGGCT
GGTGCTGGGGGCTGAGGAGTGTCTTTTGGGGGCCAAATTTTCTGTGACTTCTGTCCTGG
GGGACCTCTGACCTGAGGCCTCAGGAGAGGGCAAGGCCGCCACCCAAAAGAGATGCAGC
CATGGTTTCGCGGTGCCCTCAGCTGCCCTGGGCCAGAGCTGGGGCTAGCTTTCACCTTGTT
GAGACCCAGGACTCTGTCCCCAAGCCTGTCTTCGCCAGCGCCTTGACCCACCCCCAC
ACACTGTGTCCTGGAAAAGGTGGACACAGGAGACCACAGCCAGGGCGAGGTATCGCCCCCT
CCATCCCCCAGGCCCAATGAGAAGCAGTTGGCCAAGGTGATCCAGGTGGCAGAGGCAGC
ATCAGACCCAGTCTTCTGTGAGGCACCACCTTGGGTGCCGGTCCCCAGATGCCCTGGCGG
GGAGTGTGGATGCTCCCGGAGCCCCCAGGTACCCCCATGTGAGCCAGGCCCACAGAGCTT
GGCTCTGCAATGCCTGCTGGGCCGTGCCTATGCTCCACCCCTTCTGGGAAGCTAAAAGA
CAGCCCTTCAGTGTCCAGAGACCTGCCTGGCCTTGGAGCCTGGGTTTCACATGCCACCA
GGCTGGCAGGGGCACTCAGCTGCCTCCAGCCCCGGCGGTACCCTGGCATTGGGTCAATC
TAACTGCTCCGCAGTCACAAGGCAGCTGCTCCCCAAGGCTCCCCAAACCTGCTGGCCCCCT
CTAGAAGCCTCTGTCCATTCTGGAGGACCGAGGGCAGCCTGCATGCCATCCCGCACACA
GCCTTCTGTCTGGGCATCCTGCCTTCACACATGCTGCACAGGGAGGAACTCTTATACCA
CATTCTTAAGCAGAGACTGAAGCCTGGAGCCAGGCACATGGCACGTGCTCCACCCACC
CAGGACACACTGCGGTGTGGCTGCCTCCAGGCTGGCCCCCTAGATTGCGTCTGCTCCTGG
CATGGATAACTGGCACCTTTGCCTGGCCGTTGGGGCAGTGTTCCTTCCCGTGTGCGCA
GCAAATATTTACTGTCTCCGTCTCCAGGACTCTCCAGGCCTGAGCAGACCCTGGGGGGA
TGAGTGTGGACTCAGCGGTGCTGAGGGTAGCCCCCTGCCCTTCAGGTCCTGGTGCCAGC
AGGGGTCCAGCCCAGGGAAGAGACTGAGGCCAGGACAGGTAGTGTAAAGCCTGAGTTTC
TGGGAAAGGTAGCCCTGGGCAGAACTTGGGCCGAACGTGCGCCAGTGTCTGTCTCCAGCC
AGGCTGTGAGGTAGCTGTTTTCCAGGATGGGCACTTTTCCACACCTAGCAATGTGGCCAGG
AGCCGCCATTCTATGGGTGCGACCAGCAGATGGCATCAGAGCCTCACTTTTGATGCACTCC
AGCCACCAGCCACGGGTCCAGGCTCGGGCCACCACCCAGGGCACCACCCAGCAGCTGCAT
CCTGCCCCCTGCCGGGCACTCCCGGGGGCTGTGGGGCTTGTGGGGGCCCTGCCAGACACTC
TTGGGGGCTGTGAGGGGCTCTGCCAGGCACTCCAGGGACTATGGGGGCTGTGGGGGGCC
CTGCTGGGCACTCTGAAGGCGATGGGGCTTAGGAATGAGAGGAGCTGTCTGATGATGATG
GTGGGGGCGCTCAGAGGGCCCCGGACTGCTCAGGTCCAGTCTCGGCCCCTAAGTCAAGC
CTCAGGCCAGACTCTCACCAGCCTGGGTTTCTCAGACGGCCGGGACAAATGTTCTGGGTC
TCTAATATTCCAAGAAAGCCTCTGGCTGGACTCTGAGCCCCACCTGCGAGCCCCCTAGAAT
CACAGAGAGCCAGGGTGAGAAGACCAGGGGGACTCCGTCCCACCCTCGTCATGGCTGAGC
CCACCGTGGCCTGTGGTGAGCAGGCTGTGGCCTTTGCTGAGGGTCCCCAGGGCCCCCTGG
GGGCTACTGAGGCTGGAGGCCAGCGGTGGCCAGGAGGGTCCCTCCCTCAGCCACTCAAGC
CAGAAGGTGAGTCTTGGTTTCCATGTGAGGAGGGGGCTTCAGGGGCTGGGACCTGGGGG
CACGGAAGGCCTGGAGCTGGGGTCCAGGCGGCTGAGGGTTAGTGCCTTCCACGCTCCCC
TCCGCCAGCGCCGTGAGGAGAGGGAGGTCCACTCTGGAAAGAATGTTTGAGGGCAGGGGT
AGACAGGGTCTGGGAACGCGGAGATGACACACTGGGGTGCCGCTCCGCACCCTCCGCCC
TTCCTCCCACTATTGTTTTCTGGCCTCGGCCGAGGAACAGCAAGGCAGATTCTTTCGAAA
GTGGGGCCGGCCGGAGGCGGCACCCCCATCGTCCCTCTCGCGGGCCAGCCGTGCTGTGCTG
GCGCCAAGCTCCCCGCCGGCCTCCCACCCGCTCCTCCGCGTCTCAGCTCACTCCTCTGC
TCCCAGGTGGCCCCAGTCCCCACCCACCTCGCTTGACTCCCCACTTCTGCCCATCTG
CTGCTGTTGGGAGACCCTCTCCCTGACCCCAGGGACTGCTGAGGCAGCCTCCAAGCCTCC
CAGTGTCCCCGAGGCTGCCCCCTAGAAGCTGGCTTGGGGCTGTACCAAAAGGTCACCCAC
CTAAGTCAGAGTCGGGGTCCCTCAGGCCGAGGGTAAGGTAGGTGGTGCCTGCCAGGCTCT
CCTCGCCCAGGGAGGCTGAGAATGTCTAGACCCAGGTGGCTGTGACCCCTAGGCAGAGCC
AGCCCAGCCCTTGCCAGGGATGGAGACCGGCCTCGAGGAGGCCAAGCCCTGGGGGTCCAC
AGGCCTGTGGGCTTCGGGGAGGCTCTGCTCCCTGTGGCCCTGTGGCCCTGTGTGGCCCAG
GCTGCTGAGTCATCAGAACCTCGGGGGCGCCGCGGGCCCCACATTCCGCCCAGGCCTCTC
TCTGACCCTCTTCCAGCCCATCTGTGTTTTTGGAAAACAGAGCCAGAGCCCCCGCGGC
CCTGCCAGCTTGCGGTGCTCACGCTGGGACTCAAATCGCACCCCTTCTGTCTTCAAAGTC
CACCTTCACTTCAAAGCTCGGTCCACCCACCCAGCCGGCCTCCACAGGGCCACCACCTGCC
CACATCCAGGCCCCGCTGCTGCCAGTTTTCGGAGGGACCTTGGGCATCCCCTGATCCTCTC
TAGAGCGCGGGGTTCTGGCATGGGCCGTTACACATGGGTGGCTCGGTGGGTGGTGAGG
ACGGGGCTGGGAGGAGATCCTGGGGACCCCATGGTGGAGGCAATGAGGCACCCAAACCC

AACTCCAGTGATGGCTGCTTCCACGGGGCCCTCCGAGCCCTGACCTTCAAGGTGCAAGAA
AAGCTTTTCAGGGGCAGGGGTGAGTGGAAGGTGGGCTTCCCTCCCTTGCCACCTGGGGGGCG
GGCCCAGGACAGATGCTCCGTGAGCACTTCCCAACCTAGGCCCAGCTGTGGGGAAGGAGG
GAGCAGGCGGCTGGGCTCCAGGCAGGGGGAAGAGTTGCCTGAGAACTTAGGGAGAGAGGG
AGGGCTGGGGCACCCCATGCCAGCTCCAGCTGTACCACCAGAGCTCAAAGCAGAGGAGGC
AGGAGCCAGCAGGCTGGGAGGCTTCCCTGGAGGAGGTGTCCCAGGCCTCTGCCAGCACTGC
AGCCTTGAGCACGAGTCCACGCTGGGCTCCTGCTTCCAGGGCCGAGACCCAGCCCAGCA
CTGCCCATGCTGCCCTTGCCAGGGACCCTTACAGAGCTAGCACCTGGGTCTTCCCCCTG
GTTGGGCATGGGCTGGCCTCGACAATAAAGAAGGGGTAGCTGTGGCCTCAAGGTCTACAA
CAGGTTGTGATCTGTGGGTGCCAAGAAGTCATGGGCCCTCAGCATCCTCAGGGCCAGTGA
GTGTCCACCTGGGCCAGTGTGGGGCTGAAGGGTGAGAGCTGAGAGGCGGGGCTGACATCC
TGGCTCTGCTGCCAGTGGCTGTGTGGCCTTGGGAAAGTCACTTAGCATCTCTGGGCCTTG
GTTGCCTCATCCATCAGAAGGGAACCAAAGGAGGCCCTGCCTGCCTCAGCTGCTGGATGG
TG TAGCTCAGTGTTGTCTGTGAGAGGCAGGCCCTTGGGCAGCCGTGCCTCTGGCTTCAGT
ACACCCGTGAGCTGCTTTCTGTGCCAGGCCTTCAAATCACGGGGACAGCCCCGTGGGTG
TG GTTCAGGTCTGGGCACCACAGGCAAGGGGAAGCTGACATGTCTCCGTGGGAAGTGGC
AGTCCCCATGGCTACCGGGCATGCACAGGGCGTGTGGAAGGGACCTTGGAGGGAGGAGCA
GCTCCCCGAGGGAGGAGCCTGGGGCTGGGGTACAGAGGCCTCTGCACATCTTAGAGTAAA
ACAAGCAGGAGAGGCTGGGTGCGGTGGCTCATGCCTATAATCCCAGCACTTTAGGAGGCT
GAGGCGGGCAGATCACCTGAGGTGAGGAGTTCAAGACCAGCCTGACCAACAGGGAGAAAC
CCCATCTTTACTAAACTACAAAATTAGCTGGGTGTGGTGGCACATGCCTGTAATCCCAG
ATATTCCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAAGCAGAGGTTGCGCTGAG
CCGAGATGGCACCATTTGCACTCCAGCCTGGGCAACGAGAGCGAAACTCCGTCTCAAAAAA
ACAAAAACAAAAACAAAAAATCAAAACAATCAAAAAAACAAGCAGGAGGGGC
TCTGAGGTGCCTGCAACACCCAGGTACAATCCGTGGCCCTGAGGCCCATCACAGGGAAGG
GGTCTTTGCAGCTCTTTCAACCCCCAGCCCAGCATCCAAGGAAGCCCAGGGCAGGGAGAA
ACCTCAGCTGCACCATCAGAGCTCAGAACAGAGAAGGCAGAAATTAGCAGGGAGTGGGGC
TGTGGAGGCTTCCCTGGAAGAGGTGTCTCCCGCCTTGCTGGCACTGAGGCCTTGAGGATGG
GTCCACTACTGGGCCCCCTGCTCCAGGGATGCAGATCCGGCCCACTGCTGAAATCTGTGCT
CCTGGAGACTCCCTCCTCTTTCATGGGCCACAGGCTGTGAAAACCCCATAGTCTCTCCAGG
CAGCAAGTTTTGTTTTGTTTTTTGTTTGTATTTGTTTTTTGAGAGTCTTGCTCTGTCA
TCCAGGCTGGAGTCCAATGGCGCGATCTCGGCTCACTGCAACCCCTACCAGCCAGCAAGT
TTTGCTGGAGCAGGTGGAGGAGTCTTTGCAGCGAGCTCCCGAGCTGCCTAGGCCTGCCGA
AGAGTTGCTGAGCACATGACCACCAAAAAAGGTTTCGGCACCCGTGGGTAGGGGAGGCAG
GGGCAAATGTCACTTTACGTGAGCGGAGGATAAAATTGAACGTTTATTAACATTGATTT
TGTGCCAGAGGTGGCTACCCAGACCCATCCACTGCAGCCGTGAGGTTGGAAGGGAGAGGG
GAGGCATGAACTGGTTTGTCCAAGGTCGTTGGCAGAGCGCGGCTGTGAACTGGGTTAGCC
CCCAGGCCCCCTCCTGCCCATCCATCCTCCAAGCACTGATGTGTGAGTGGCACAGTGGTCC
CCTTGGCCCTGCCAGCCCCCTCGGGCTGTGGCCCTATAGCCTTTGAAGGACGGCCTCACTT
CCCCAGCAGAGGCAGCCTGGCCTTCCAGTACCTCGATGGTGTAGGGCTCTGGGGCCACAA
TCTGGTGTTTTTTCAGGACTAGGTTGTGCGTTACTTCTCCTCTTCTCTCAATTCCCAAGGT
TTGTTAAGGGGCCCTGTGGCTGTGGTCATGGCAATATGGATTTACTGGCGGTGCTTTCTG
AAGCTGTGGATGCAATTCTCACAAGTGGCCACATGGTGTGAGGCTCGGGCCAGATTTGGA
AATCTCCGCTTTACAGGTGGAGCAGGGTGGGGCTGGGAAGCGAGTTAAGAACGAAATTCA
GCCCATCTGGGCTCATCCTAGGGACGCTGTCTCTGAAAGGGGGTCTCAACGTTCTGCCCC
GGGCCTCTCAGTAATATATCTCATTTTTATATTGCTAATGTAGTGGTTTCTCTCTTA
GAACAAAACGGATGCTTCAGAAAGGTGTTCCAGACCTATCCTGTGGCTTAGGCCCTAAGC
TTTAGGATGCTCACCAGGCCACAGCCCCCTCGGGTTAGGAGACACCCGAAGGGAGGAATT
CCCTCAGCGCCCCCTCTCAGTCCTGTAATGCCAGGACAGAAGCAGATAAGAGCCCCGGTT
CCAAACACAAAGGTGCAGGACCCATGGCACCCCTGACACTGTAGGGAGGCGTCTGACTTT
GTCCTCAGAGTCCCCCTCCAGAGTGCCCCAAATTGCGGCGCCCAAGACCTTTGGGTTATG
CCACTGTAATTCTAGAAATGGCCGCCAGCGGTGAGCATAAGGCCACTTAAGAAGCTAAAT
GGCGAAAGCTTGGTGTAGCAGATTTCTGGGGGTTCCCAAGACTCTCATGCATGAACAAGG
GTGGGCTGAGCCTCTCCCAGGCCTAAGAGGCAGTGGAGGAAGGAGGTTTTTCTATGTCGT
GTCCCTCAATCCTGCACCCCAGTTTTTGTCCATTTTTCGGTCCCTCTCACCTGTCTCCAT
AAAGCATGTGTCCCCCAGACCCTTGCGTGGAATCTGTGGAATGTAGTGCTGTGGGGCC
ACAGCCCTCACACCCGAACCCACTGATCCCCAATGCCTCTAGTCCTGCCTCGGTGGCACC

ATGGCTTAGCTGCTCCCCGGGGGCCCTTTCTATCCTGACCGCTTCTCCCTGCTCTCCTTC
TCCTGTTTTACACACCCTGGGTTCAGATCCTGATGGGATAAAATAGCCTTGGAGACTAG
AGTGAGCTCCTCCTGGGCCCATCCAGAGATGGAATTCCAGTCCAGGGCAGAGTTTGGATT
CAAGACACCATTGTGGGCCTGGGAGGCTCCAGGTTTCCAGGAGTTTGACCTGTGGGCTGC
TTCCCTTTTCAGCCCCGTAGCTGCCCACCTCCTCAACCTGGGGACTCTTCATTTGGGGGGC
ATCCCATGGAATCACTTCATCTGGCTGGGTGTGAGCCCGTCACTCTCTCTAGGTCCATCT
GCAGCTTGGAAGTCCCAGCCAAGTGTCTCCAGAGGGCTCTGGCAGTCGCTGTTTGCAAC
ATTTGGGCATCTTGCAGCAGACACTGGGACCAGCACATGCTACGGGGTCTCTAAGCCATA
GAGCCCTGGAGCAAGGGCTGCTTCATGCTGACCTCCTTGTATCTTTGTTTTGTTTCTTTC
CCTGAGTCTTTGTTTTGTGCTTTCTCTTCCCTGCAGGCTTGAGGAAGAGGTGTGCAATTCC
CAGTAAGACTTGATCTAATCCAATGGGTGATTCTATCTGAGCCCTCCCTGAGCTGCTGCC
AGCTCTGAGTCTGTGCTTTAGCTGGAAGACTGCGAACCATTTCAGTGTCTGGACCTGA
CAGGGCCATTAAACATCCCTCCCCACTCAAAGCCGAGTGAAAAATCCATTTTCCCTCCT
CCCATGGGAGGGCATGGCTCTGGGGCATCTACGTCTTATTTGAGACATCCCCAACACAT
GCCAGATTTGAATCCCCTTCAAGTCTAACTCTCTTCCCTGCTGAAATCCTTGTTATCCCC
ACTAGATGTCCTATTCTCCAGCTAATGAAGGGATTTAAGACGAGGGCTTGAAAGAAAAA
GGGAGAGGGGACAATTAAATCCCAACACAGGCTAGTAAACTAATCACACTGAAGGTGGT
TGGGGTGCATGGATGTTGGGAGGGTAGGGACTGGTGGCTCTCTGGAAAGTACACGTCTTT
GCCTCTTCCCTGAGTCTTTGCTAGCTTGGGGAAGAGGTTAGAAATTTCCAGTGTGATTTAG
TCCAATCTGGTGGATGCTTCTATATGAACCCTCACCCAGCTGCTGCCAGCTTTGAGTCTG
TGCTTTAATTAGAAAGGCTGCGAACCTGTCCAATCGCCCAATCCAGAAATGCCTGCAAGT
GCCCCCTGCAGCAGTGTCAATTAGAGCTTAGTACCACACAGGGTGTGTCCGAGGTGAGATC
AGACGTGACCAGTGAAGTGAAAACCCCCCAGCTGAGTACCTTGCTCATGGAAGATGGAA
GGACCTAACTGCACAGTACATAACTGTACTGGGAGCACCTCAAATCCAATTTTCTGCTC
ATTCTTGAATGCTGCTTTGTCCCCCTTATCCCCCCCCCCTTTTACCAGTGCCACGTCC
ACCAACATTCCGGGGTGTCAAGTAACTGCCAAGTGTCACTCTAAGTAAAGCTACACCCAC
TCCCCACCACCTCCACATAGCCCCCACCTCCTAGCTGGCAGGGAGCTTCTGGCTTATGCC
CACGCCCCACAGGCGCCTTTCTGCCAGGTGAGGGGTGGGCCAAACCTCCACCCCTAATGT
ACCATGCCCTGGTGCTGGAAAGTGCCTGAGCCAGCTGCCCCAGCGGCCTCAGCACTACCA
AGTTGGCACAAGTCCCCAAATTCGGAGGGGCTCAGGGAACGAGTGGAGGGGATGAGG
AGGTGAGGGGTAAACCCATCATTTTCAGTTGGCATTGAGCAGGTGCCATGCTCAGCGGAG
ATGAGGCTCTCCCATCTGTAGGGGCCGTATTAACATGCACACTCTAAAAGTGCCCTTCGT
GTCTCCAG**CCTCAGCTTTGTCCCTCTCCTCCTCCACGTCAACCTGGCCAGAGGGTCTGGA**
CGCCACAGCCAGGGCACCCCTGCTTTGGTGGTGA**CTGCTAATATCGGCCAGGCCGGCGG**
ATCATCATCCAGGCAGTTTCGGCAGAGAGCCTTGGAACAGTGA**CTCCCCGGTCTCTT**
TATCCACTGTCCAGGAGCTGCGGGGACTGCGCAGGGACTAGAGTACAGGTAACCTGGGCTC
CCATCCCTTGAGCGTGAGAGAAAAGCTGCACCTTAAAGTGTTTCAGGCGTGGGTGGGCCGGG
AGCTTGGGGCAGCTGCCAGGATCCTGGTTTCTGAAGGAGGGGAAGAACTGCTGCTGCTGG
AGGGTGCAGGGAAGCCTCCTGAGAGCAGCCTCAACTTCAGGGATGGGGTGTGCAGGAAAG
GCCATTGTGGAGAGGGTTCTTCTTTAGGGCTGCACAAAGCCACTGAGGCTTTTGCAAGGA
AAATAGGTTTTCTTGACTAATTCACCAAGCAAAATGGGAGGGGTAGGGGAGGAGGGCTA
GGCTGCTCTTCCCAGCGGGAACACACAGCTGTCTTCACAAGTGTGAAAGGAAGAGTCTTT
CTGTGTGAAAAGTTTCTCCCGTTGCATCCCCCATCCCATTCCCAGAGACAAACAGGAGA
CTTTGCAGAGGAGCCAGGGGCCCCGAGATTCTGGCGCAGATATTTTATTTATATATATA
TATATATACACCATTTTAAAGGTAAAGCTTCCTTCCCTCCTGCCTCCCTATGCCTCCTGA
CCACCAGCAAGAAATTGGACAGGAGACTGAGGAGAAATGCCGGGAGAGGCAACAACGGCC
CTCCATGTCCCCCTAGGTTTAGCTTCTCTCCTCCTGATGGCGCACCTGGTCCCCCTTGCT
GCCCTCCCAGCCTCCCTGGCACAGAGAGGGGCCCTGGGGCCAAGGCAGTTTCCCTGGGAA
TGCTCATTCATGCATGAAGTTTTTCTCTGTGTCACCCTGGACCCAGACTCCTCAATCCAC
CCAGGGTGGTGTCTGTGGGGCGGGGTTTCAATTTCCCCAGGAAGCACAGCCACCCCTCCC
TCACTGGCCTCGTCAAGCAGAGCTGTGTGTCCAGTGGCTTTTGCTGGGGCCCCCTCCTTA
TCTCCTTCCAAGGTGGGGGTGTTTGGAGGTGGAGGAGGCTTTCATATTCTGTGCCATGAC
CCCTCAAGGCGGGCCATTCTGTGTGCACCCTCCACCCCAAGTGCCAGGCAGAAGCCCATCC
TCACCCAGGAACAGGGCAGCTGTCCAACAGAAGGGTCTCGGCCTCTCCATCAGCACCGGG
AAGCCCTTTCTAGGCAAACTTCTCACCATTCTTCCCTCCCTTATACTTTGAAAACGGGA
GCTCTAGGCGGGGGAGGGGCTAGAGGGGGAGCCGCTGCCCAAATCCTGACAAGGTGACCT
GAAGGAACCCGGGGAGGGGGATGGCACAGGGCTCAGGCTTGGGGTGTATGGGGAGGGGG

GCTTTGCTTTTAAAAGAGGTCATCTCAGCAATATCTTTTTGTTTTTCCCCAG**GGGCCGAA**
GAGTCACCACCGAGCTTGTGTGGGAGGAGGTGGATTCCAGCCCCAGCCCCAGGGCTCTG
AATCGCTGCCAGCTCAGCCCCCTGCCAGCCTGCCCCACAGCCTGAGCCCCAGCAGGCCA
GAGAGCCCAGTCCTGAGGTGAGCTGCTGCGGCCTGTGGCCAGGCGACCCAGCGCTCCC
AGAACTGAGGCTGGCGGCCAGCCCCAGCCTCAGCGCCAAGTGCAGAGGCAGAGAGGTGAG
TGTCTCAGGCACCCTGAGGCCTGGCAGAGAGGGCCACAGGCTCTGCGCGGGAGTCTTCGA
ACTGGGATCTCCCCCTTCTGCAAGCAGCTTTGGCTCAGAGAGGCTGGCGTGGATTTCAGTC
ACACAGCTGGGATCTGGAGTTCCGTGGTTGGCTCCAGGTGCTTCCGTCCAGGGGCCAGAG
CAGGTGTGGGCAGAGCAGGTTCCCCGCAGTCTCCACGGCACCGAGGTCTTGGCAGGGGAG
CTCCTGGGAGAGGAAAGAGGGCAAAGAAGGGGAGAGGGGCAGGGAGAGAGCGGGCAGCCA
AAGGGGAGAAGATGGGGGGCAGAAAGTGGGTAGAGAGGGGAAAAAGGGAAAAATATCATTGG
GGAAGAACCCAAAAACCAAGGAAAGCTGGGCTCTGCTGGGGGCTGTGAGACCCCCGGGT
TCTCCCCGCCCCAGGCTGCTGGCCATGGGGTCTTGACCAATGGCCTGACCTTTCTGTCA
GTCTGTATTTATCAAAGTGGGTGACAGTCTCAGGCCTCCTGGCTGTTTCAAGATTGAGGTA
ATAACCAG

Η ζητούμενη αλληλουχία του γονιδίου INS (insulin I) στον ποντικό:

>chromosome:GRCm39:19:52252135:52254514:1
GAGGTCTGATACTTTGGGGAAAAATTAGGATGAGTCCAAATATTCTACCTAATTCAAATG
ATTGTTTTCAAAAAAAAAAGGAATTTGAGCATGTATACAAAGAATTCTGTAATAACTATATA
GAACTCTTCTTATATATGCTCAAATTTTACATGCTAGCCTTCAGGTACATATCTTGGGTT
GTTGGGTATTGTAGAAGAATGTACTACAGGGCTTCAGCCCAGTTGACCAATGAGTGGGCT
ACGGGGTTTTGTGAAAGGAGAGATGGAGAAGGAGGGACCATTAAGTACCTTGCTGCCTGAG
TTCTGCTTTTCTCTCCCTCTGAGGGTGAGCTGGGATCTCATCTGAGTTAAGGGCCCAGC
TATCAATGGGAAGTGTGAAACAGTCCAAGGGACATCAATATTAGGTCCCTAACAACTGCA
GTTTCCTGGGGAATGATGTGGAATGCTCAGCCAAAGATGAAGAAGGTCTCACCTTCTG
GGACAATGTCCCCTGCTGGGAAGTGGTTCATCAGGCCATCTGGTCCCTTATTAAGACTAT
AATAACCCTAAGACTAAGTAGATGTGTTGATGTCCAATGAGTGCTTTCTGCAGACCTAGC
ACCAGGCAAGTGTGTTGGAACTGCAGCTTCAGCCCCTCTGGCCATCTGCCTACCCACCCC
ACCTGGAGACCTTAATGGGCCAAACAGCAAAGTCCAGGGGGCAGAGAGGAGGTACTTTGG
ACTATAAAGCTGGTGGGCATCCAGTAACCCCCAGCCCTTAGTGACCAGCTATAATCAGAG
ACCATCAGCAAGCAGATGTACTCTCCTCTTTGGGCCTGGCTCCCCAGCCAAGACTCCA
CGCACTTTAGGGAGAATGTGGGCTCCTCTCTTACATGGATCTTTTGCTAGCCTCAACCT
GCCTATCTTTTCAG**GTCAATTGTTTTCAACATGGCCCTGTTGGTGCACTTCCTACCCCTGCTG**
GCCCTGCTTGCCCTCTGGGAGCCCCAAACCCACCCAGGCTTTTTGTCAAACAGCATCTTTGT
GGTCCCCACCTGGTAGAGGCTCTCTACCTGGTGTGTGGGGAGCGTGGCTTCTTCTACACA
CCCAAGTCCCGCCGTGAAGTGGAGGACCCACAAGTGAACAACCTGGAGCTGGGAGGAAGC
CCCGGGGACCTTCAGACCTTGGCGTTGGAGGTGGCCCGGCAGAAGCGTGGCATTGTGGAT
CAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAAGGCCCA
CCTCGACCCGCCCCACCCCTCTGCAATGAATAAACTTTTGAATAAGCACCAAAAAAAG
AGTTCTATAATGAATGAAAAGGATTGTGTATATAGACATCTTTTTCTCTGGCATTATTT
GTCATGTTAGCATACTATTAAACCATTGTTAGGTTGGATGATTATATAATCATGTATGAA
GCTTGTGATAAAACACCAGGAATAATTCAAGTATCTGGAATTCTGCTTCCTGCCAAGAA
GGTAGGCAACCGTGAAATGCCACTGAAGCTACTAGTCTAAAAGTGAGTTATCTCTGTCT
TTGTCTTACCCCTGATGCTGTGATAAAACCTGACAAGAGCAACTGACTCCTGAGAGGA
AGGTTTATTCTAGCTCACAATTCAGGTTACAAACAGTCCATCCGTAGCAGGGGAGTCAC
AGCAACAGGAACCTCAGGGAAGTCTCCTATTATCCCCACAATCAAGAATAGTGACCAAT
AAATAAGTGGATCTTTTCTCAAAAAAAAAAAAAAAAAAAAAAAGACCAGAAACATTGGC
TACTCTTCAAGCATGAGTGATAGGCTGCTCATTGTCTAAGTTAATCCCATCTGTGGATT
GGGATATAACTTTTACTTGGTCTCTGGTCCATTTCAGTAATTTTCAAATGCAAAGTATTAT

```
GGTAAAATCCCCTCTCAGCAGTGGTGTAATGGCTGCTGTGCTGGGGAGCCTAGGCGAGGG
GGAAATGGATCAGGAAGATTAATCAGGAGCTACTTCCATTACCTAAGGCCGTTGTTTTCT
TAGACTTCTACATTTAGTGATCCAGAAAGCTAGTGGACATTTCAAGTAAAGCGTCATCTG
TCACTGGCAACAAATATTTATCAAACCCAGAATCATGTTGAAAGGGACACAAAAAACTC
CTTACTTCTTCACTGCTGAAAAATAGGTGAAAACAAGGCCAAACAAATTAACACAAAGGA
GATAATAAAAACTGTGAGAATGTAAAGAAAAAAAATTGCAAGTTGCCCCACATGCATTG
AAATTCCTTCCGATTGTGCCCTTATGTAGAATAGTAAAAGTAATATAAGAATCATATGGA
AACAGATCTGATGGTTTTATCCATTGTCTGCAATTTGACC
```

Σημειώνεται ότι η αλληλουχία DNA για το γονίδιο Insulin II του ποντικού δεν αναγράφεται στην αναφορά (λόγω του μεγάλου μήκους της), αλλά συμπεριλαμβάνεται σε μορφή .fasta στο παραδοτέο zip file.

4. Εύρεση θέσεων αναπαραγωγής του DNA για τη γονιδιακή θεραπεία μεταβολικών ασθενειών

4.3 Πρακτικό μέρος

Ζητούμενο 4.1:

Αρχικά, η δοθείσα ακολουθία αντιγράφηκε σε .txt αρχείο, πρωτού διαβασθεί από το script που γράφτηκε σε python. Για το εν λόγω ερώτημα, υλοποιήθηκε η συνάρτηση find_substring(string,sub) η οποία βρίσκει το πλήθος που εμφανίζεται η ακολουθία sub εντός της ακολουθίας string. Παρακάτω ακολουθεί το αποτέλεσμα που δίνει η κλήση της συνάρτησης για την αλληλουχία που δίνεται στην εκφώνηση και sub = 'atc'. Το αποτέλεσμα σημαίνει ότι η λέξη “atc” βρέθηκε 21 φορές στο τμήμα γονιδιώματος εισόδου.

```
#question 4.1
#open text file in read mode
text_file = open('sample_sequence.txt', "r")
#read whole file to a string
data = text_file.read()
#close file
text_file.close()
print(data)
def find_substring(string, sub):
    return int(string.count(sub))
find_substring(data,'atc')
```

Python

```
atcaatgatcaacgtaagcttctaagcatgatcaaggtgtcacacagtttatccacaacctgagtgatgacatcaagtaggtctgttatctcttctctgtactctcat
```

21

Ζητούμενο 4.2:

Για το ζητούμενο αυτό, υλοποιήθηκε σε python η συνάρτηση `most_frequent_substrings(string,k)` που λαμβάνει ως εισόδους της αλληλουχία ενδιαφέροντος(`string`) και το μήκος των συχνότερα εμφανιζόμενων λέξεων(`k`). Πιο αναλυτικά, η συνάρτηση ξεκινά από την αρχή του `string` εισόδου και αυξάνει κατά 1 κάθε νέα ακολουθία μήκους `k` (είτε αυτή είναι καινούργια είτε έχει ξαναβρεθεί). Το σύνολο των διακριτών `substrings` μήκους `k` που βρέθηκαν καθώς και οι φορές που εμφανίστηκαν στην αλληλουχία εισόδου αποθηκεύονται σε ένα `dictionary` το οποίο-αφού γίνει `sorted` κατά φθίνουσα σειρά-αποτελεί την έξοδο της ρουτίνας. Αποτελέσματα από την κλήση της `most_frequent_substrings(string,k)` φαίνονται στην κάτωθι εικόνα:

```
#question 4.2
from collections import defaultdict

def most_frequent_substrings(string, k):
    # Create a dictionary to store the substrings as keys and their frequency as values
    substring_counts = defaultdict(int)

    # Iterate through the string, taking substrings of length k and adding them to the dictionary
    for i in range(len(string) - k + 1):
        substring = string[i:i+k]
        substring_counts[substring] += 1

    # Find the most frequent substring by finding the maximum value in the dictionary
    most_frequent = max(substring_counts, key=substring_counts.get)
    sorted_dict = dict(sorted(substring_counts.items(), key=operator.itemgetter(1), reverse=True))

    return most_frequent, sorted_dict

a,b = most_frequent_substrings(data,9)
print(b)
```

Python

```
{'atgatcaag': 3, 'ctcttgatc': 3, 'tcttgatca': 3, 'cttgatcat': 3, 'aatgatcaa': 2, 'aagcatgat': 2, 'agcatgatc': 2, 'gcatgatca': 2, 'ttgatcatc': 2, 'tgatcatcg': 2, 'gctcttgat': 2, 'atcaatgat': 1, 'tcaatgatc': 1, 'caatgatca': 1, 'atgatcaac': 1, 'tgatcaacg': 1, 'gatcaacgt': 1, 'atcaacgta': 1, 'tcaacgtaa': 1, 'caacgtaag': 1, 'aagctaagc': 1, 'acgtaagct': 1, 'cgtaagctt': 1, 'gtaagcttc': 1, 'taagcttct': 1, 'aagcttcta': 1, 'agcttctaa': 1, 'gcttctaag': 1, 'cttctaagc': 1, 'ttctaagca': 1, 'tctaagcat': 1, 'ctaagcatg': 1, 'taagcatga': 1, 'catgatcaa': 1, 'tgatcaagg': 1, 'gatcaaggt': 1, 'atcaaggtg': 1, 'tcaaggtgc': 1, 'caaggtgct': 1, 'aaggtgctc': 1, 'aggtgctca': 1, 'ggtgctcac': 1,
```

5. Ανάλυση δεδομένων γονιδιακής έκφρασης από μικροσυστοιχίες DNA και πρακτική εξάσκηση στο πεδίο της Διατροφογενωμικής

5.6 Πρακτικό μέρος

Ζητούμενο 5.1:

Στο πείραμα με GEO accession id: GSE7117 μετρήθηκαν συνολικά 31 δείγματα από υπέρβαρες μεσήλικες γυναίκες. Τα δείγματα χωρίστηκαν σε 2 διαφορετικά groups: το intervention group τα 12 μέλη του οποίου τοποθετήθηκαν σε υποθερμιδική δίαιτα ΑΗΑ σταδίου I με προτεινόμενη ημερήσια πρόσληψη ενέργειας 5,0 MJ, με στόχο την απώλεια 0.5kg σωματικού βάρους την εβδομάδα. Από την άλλη, τα 19 μέλη του control group κλήθηκαν να συνεχίσουν την συνήθη δίαιτά τους και να μην χάσουν βάρος. Στην συνέχεια, επιλέχθηκαν δείγματα RNA 4 ατόμων από κάθε group. Η πλατφόρμα μικροσυστοιχίας DNA με την οποία έγιναν οι μετρήσεις ήταν η [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

Ζητούμενο 5.2:

- i) Ο πίνακας με τον τίτλο κάθε δείγματος και το id του ακολουθεί παρακάτω

Τίτλος	Sample_geo_accession id
Liver 1	GSM162954
Liver 2	GSM162956
Liver 3	GSM162957
Liver 9	GSM162958
Liver 12	GSM162959
Liver 25	GSM162960
Liver 26	GSM162961
MHLiv	GSM162962

- ii) Όλα τα δείγματα προέρχονται από το είδος οργανισμού *Homo sapiens*, όπως αναγράφεται στο domain “Sample_organism_ch1”.
- iii) Παρακάτω ακολουθεί ο πίνακας του ερωτήματος (i), μαζί με το group στο οποίο ανήκε κάθε subject

Τίτλος	Sample_geo_accession id	Group
Liver 1	GSM162954	Control
Liver 2	GSM162956	Interv.
Liver 3	GSM162957	Interv.
Liver 9	GSM162958	Control
Liver 12	GSM162959	Interv.
Liver 25	GSM162960	Control
Liver 26	GSM162961	Interv.
MHLiv	GSM162962	Control

- iv) Τα ορίσματα που υπάρχουν στην συνάρτηση `read.table()` είναι τα εξής:
 - `file.path("GSE7117_series_matrix.txt")` -> για το άνοιγμα του αρχείου με τα δεδομένα του πειράματος προς ανάλυση
 - `skip=68` -> για αγνόηση των πρώτων 68 γραμμών (ξεκινά η ανάγνωση του αρχείου από την γραμμή 69).
 - `header=TRUE` -> Η πρώτη γραμμή των δεδομένων περιλαμβάνει τα ονόματα των δεδομένων ανά στήλη.
 - `sep="\t"` -> για τον χωρισμό των στοιχείων όπου συναντώνται tab
 - `row.names=1` -> για την ανάγνωση των ονομάτων των γραμμών από την 1^η στήλη
 -
- v) Εκτελώντας τις εντολές `setws()`-με το τοπικό directory που απαιτούνταν-και `read.table()` με τα προαναφερθέντα ορίσματα λάβαμε την μεταβλητή `x`, η οποία αποτελεί έναν πίνακα δεδομένων με **54675 γραμμές και 8 στήλες**.

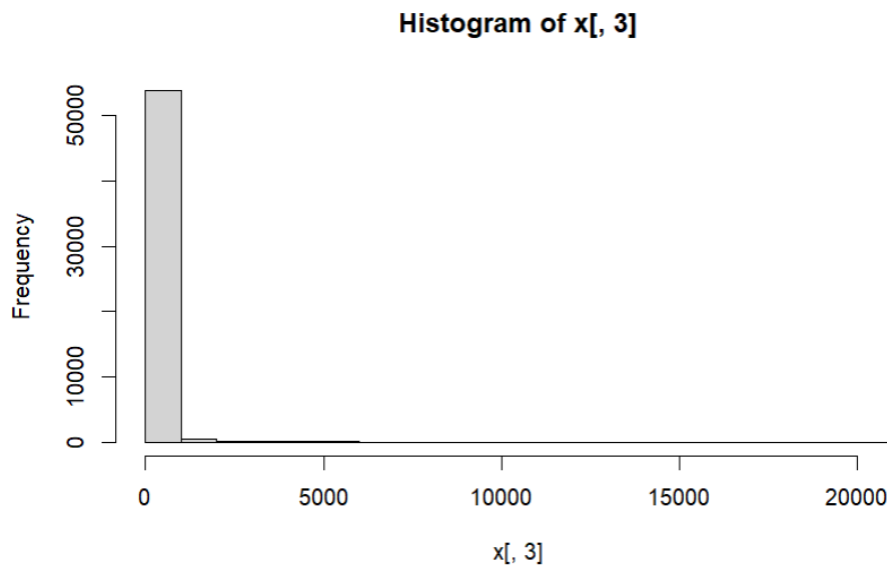
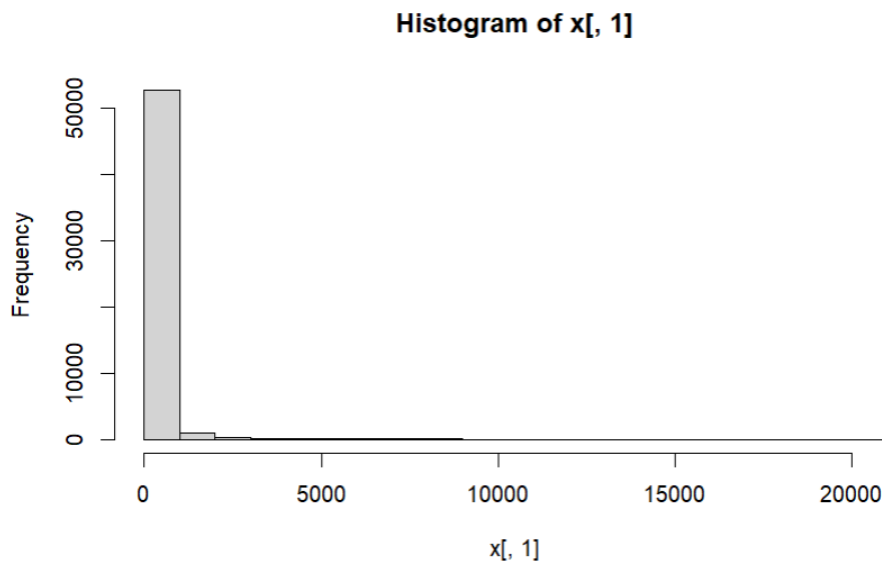
- vi) Μέσω της εντολής `colnames(x)` λάβαμε τα ονόματα των 8 στηλών τα οποία αντιστοιχούν στα διαφορετικά ids των 8 subjects από τα οποία ελήφθησαν τα RNA samples. Τα αποτελέσματα από την κλήση της εντολής φαίνονται παρακάτω:

```
> colnames(x)
[1] "GSM162954" "GSM162956" "GSM162957" "GSM162958" "GSM162959" "GSM162960" "GSM162961"
[8] "GSM162962"
```

- vii) Μέσω της εντολής `rownames(x)[1:15]` λάβαμε τα ονόματα των 15 πρώτων γραμμών του πίνακα δεδομένων, που αντιστοιχούν στα διαφορετικά γονίδια των οποίων την συγκέντρωση επιθυμούμε να μελετήσουμε σε καθένα από τα 8 δείγματα. Τα αποτελέσματα από την κλήση της εντολής φαίνονται παρακάτω:

```
> rownames(x)[1:15]
[1] "1007_s_at" "1053_at" "117_at" "121_at" "1255_g_at" "1294_at"
[7] "1316_at" "1320_at" "1405_i_at" "1431_at" "1438_at" "1487_at"
[13] "1494_f_at" "1552256_a_at" "1552257_a_at"
```

- viii) Από το command window του Rstudio, λάβαμε $x[3,5] = 42.91798$, που αντιστοιχεί στην μέτρηση συγκέντρωσης του 3^{ου} γονιδίου-δείκτη ενδιαφέροντος στο 5^ο δείγμα.
- ix) Η 200^η γραμμή αναφέρεται στις συγκεντρώσεις του 200^{ου} γονιδίου-δείκτη σε καθένα από τα 8 υποκείμενα προς εξέταση. Από την άλλη, η 7^η στήλη μας δίνει τις συγκεντρώσεις όλων των γονιδίων-δειγμάτων για το 7^ο υποκείμενο προς εξέταση.
- x) Μέσω της εντολής `hist(x[,i])` εξήχθησαν τα ιστογράμματα συχνοτήτων για το 1^ο ($i=1$) και για το 3^ο δείγμα ($i=3$), με τα αποτελέσματα να αποτυπώνονται παρακάτω:



Ζητούμενο 5.3:

- i) Όσον αφορά τα ορίσματα της εκτελούμενης εντολής `round()`, αυτά είναι τα εξής:
- `apply(x,2,summary)` -> για την λήψη των συνοπτικών (“summary”) στατιστικών στοιχείων επί των στηλών του πίνακα δεδομένων x. Σε

περίπτωση που επιθυμούσαμε summary στις γραμμές, θα έπρεπε να εκτελεστεί η εντολή `apply(x,1,summary)`.

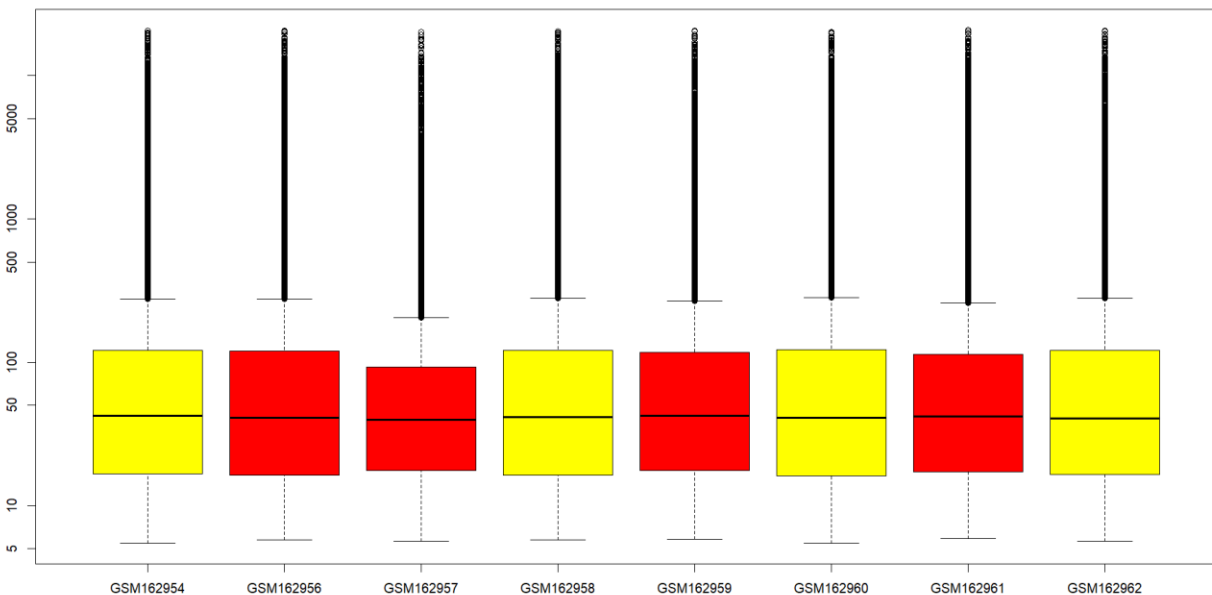
- `digits=2` -> για αποτύπωση των στατιστικών αποτελεσμάτων με ακρίβεια 2 δεκαδικών ψηφίων.

Τα αποτελέσματα αναγράφονται παρακάτω:

```
> round(apply(x,2, summary),digits=2) # explain in your assignment the functionality/arguments in this line of code
```

	GSM162954	GSM162956	GSM162957	GSM162958	GSM162959	GSM162960	GSM162961	GSM162962
Min.	5.44	5.74	5.65	5.76	5.81	5.48	5.90	5.64
1st Qu.	16.63	16.33	17.55	16.24	17.54	16.18	17.14	16.49
Median	42.31	41.06	39.89	41.26	42.33	40.88	41.93	40.56
Mean	213.47	217.89	129.67	213.59	207.06	216.42	197.97	209.00
3rd Qu.	120.77	119.92	92.73	121.73	117.34	122.76	113.54	121.52
Max.	20596.82	20710.01	20130.98	20276.33	20518.67	20142.23	20892.18	20513.43

- ii) Μέσω κλήσης της εντολής `boxplot()` έγινε δημιουργία των ζητούμενων θηκογραμμάτων για καθένα από τα 8 δείγματα, με την σχετική γραφική να ακολουθεί παρακάτω:



Με κίτρινο έχουν χρωματιστεί τα θηκογράμματα για τα control samples και με κόκκινο τα θηκογράμματα για τα test samples. Παρατηρούμε ότι οι αποκλίσεις στις τιμές ανά δείγμα είναι σχετικά μικρές, γεγονός που επαληθεύει το γεγονός ότι έχει προηγηθεί κανονικοποίησή τους.

Ζητούμενο 5.4:

- i) Εφόσον επιθυμούμε τα δείγματα ελέγχου να έχουν τιμή 0 και τα δείγματα με διαιτητική παρέμβαση να έχουν τιμή 1, έχουμε $c=(0,1,1,0,1,0,1,0)$, τιμή την οποία λαμβάνει και η `xsampleslabels`
- ii) Ο κώδικας για την υλοποίηση του εν λόγω βήματος περιλαμβάνει την απομόνωση-για κάθε ανιχνευτή-της τιμής του (4 τιμές για το control group και άλλες 4 για το interv.group). Στην συνέχεια, τα 2 groups διαχωρίστηκαν σε 2 arrays (`c_control` και `c_interv`) τα οποία αποτελούν είσοδο στην συνάρτηση `t.test()` της R. Στον κάτωθι πίνακα ακολουθούν τα p-values για τους 15 πρώτους ανιχνευτές

Ανιχνευτής	p-value
1007_s_at	0.1235861
1053_at	0.2768388
117_at	0.6168126
121_at	0.1692107
1255_g_at	0.1741585
1294_at	0.5607194
1316_at	0.8137377
1320_at	0.2075931
1405_i_at	0.395651
1431_at	0.5966297
1438_at	0.5704014
1487_at	0.5788862
1484_f_at	0.7940872
1552256_a_at	0.7325208
1552257_a_at	0.9863738

- iii) Για την εύρεση των ανιχνευτών με $p_values < 0.001$ έγινε χρήση της εντολής `which()` για τον εντοπισμό των indices που πληρούν την επιθυμητή συνθήκη για τα p-values τους. Στην συνέχεια, για κάθε

index, αποθηκεύεται τόσο η τιμή του p_value όσο και το όνομα του ανιχνευτή. Τέλος, indices, names και p_values αποθηκεύονται σε ένα συνολικό dataframe, το οποίο παρατίθεται παρακάτω:

	sign_idx	sign_name	p_values[p_values < 0.001]
1	467	1552912_a_at	5.930648e-04
2	485	1552935_at	9.469452e-04
3	694	1553222_at	9.627742e-04
4	2471	1555645_at	1.380162e-04
5	2754	1556049_at	6.507395e-04
6	6441	1562173_a_at	1.987993e-04
7	10380	200931_s_at	8.925807e-04
8	13467	204019_s_at	6.827508e-05
9	13562	204114_at	3.844311e-04
10	14129	204681_s_at	3.303579e-04
11	14309	204861_s_at	4.276867e-04
12	14619	205171_at	5.160510e-04
13	18214	208798_x_at	9.010826e-04
14	21722	212416_at	8.487104e-04
15	24295	214999_s_at	6.587829e-05
16	24757	215463_at	1.335609e-04
17	25164	215870_s_at	3.812513e-04
18	30509	221225_at	4.155650e-04
19	33838	224578_at	2.624828e-04
20	34432	225174_at	8.433338e-04
21	36562	227306_at	9.672951e-04
22	36917	227662_at	2.642549e-04
23	38279	229024_at	6.726973e-04
24	40388	231133_at	6.688893e-04
25	41859	232604_at	2.522727e-04
26	42184	232930_at	3.920250e-04
27	44001	234751_s_at	4.826675e-04

28	44080	234830_at	7.787919e-04
29	44707	235457_at	7.713711e-04
30	44926	235676_at	7.537940e-04
31	45502	236252_at	2.810705e-04
32	48618	239368_at	8.203637e-04
33	48798	239548_at	4.264592e-04
34	49427	240177_at	2.759772e-04
35	52587	243337_at	2.685686e-05
36	52904	243654_at	9.011096e-04
37	53044	243794_at	5.802618e-04
38	54239	36129_at	2.664764e-04

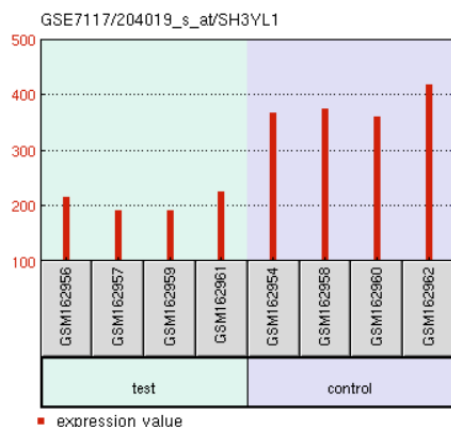
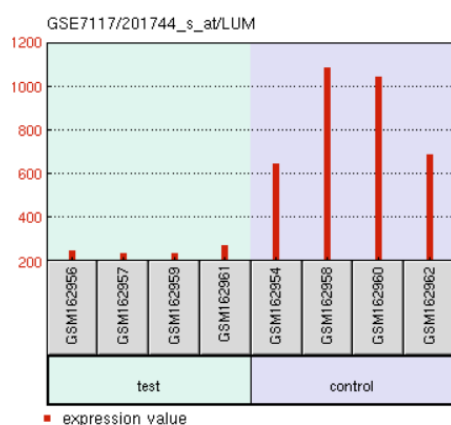
Για την εύρεση του ανιχνευτή με την χαμηλότερη τιμή, p_value έγινε χρήση της εντολής `which.min()` για τον προσδιορισμό του index, ενώ με παρόμοιο τρόπο με προηγουμένως έγινε η ενοποίηση ονόματος, index και τιμής σε ένα dataframe διάστασης 1x3:

	<code>which.min(p_values)</code>	<code>rownames(x)[(min_idx[1, 1])]</code>	<code>p_values[(min_idx[1, 1])]</code>
1	52587	243337_at	2.685686e-05

Ζητούμενο 5.5:

Αφού έγινε διαχωρισμός των δειγμάτων σε control group και σε diet intervention group (μέσω της εντολής “Define groups”), πραγματοποιήθηκε ανάλυση διαφορικής έκφρασης για τον εντοπισμό των 250 πιο σημαντικών γονιδίων.

- i) Τα αποτελέσματα της ανάλυσης για τα 2 πιο σημαντικά γονίδια -με 1^ο τον ανιχνευτή 201744_s_at και 2^ο τον 204019_s_at - παρατίθενται παρακάτω:



ID	adj.P.Val	P.Value	t	B	logFC
201744_s_at	0.4	0.00000731	-9.4	0.8482	-1.751
204019_s_at	0.544	0.00002013	-8.27	0.5363	-0.883

Αρχικά, παρατηρούμε ότι πράγματι και οι 2 ανιχνευτές έχουν στατιστική σημαντικότητα, μιας και έχουν $p_values \ll 0.001$, γεγονός που επιτρέπει την απόρριψη της μηδενικής υπόθεσης, μιας και η πιθανότητά της να ισχύει είναι χαμηλότερη από το καθορισμένο threshold.

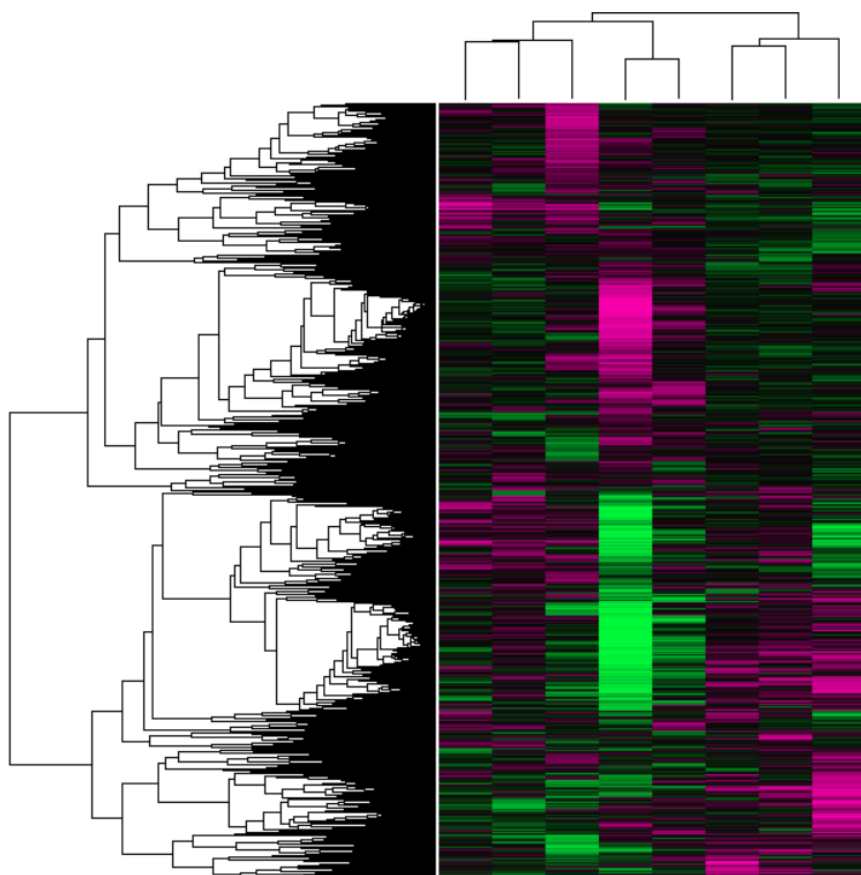
Όσον αφορά τα logFC values, παρατηρούμε ότι $\log FC(\text{test/control}) < 0$ γεγονός που συνεπάγεται ότι $(\text{test/control}) < 1$ ή $\text{test}_i < \text{control}_i$ $i \in \{1,2\}$, κάτι το οποίο επαληθεύεται και από τα αποτελέσματα της διαφορικής ανάλυσης παραπάνω, όπου τα δείγματα στο test group έχουν αισθητά χαμηλότερη έκφραση των ανιχνευτών 201744_s_at και 204019_s_at σε σχέση με τα δείγματα του control group. Συνεπώς, οδηγούμαστε στο συμπέρασμα ότι **η έκφραση των εν λόγω γονιδίων καταστέλλεται**.

- ii) Το γονίδιο με τον στατιστικά σημαντικότερο βαθμό διαφορικής έκφρασης είναι το 201744_s_at ή αλλιώς *luminar*. Το εν λόγω γονίδιο βρίσκεται στο 12^ο χρωμόσωμα (περιοχή q21.33) και

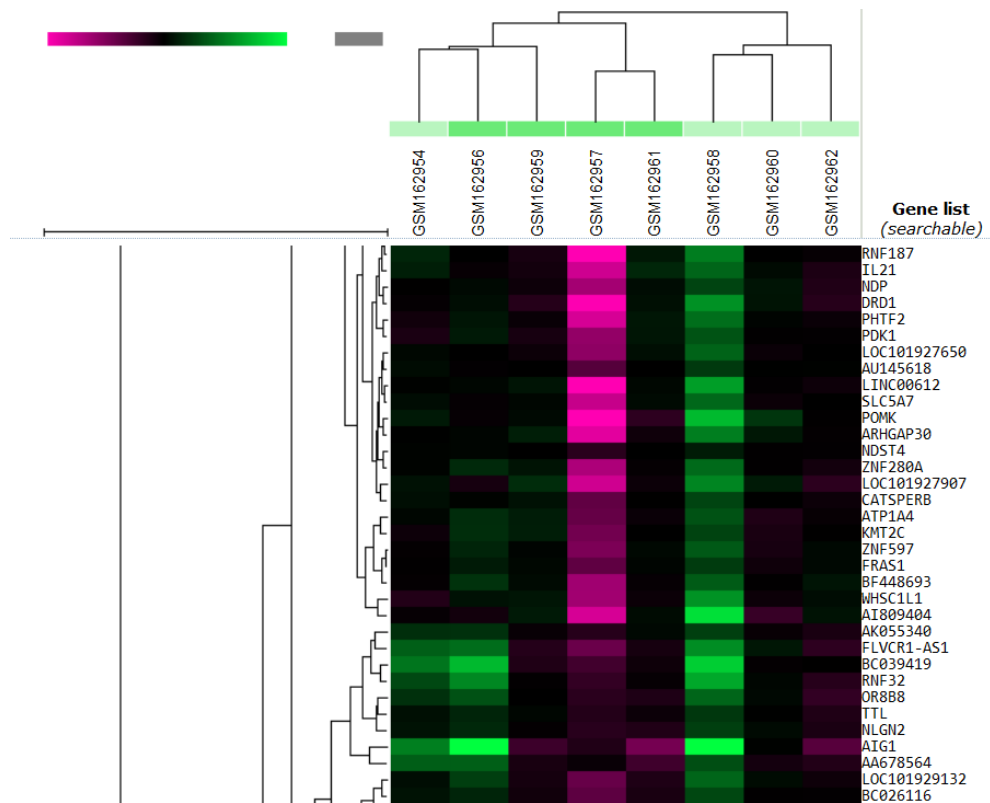
κωδικοποιεί ένα μέλος της μικρής οικογένειας πρωτεογλυκνών πλούσια σε λευκίνη (SLRP) που περιλαμβάνει ντεκορίνη, διγλυκάνη, ινομοντουλίνη, κερατοκάνη, επιφυκάνη και οστεογλυκίνη. Σε αυτά τα διλειτουργικά μόρια, το τμήμα πρωτεΐνης δεσμεύει τα ινίδια κολλαγόνου και οι υψηλά φορτισμένες υδρόφιλες γλυκοζαμινογλυκάνες ρυθμίζουν τις αποστάσεις μεταξύ των ινιδίων. Το Lumican είναι η κύρια πρωτεογλυκάνη θεικής κερατάνης του κερατοειδούς, αλλά κατανέμεται επίσης σε διάμεσες κολλαγόνες μήτρες σε όλο το σώμα. Το Lumican μπορεί να ρυθμίζει την οργάνωση και την περιφερειακή ανάπτυξη των ινιδίων κολλαγόνου, τη διαφάνεια του κερατοειδούς και τη μετανάστευση των επιθηλιακών κυττάρων και την επιδιόρθωση ιστών.

Ζητούμενο 5.6:

Έπειτα από αναζήτηση του πειράματος GSE7117 στον Dataset Browser του NCBI και επιλογή “Cluster Analysis”, έγινε εξαγωγή του πλήρους θερμικού χάρτη ο οποίου παρατίθεται παρακάτω:



- i) Ο άνωθεν heatmap έχει 54675 γραμμές και 8 στήλες.
- ii) Με την επιλογή “Stack up”, πραγματοποιείται απεικόνιση των λεπτομερειών του θερμικού χάρτη:



Όσον αφορά τις διαστάσεις του, ο χάρτης έχει 54675 γραμμές, που αντιστοιχούν στους διαφορετικούς ανιχνευτές προς διερεύνηση, και 8 στήλες, που αντιστοιχούν στα διαφορετικά δείγματα. Επιπλέον, η θέση των διαφορετικών γονιδίων στις γραμμές είναι τέτοια, ώστε γονίδια με παρεμφερή επίπεδα έκφρασης για τουλάχιστον ένα δείγμα να βρίσκονται σε κοντινές γραμμές.

- iii) Η χρωματική ποσοτικοποίηση της γονιδιακής έκφρασης λειτουργεί ως εξής: γονίδια με υψηλά ποσοστά έκφρασης (σε σχέση με τον μέσο όρο) έχουν μωβ χρώμα, ενώ γονίδια με χαμηλά ποσοστά αντιστοιχούν σε πράσινο χρώμα. Το πρώτο γονίδιο στην λίστα είναι το RNF187 και η έκφρασή του στο 1^ο δείγμα (GSM162954) έχει πολύ αχνό πράσινο χρώμα. Αυτό συνεπάγεται ότι έχει **μέση προς χαμηλή τιμή** έκφρασης στο εν λόγω δείγμα.

- iv) Μια πρώτη συσχέτιση αφορά την τοποθέτηση των δειγμάτων κατά στήλες. Συγκεκριμένα, τα δείγματα testing έχουν τοποθετηθεί στις στήλες 2-5 και τα δείγματα ελέγχου στις στήλες 1 και 6-8. Στην συνέχεια, παρατηρούμε ότι για τα πρώτα 10-15 γονίδια τα περισσότερα δείγματα έχουν σχετικά ενδιάμεσες τιμές έκφρασης, με εξαίρεση το δείγμα GSM162957 του test group - το οποίο παρουσιάζει υψηλές τιμές έκφρασης - και το δείγμα GSM162958 του control group, το οποίο παρουσιάζει χαμηλές τιμές έκφρασης στα προαναφερθέντα γονίδια. Εξαίρεση στα παραπάνω αποτελεί το γονίδιο NDST4 του οποίου οι τιμές έκφρασης παραμένουν σε ενδιάμεσα επίπεδα για όλα τα δείγματα. Τέλος, αξίζει να αναφερθεί ότι η ομαδοποίηση και συσχέτιση των γονιδίων σε διαδοχικές γραμμές ανάλογα με τα επίπεδα έκφρασης στα 8 δείγματα πρέπει να επαληθευτεί και μέσω αναζήτησης θεωρητικών στοιχείων και πληροφοριών για αυτά (π.χ. εύρεση πλήθους βάσεων αλληλουχίας, χρωμοσώματος και περιοχής, σύγκριση των εκφραζόμενων πρωτεϊνών και της λειτουργίας του κ.α.)