GBIO0002

Homework 2

Genetics and bioinformatics

Important Points:

- Prepare reports for HW-2 part1. Reports corresponding to literature based homework's will be considered for fine-tuning the marks based on the in vivo presentations (slides).

- No report is needed for the Q&A part2 in HMW2.

Important dates:

- Submit report and presentation files before 18 December 2017, 18:00h (HW2-part1)

- The presentation will be on 19 December 2017

Marks:

- Part 1 – 15 points
- Part 2 – 15 points

Evaluation:

- For report, your work will be evaluated based on the accuracy and completeness of answers and report structure.
- For presentation, you will be evaluated based on the completeness of slides, your presentation, your understanding, and your answers to questions.

Instruction:

- Form a group of 2-3 persons and complete the homework in both parts (1&2).
- For presentation, you have 15 minutes to present your work for each part and everybody in a group needs to present.

Part 1:

A: Select ONE of these papers to study and create a presentation. Your presentation should cover objective, method, result (if available), literatures (if needed), your own discussion, and your own conclusion.

- Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev. 2007 May 1;21(9):1010–24.
- Wolf JBW. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Mol Ecol Resour. 2013 Jul 1;13(4):559–72.

- Daakour, Sarah, Hajingabo, Leon Juvenal, Kerselidou, Despoina, Devresse, Aurelie, Kettmann, Richard,Simonis, Nicolas,Dequiedt, Franck,Twizere, Jean-Claude, Systematic interactome mapping of acute lymphoblastic leukemia cancer gene products reveals EXT-1 tumor suppressor as a Notch1

and FBWX7 common interactor. BMC Cancer 2016, 16(1): 335,doi10.1186/s12885-016-2374-2

Dec-2017 ULg

Part 2 :

- **Submit answers electronically by <u>1 January 2018, 18:00h</u>**

Answer the following questions :

(1) (a). Predict the chromosomal details (chromosome number, start position and end position) of the following gene list (separated by commas)

LIN28B,HNF1A,UGT1A6,APOC3,BDNF,WWOX,SLC22A5,UGT1A9,APOC1,GDF5,GCKR,BMP2,FADS1,MICA,AHR,ESR1,PCSK9,NPPC,IL10,ACADM,COMT,PPID,GNA12,CPS1,IFNG,CXCR4,KLKB1,ALPL,PAPPA,UGT1A7,IKZF1,PON1,CCL2,HHIP,SH2B3,HLA-A,NPR3,PLAG1,F12,CXCL12

(b) Perform functional analysis of the given gene list using the Enrichr web server and predict "enriched pathways" as well as draw the corresponding network. Draw conclusions based on the predicted network such as "which pathway is highly enriched" and "how it relates to human biological mechanisms".

(c) Perform Ontology Analysis using the Enrichr webserver and predict GO Biological Process 2017b, GO Molecular Function 2017, GO Cellular Component 2017b. Provide the tabular output obtained. Also check genes involved in the "significant enriched (FDR crossed) GO terms".

2. (I) Perform BLAST of following sequence using the NCBI blastn server against NR (Non-redundant) database. Based on BLAST results, estimate to which species these sequences belongs. Copy the BLAST output on which your assertion is based.
(a) Human
(b) Mouse
(c) none of the above

>seq1
CGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCCAGACTGCCTTCCGGG
TCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCA
GGAAACATTTTCAGACCTATGGAAACTACTTCCTGAAAACAACGTTCTGTCCCCC
TTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAAT
GGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTC
CCCGCGTGGCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCCTGCACCAGCCCC
CTCCTGGCCCCTGTCATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTAC
GGTTTCCGTCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACGT
ACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCA
GCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATC
TACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCCACCATGAG
CGCTGCTCAGATAGCGATGGTCTGGCCCCTCCTCAGCATCTTATCCGAGTGGAAG
GAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCGACATAGTGTGGT
GGTGCCCTATGAGCCGCCTGAGGTTGGCTCTGACTGTACCACCATCCACTACAAC
TACATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCA
TCATCACACTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGG

TGCATGTTTGTGCCTGTCCTGGGAGAGACCGGCGCACAGAGGAAGAGAATCTCC
GCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCCAGGGAGCACTAAGCGAGCA
CTGTCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTGGATGGA
GAATATTTCACCCTTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGC
TGAATGAGGCCTTGGAACTCAAGGATGCCCAGGCTGGGAAGGAGCCAGGGGGG
AGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGC
CATAAAAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGACATTCTCCA
CTTCTTGTTCCCCACTGACAGCCTCCCACCCCCATCTCTCCCTCCCCTGCCATTTT
GGGTTTTGGGTCTTTGAACCCTTGCTTGCAATAGGTGTGCGTCAGAAGCACCCAG
GACTTCCATTTGCTTTGTCCCGGGGCTCCACTGAACAAGTTGGCCTGCACTGGTG
TTTTGTTGTGGGGAGGAGGATGGGGAGTAGGACATACCAGCTTAGATTTTAAGGT
TTTTACTGTGAGGGATGTTTGGGAGATGTAAGAAATGTTCTTGCAGTTAAGGGTT
AGTTTACAATCAGCCACATTCTAGGTAGGGGCCCACTTCACCGTACTAACCAGGG
AAGCTGTCCCTCACTGTTGAATTTTCTCTAACTTCAAGGCCCATATCTGTGAAATG
CTGGCATTTGCACCTACCTCACAGAGTGCATTGTGAGGGTTAATGAAATAATGTA
CATCTGGCCTTGAAACCACCTTTTATTACATGGGGTCTAGAACTTGACCCCCTTG
AGGGTGCTTGTTCCCTCTCCCTGTTGGTCGGTGGGTTGGTAGTTTCTACAGTTGGG
CAGCTGGTTAGGTAGAGGGAGTTGTCAAGTCTCTGCTGGCCCAGCCAAACCCTGT
CTGACAACCTCTGGTGAACCTTAGTACCTAAAAGGAAATCTCACCCCATCCCACA
CCCTGGAGGATTTCATCTCTTGTATATGATGATCTGGATCCACCAAGACTTGTTTT
ATGCTCAGGGTCAATTTCTTTTTTCTTTTTTTTTTTTTTTTCTTTTTCTTTGAGACT
GGGTCTCGCTTTGTTGCCCAGGCTGGAGTGGAGTGGCGTGATCTTGGCTTACTGC
AGCCTTTGCCTCCCCGGCTCGAGCAGTCCTGCCTCAGCCTCCGGAGTAGCTGGGA
CCACAGGTTCATGCCACCATGGCCAGCCAACTTTTGCATGTTTTGTAGAGATGGG
GTCTCACAGTGTTGCCCAGGCTGGTCTCAAACTCCTGGGCTCAGGCGATCCACCT
GTCTCAGCCTCCCAGAGTGCTGGGATTACAATTGTGAGCCACCACGTCCAGCTGG
AAGGGTCAACATCTTTTACATTCTGCAAGCACATCTGCATTTTCACCCCACCCTTC
CCCTCCTTCTCCCTTTTTATATCCCATTTTTATATCGATCTCTTATTTTACAATAAA
ACTTTGCTGCCAAAAAAAAAAAAAAAAAA

>seq2
GACATCGTGATGACCCAGTCTCCTTCCACCCTGTCTGCATCTGTAGGAGACAGAG
TCACCATCACTTGCCGGGCCAGTCAGAGTATTAGTAGCTGGTTGGCCTGGTATCA
GCAGAAACCAGGGAAAGCCCCTAAGTTCCTGATCTATAAGGCGTCTAGTTTAGA
AAGTGGGGTCCCATCAAGGTTCAGCGGCAGTGGATCTGGGACAGAATTCACTCT
CACCATCAGCAGCCTGCAGCCTGATGATTTTGCAACTTATTACTGCCAACAGTAT
AATACGTTCGGCCAAGGGACCAAGCTGGAGATCAAAC

>seq3
CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAACCCTGGGGCCTCAGTG
AAGGTCTCCTGCAAGGCTTCTGGATACACCTTCACCGGCTACTATATGCACTGGG
TGCGACAGGCCCCTGGACAAGGACTTGAGTGGATGGGCTGGATCAACCCTAACA
GTGGTGGCACAAACTATGCACAGAAGTTTCGGACTGGGTCACCATGACCAGGGA
CACGTCCATCAGCACAGCCTATATGGAGCTGAGCAGGCTGAGATCTGACGACAC
GGCCGTTTATTACTGTGCGAGAAATATAGCAACAACTGGTGATGCTTTTGATATT
GGGGCCAAGGGACAATGGTCACCGTCTCCTCAG