Isoform Switch Analysis Exercise

Part 1 : An introduction

Connet to server

- Log into the ricco server via a terminal
- ssh -X username@ricco.popgen.dk
    - REMEMBER THE “-X” !!!
      (else you cannot make plots)
- Your username is your KU ID (ex: abc123)
  password: bohta17

Open R and check -X works

- Open R by copy/pasting:
  nice /home/bohta/.linuxbrew/bin/R
- make sure it is R version 3.4
- Test plotting:
  Type: hist(rnorm(10))
- A new window should open with a histogram. If it does not:
  - Windows: use http://mobaxterm.mobatek.net/
  - Mac: Install https://www.xquartz.org/
  - Work with your neighbour

Overal Exercise Idea

- Aim:
  Get familiar with the overall workflow of IsoformSwitchAnalyzeR
- Approach:
  Use IsoformSwitchAnalyzeR and answer questions along the way.

Assignment

Individually read/work (meaning run the R code and check output) through the following sections of the vignette:

- “Abstract”
- “Workflow Overview”
- “Short Example Workflow”

And answer the questions in the rests of the slides. Remember to look up the documentation of the functions (run ?functionName) as all the details are there.

Open vignette

http://people.binf.ku.dk/rtl144/IsoformSwitchAnalyzeR.html

Questions

- Q1: What is the core functionalities of IsoformSwitchAnalyzeR?

advanced post analysis of full length, RNA-seq derived transcripts, with a focus on finding, annotating and visualizing isoform switches with functional consequences. IsoformSwitchAnalyzeR therefore performs 3 specific tasks:

- Identifying isoform switches
  Annotate the transcripts involved in the isoform switches
  Visualize the consequences of the isoform switches, both individually and combined.

- Q2: What external sequence analysis tools are currently supported

CPAT : The Coding-Potential Assessment Tool, which can be run either locally or via their webserver.

- Pfam : Prediction of protein domains, which can be run either locally or via their webserver.
  SignalP : Prediction of Signal Peptides, which can be run either locally or via their webserver.

- Q3: Why is it smart to identify isoform switches as the first step in a workflow? (hint look at the documentation of isoformSwitchAnalysisPart1())

The importance of analyzing isoforms instead of genes has been highlighted by many examples showing functionally important changes that cannot be detected at gene level. Since we know the exon structure of the full-length isoform,

- IsoformSwitchAnalyzeR can extract the underlying nucleotide sequence from a reference genome. This enables integration with the Coding Potential Assessment Tool (CPAT) - which predicts the coding potential of an isoform and can also be used to increase accuracy of ORF predictions. By combining the CDS/ORF isoform positions with the nucleotide sequence we can also extract the (most likely) amino acid (AA) sequence of the CDS/ORF. The AA sequence enables integration of analysis of protein domains (via Pfam) and signal peptides (via SignalP) -

Questions

- Q4: How many high-level functions (functions that automatically performs multiple step of the pipeline) are there in IsoformSwitchAnalyzeR?

- isoformSwitchAnalysisPart1(): Extract Isoform Switches and Their Sequences
  isoformSwitchAnalysisPart2(): Plot All Isoform Switches and Their annotation

- Q5: What is a BSgenome object and why is it needed

- Infrastructure for Biostrings-based genome data packages and support for efficient SNP representation
  A BSgenome data package contains the full genome sequences for a given organism.
  needs the genomic sequence to predict ORFs.

Questions

- Q6: What is the result if you change the dIFcutoff parameter to 0.5 in the isoformSwitchAnalysisPart1()
-
- Q7: What is the main functionality of isoformSwitchAnalysisPart1() vs isoformSwitchAnalysisPart2()

  The dIFcutoff argument indicating the (absolute) change in isoform usage (dIF) a change must be larger than.
-

Questions

- Q8: How many switches with consequences did you identify (when using a dIF cutoff as 0.5):
- The number of isoform swithces found were:

	Comparison	nrIsoforms	nrGenes
1	hESC vs Fibroblasts	6	4

- Q9: What is the consequence of the switch in the “LDLRAD2” gene? (identify the predicted functional consequence)
-

Questions

- Q10: Why is a cutoff on both the q-value (alpha) and the dIF necessary?
-
- Q11*: Re-load the example data by copy pasting the following: data(“exampleSwitchListAnalyzed”). Use this example data to identify which consequence is the one with the most certain (aka smallest p-value) global changes in isoform usage? Hint: see extractGenomeWideAnalysis()
-

Isoform Switch Analysis Exercise

Part 2: Improved understanding

Open vignette

http://people.binf.ku.dk/rtl144/IsoformSwitchAnalyzeR.html

Overal Exercise Idea

- Aim:
  To dig into the details of an IsoformSwitchAnalyzeR workflow
- Approach:
  Use IsoformSwitchAnalyzeR and answer questions along the way.

Assignment

Thoroughly read/work (meaning run the R code and check output) through the following sections of the vignette:

- “Detailed Workflow”

And answer the questions in the rests of the slides.

Remember to look up the function documentation for all functions you use as all the details are there.

Questions

- Q1: How many functions are isoformSwitchAnalysisPart1() and isoformSwitchAnalysisPart2() internally using?
- Q2: What is the structure of the switchAnalyzeRlist object?
-
- Q3: What is the name of the central entry in the switchAnalyzeRlist object, how is it accessed and what does each row correspond to?
-

Questions

- Q4: What does all the analyzeXXX and extractXXX functions do?
-
- Q5: Which functions are used for importing RNA-seq quantification data and creating the switchAnalyzeRlist?
-

Questions

- Q6: Why is the summary statistics of these two commands not different:
  e1 <- preFilter(exampleSwitchListAnalyzed, keepIsoformInAllConditions = TRUE)
  e2 <- preFilter(exampleSwitchListAnalyzed, keepIsoformInAllConditions = FALSE)

  When the resulting switchAnalyzeRlists are different:
  > nrow(e1) == nrow(e2)
  [1] FALSE

-

- Q7.1: How many different ways of testing for isoform switches are supported by IsoformSwitchAnalyzeR and which is the (current) recommended?

-

- Q7.2: Why is it smart to identify isoform switches as the first step in a workflow? (hint look at the documentation of the functions for isoform switch testing)

-

- Q8*: Which algorithms for identifying ORFs are supported? Which is the default and what do you think of that?

Questions

- Q9: Which consequences are affected by the ‘ntCutoff’ and the ‘AaJCsimCutoff’ partner in the analysis of switch consequences, and what does the cutoffs do.
-
- Q10: How many different types of consequences can be predicted from the domain annotation and why is there a need for all of them?
-
- Q11: How do you access the details of the consequence analysis and which is the first two isoforms compared in the example?
-

Questions

- Q12*: Which gene (with consequences) is on the 4th place on the top switching list when sorting on respectively q-values and dIF values? Comment on the gene name.
-
- Q13: Which isoform (with consequences) is on the 5th place on the top switching list when sorting on respectively q-values and dIF values?
-
- Q14: What is the consequence of the switch in the gene associated with TCONS_00004208?
-
- Q15: What is the relation between isoform switches and gene log2 FC?
-

Questions

- Q16*: What is the relation between isoform switches and gene expression
-