NBIC: Support Project Application Galaxy RNASeq platform

> Contact Guido Jenster

> E-mail address [g.jenster@erasmusmc.nl](mailto:g.jenster@erasmusmc.nl)

> Phone number +31 10 704 3672

> NGI Centre

> Choose one of the following answers

> No NGI Centre, Cancer Genomics Centre; Celiac Disease Consortium; Centre for BioSystems Genomics; Centre for Medical Systems Biology; Ecogenomics Consortium; Forensic Genomics Consortium Netherlands; Kluyver Centre for Genomics of Industrial; Netherlands Consortium for Healthy Ageing; Netherlands Consortium for Systems Biology; Netherlands Metabolomics Centre; Netherlands Proteomics Centre; Netherlands Toxicogenomics Centre; Nutrigenomics Consortium; VIRGO Consortium

>Other affiliation (E.g. CMBI, TUDelft.) Erasmus MC

Project description

Title: Galaxy RNASeq platform

> Please describe in some detail what the goal of the project is. What is the current situation and what is the desired outcome?

Provide a list of desired deliverables.

The aim is to provide CTMM (PCMM) and other scientists with a simple analytical workflow in Galaxy to determine gene expression of spliced genes and fusion genes from RNASeq data.

The “state of the art” applications for RNASeq analysis include tophat/Cufflinks and oQtans. Tophat/Cufflinks are tools with Galaxy and we have imported oQtans into the same Galaxy instance @ Erasmus but there is no validated workflow for either method. In addition we have python code to compare results from the two methods and wish to determine a comparison between the methods with test data. We will process the CTMM data using this validated workflow

**Proposed Deliverables**

1. Galaxy Workflow for RNAseq @ Erasmus
   1. for two (multi-) class gene expression comparison
   2. for gene splicing and fusion gene detection
2. Application documentation
3. Method evaluation report of Case Study
4. User manual with Case Study

>Please describe why you would want support on this project? (E.g. missing expertise at our institution, use of high-performance computing resources or other infrastructural needs.)

The student working on this project is now continuing their studies into bioinformatics so we have no programmers to complete this project.

Workflow validation is paramount to this project and we are lacking software engineers with those skills on this project.

> Are you planning on publishing about this project?

Choose one of the following answers Yes, No, Maybe

Yes, as an addition to prototyping already completed in our department on RNAseq analysis by our student. The work will be a joint NBIC/Erasmus publication

> Any other remarks related to the project.

This work is part of the CTMM/NBIC TraIT WP4-NGS project

We will provide expertise on the test case, use cases and application development

We will migrate this workflow to NBIC/CTMM Galaxy for public use

We will use this validated workflow to analyse our CTMM data

NBIC: Support Project Application R-package for Structural Variation Visualization

> Contact Guido Jenster and Andrew Stubbs

> E-mail address [g.jenster@erasmusmc.nl](mailto:g.jenster@erasmusmc.nl); a.stubbs@erasmusmc.nl

> Phone number +31 10 704 3672; +31-10-704 4776

> NGI Centre

> Choose one of the following answers

> No NGI Centre, Cancer Genomics Centre; Celiac Disease Consortium; Centre for BioSystems Genomics; Centre for Medical Systems Biology; Ecogenomics Consortium; Forensic Genomics Consortium Netherlands; Kluyver Centre for Genomics of Industrial; Netherlands Consortium for Healthy Ageing; Netherlands Consortium for Systems Biology; Netherlands Metabolomics Centre; Netherlands Proteomics Centre; Netherlands Toxicogenomics Centre; Nutrigenomics Consortium; VIRGO Consortium

>Other affiliation (E.g. CMBI, TUDelft.) Erasmus MC

Project description

Title: R-package for Structural Variation Visualisation

> Please describe in some detail what the goal of the project is. What is the current situation and what is the desired outcome?

Provide a list of desired deliverables.

The aim is to provide PCMM scientists with a R-package that will display structural variation events with associated allele specific CNV across the SV region. The “state of the art” applications for SV representation include, inGAPSV, Circos and Gremlin (linear circus plot). The first method, inGAPSV, requires that the user processes the data with the methods supplied in the application and is not open to other detection methods, Circos requires programming skills and cannot zoom in to the break point sequence. Gremlin is a linear visualization of SV but similar to Circos the user cannot zoom in to the break point and we have been unable to access the code even after contact with the authors. There are several commercial genome viewers, Illumina, CLC and public IGV which will show the granularity and but since they use the underlying genome build cannot put two regions together in a single visualization. There are several R-packages which deal with allele specific CNV visualization but do not allow cross gene SV to be visualized since the order of the events cannot be changed. In addition GenomeGraphs which uses the underlying genomic organization also does not allow any change in genomic order. To circumvent these problems we have adopted a simple approach to use existing R libraries to solve this problem and resulted in a basic view.

**Proposed Deliverables**

1. R-package to visualize SV and CNV events
2. Optional – make view dynamic?
3. Application documentation
4. Method evaluation report of Case Study
5. User manual with Case Study

>Please describe why you would want support on this project? (E.g. missing expertise at our institution, use of high-performance computing resources or other infrastructural needs.)

This work has been prototyped by myself but I do not have the expertise to complete this project.

> Are you planning on publishing about this project?

Choose one of the following answers Yes, No, Maybe

Yes, as an addition to prototyping already completed in our department. The work will be a joint NBIC/Erasmus publication

> Any other remarks related to the project.

This work is part of the CTMM/NBIC TraIT WP4-NGS project

We will provide expertise on the test case, use cases and application development.

We will migrate this workflow to NBIC/CTMM Galaxy and/or to iFUSE for public use

We will use this validated workflow to analyse our CTMM data