A Shiny based tool for NGS data integration in Reproducible Spirit

Righelli D., Franzese M., Angelini C.
Background

- Sequencing techniques are developed to study cellular –omics

- Useful to inspect the cellular behaviour from many different point of view.

Image from: evodenomics, Coraline Petit, Master Biosciences, departement de Biologie, Ecole Normale Superieure de Lyon. 01-10-2012
Sequencing techniques are developed to study cellular -omics.

Useful to inspect the cellular behaviour from many different point of view.

Single -seqs are like a camera view in a multi-view camera system.

Single views need to be integrated to obtain the 3d model reconstruction.
How Analysis works

Single-omic pipelines
- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent
How Analysis works

**Single-omic pipelines**
- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent

**Multi-omic Integration**
- Emerging novel methods
- No past examples as guide
- No standards reached, yet
- Few implemented tools
How Analysis works

Single-omic pipelines
- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent

Multi-omic Integration
- Emerging novel methods
- No past examples as guide
- No standards reached, yet
- Few implemented tools

Focus on Next Generation Sequencing data
How to Integrate

Data
- Experimental data
- Online repositories
  - Consortium data
  - Annotation data
  - ...

Several integration levels
- Graphical level
- Pathway/GO Analysis
- Statistical methods
- Regulation Networks

References
- TUNCBAK, Nurcan, et al. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol,
- ...

[Diagram showing integration process with ChIP-Seq, RNA-Seq, BS-Seq, and online repository inputs leading to integrated data with graphical exploration, functional annotation, statistical methods, and integration network levels]
How to Integrate

Data
- Experimental data
- Online repositories
  - Consortium data
  - Annotation data
- ...

Several integration levels
- Graphical level
- Pathway/GO Analysis
- Statistical methods
- Regulation Networks

References
- TUNCBAĞ, Nurcan, et al. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol,
Reproducible Background

Handling problems
- Trace complex analysis
- Trace used methods
- Trace input/output data

Third party reproducibility
- Markdown language
- Data Caching
- Analysis Report

How it’s Implemented

**RNASeqGUI**
- Analyse RNA-Seq data
  - Preprocessing/normalization
  - DE methods
  - Pathway/GO analysis
  - Data inspection plots
- Reproducible Research
  - R-Markdown
  - Data Caching
    - filehash package
  - HTML Report
    - knitr package

How it’s Implemented

User Interactive Layer

RR Hidden Layer
How to Improve

Integration of High-throughput Omics data

Multi-omic Integration
How it’s Made

Graphical User Interface
- HTML based (shiny)
- Stand-alone
- No pipeline oriented
- Interactive plots (plotly)
- Multi-omic analysis
- Multi-omic integration
- Reproducible Research
- open-source

On-going implementation
- RNA-Seq data Analysis
- ChIP-Seq data Analysis
- BS-Seq data Analysis
- Live Markdown editing
How to Use

- Sequencing
  - RNA-Seq Gene Expression
  - RNA-Seq Isoform Analysis
  - ChIP-Seq
  - Methyl-Seq
  - Others

DE Genes
How to Use
How to Use DE Genes Called Peaks

- RNA-Seq Gene Expression
- RNA-Seq Isoform Analysis
- ChIP-Seq
- Methyl-Seq
- Others

DE Genes

Gene Annotation
- Gene Ontology
- Pathway Analysis
How to Use

Integration Level 0

Called Peaks

DE Genes

RNA-Seq Gene Expression
RNA-Seq Isoform Analysis
ChIP-Seq
Methyl-Seq
Others

Gene Annotation
Gene Ontology
Pathway Analysis
How to Use

Integration Level 0

Called Peaks

DE Genes

Gene Annotation
- Gene Ontology
- Pathway Analysis

Funct & Ann

Sequencing
- RNA-Seq Gene Expression
- RNA-Seq Isoform Analysis
- ChIP-Seq
- Methyl-Seq
- Others

Data Exploration
- Graphics
- Explore BAM files
- Explore BED files
- Explore Counts file
- Genomic Plots
- Data Integration
How to Use

Integration Level 0

Called Peaks — DE Genes — Gene Annotation — Data Exploration

RNA-Seq Gene Expression — RNA-Seq Isoform Analysis — ChIP-Seq — Methyl-Seq — Others

Graphics
- Explore BAM files
- Explore BED files
- Explore Counts file
- Genomic Plots
- Data Integration

Gene Ontology — Pathway Analysis
How to Use

Integration Level 0
How to Use

Integration Level 0

Integration Level 1
How it Reproduces

- Dedicated interface
- Fast R-markdown editing
- Fast report compilation
- Packages
  - shinyAce
  - rmarkdown/knitr
How it Reproduces

- Dedicated interface
- Fast R-markdown editing
- Fast report compilation
- Packages
  - shinyAce
  - rmarkdown/knitr
How it Caches

Motivations
- Helpful to reproduce analysis
- Stores input/output objects
- Useful to speed-up time demanding steps
- Useful to share data

Implementation
- Used to speed-up report compilation
- Implemented with R base data management system
- No external packages required
How it’s Improved
side mini projects

<table>
<thead>
<tr>
<th>Rmarkdown wrapper functions</th>
<th>Caching wrapper class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed up code reporting</td>
<td>Based on <em>RData</em> files</td>
</tr>
<tr>
<td>Standard rmarkdown based</td>
<td>No external dependencies</td>
</tr>
<tr>
<td><img src="https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf" alt="Image" /></td>
<td>R base implementation</td>
</tr>
<tr>
<td><em>rmarkdown</em> package</td>
<td></td>
</tr>
<tr>
<td>No Markdown knowledge necessary</td>
<td></td>
</tr>
<tr>
<td>If caching is reported</td>
<td></td>
</tr>
<tr>
<td>Some functions depends on caching wrapper class</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Online Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://github.com/drighelli/RR" alt="Image" /></td>
</tr>
</tbody>
</table>

(on-going development)
**What’s next**

**Integration methods**
- Graphical methods
- Statistical methods
- Integration networks

**User helping functionalities**
- Preselected analysis workflow
- Live help
- Live results inspection

**Reproducible Research**
- Live browsing of caching file
- Analysis step tracking
  - Graph construction
Thank you for Your Attention!
How to Use

DE Genes

Called Peaks

Funct & Ann

RNA-Seq Gene Expression
RNA-Seq Isoform Analysis
ChIP-Seq
Methyl-Seq
Others

Gene Annotation
Gene Ontology
Pathway Analysis

Seq & Ann

Sequencing

IntegrHO