A workflow for multi-omic data analysis

Vincent Henry
The first omic revolution: Unprecedented perspectives for life-science research

Technological advances

Experimental approaches

Analytical methods

Storage

Organization

Development of specific structures for bioinformatics experts
The second omic revolution: The fall of constraints

Technological advances

Experimental approaches

Analytical methods

- 3rd generation sequencers
- Cloud computing

Regular use for life-scientists from other fields of expertise ("newcomers")
Current challenge

- Transformation of raw data into interpretable data
- Biological interpretation with high level of relevance
- Development of integrated approaches

Promises based on by a wide range of bioinformatics tools:
- analytical software
- knowledge- and data- bases
The large field of omic science

Technologies

- NGS
- Bioarrays
- PCR
- MS
- RMN
The large field of omic science

Technologies:
- NGS
- Bioarrays
- PCR
- MS
- RMN

Applications:
- Genomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics
- Drug discovery

Biological systems:
The large field of omic science

Technologies
- NGS
- Bioarrays
- PCR
- MS
- RMN

Applications
- Genomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics
- Drug discovery

Analysis
- Genome
- Ex 1 Ex 2

Biological systems
- Genomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics
- Drug discovery
The large field of omic science

Technologies
- NGS
- Bioarrays
- PCR
- MS
- RMN

Applications
- Genomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics
- Drug discovery

Functional analysis

Biological systems
- Genomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics

OMIC tools
OMICtools: an answer to the need of organization

OMICtools: an unique website for indexing and classification of analytic software and databases

A didactic knowledge base free of access

Wide range of editorial topic

Genomics → Epigenomics → Transcriptomics → Proteomics → Metabolomics → Drug discovery

Aim to:
- Help newcomers to find appropriate tools for their needs
- Help developers to stay up to date and to avoid redundancy
- Promote communication between newcomers and developers

Specificities: Classification, didactics, traceability
Tools classification

Analytical software and databases are classified across more than 700 hierarchical categories

- Technologies
- Biological systems

Didactics

- Iconography
- Interconnection
- Workflow
Tools classification

Analytical software and databases are classified across more than 700 hierarchical categories.

- Applications
- Functional analysis

Didactics

- Iconography
- Interconnection
- Workflow
Tools classification

Analytical software and databases are classified across more than 700 hierarchical categories

- Analytic steps

Didactics

- Iconography
- Interconnection
- Workflow
Tools classification

Analytical software and databases are classified across more than 700 hierarchical categories.

- Workflow

Didactics

- Iconography
- Interconnection
- Workflow

In recent years, the continuing technical improvements and decreasing cost of next-generation sequencing technology have made RNA sequencing (RNA-seq) a popular choice for gene expression studies (Wang et al., 2008; Ozsolak and Milos, 2011). Such sequence-based methods have revolutionized studies of the transcriptome by enabling a wide range of novel applications, including detection of alternative splicing isoforms and transcript fusion detection. In addition, RNA-seq has become an attractive alternative to microarrays for the identification of differentially expressed genes between several conditions or tissues.

Curators: Abdoulaye Baniré Diallo, Rekin's Janky, Bogdan Mirauta and Mikhail Shugay
Tools classification

More than 9000 references

**A comparative study of techniques for differential expression analysis on RNA-Seq data**


Abstract: Recent advances in next-generation sequencing...

**A comparison of methods for differential expression analysis of RNA-seq data**

Authors: Soneson, C., and Delorenzi, M.

Abstract: BACKGROUND: Finding genes that are differentially expressed between conditions is an integral part of understanding the molecular basis of phenotypic variation. In...

**A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data**

Authors: Kvam, V.M., Liu, P., and Si, Y.

Abstract: RNA-Seq technologies are quickly revolutionizing genomic studies, and statistical methods for RNA-seq data are under continuous development. Timely review and...

**A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis**

Abstract: High-throughput sequencing assays and tests for differential expression...

---

Classification and didactics
Tools classification

-Purpose
-Metadata

antaRNA

It is capable of compiling RNA sequences for a given structure that comply in addition with an adjustable full range objective G/C content distribution, specific sequence constraints and additional fuzzy structure constraints. antaRNA applies ant colony optimization meta-heuristics and its superior performance is shown on a biological datasets.

Visit antaRNA

<table>
<thead>
<tr>
<th>RRID</th>
<th>OMICS_08738</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of tool</td>
<td>Program</td>
</tr>
<tr>
<td>Nature of tool</td>
<td>Standalone</td>
</tr>
<tr>
<td>Interface</td>
<td>Command line</td>
</tr>
<tr>
<td>Language</td>
<td>Python</td>
</tr>
<tr>
<td>Depends</td>
<td>RNAfold of the ViennaRNA Package version 2.1.3, Python 2.7.3, numpy</td>
</tr>
<tr>
<td>Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany</td>
<td></td>
</tr>
<tr>
<td>Maintainer</td>
<td>Backofen Rolf</td>
</tr>
<tr>
<td>Email</td>
<td>backofen at informatik.uni-freiburg.de</td>
</tr>
</tbody>
</table>

Specific supports for communication
OMICtools allows the automatic detection of break URLs.
A scientific publication in “Database” (Henry et al., 2014)

Collaboration
France

- Laboratoire NeoVasc ERI-28 (Université de Rouen)
- Laboratoire LITIS and CISMef team (Rouen Hospital)
- BiblioInserm portal
- Collaborative convention between Université de Rouen, InsermTransfert and NéoVasc

World

- Center for Research in Biological Systems, University of California, San Diego
- Consortium Force 11
Community

Volunteer curators

Collaborations
**OMICtools metrics**

- Page views per month: 102,091
- Visits per month: 41,406
- Unique visits per month: 29,942
- Avg. session duration: 00:02:28
- Pages/Session: 2.47

*Period: 01 May 2015 - 31 May 2015*

**Unique visitors**

- Bar chart showing monthly visitors from August 2013 to May 2015.

**Geographic breakdown (%)**

- Europe: 35%
- America: 33%
- Asia: 29%
- Africa: 1%
- Oceania: 2%

**Period: 2014**
Key points

An original website free of access

- More than 9000 classified analytical software and databases for “omic” data analysis and interpretation

- A publication in DATABASE

- 30 international volunteer curators who bring expertise in their categories

- 12 databases which bring their specificities

- Unique identifiers (RRID) used by publishers

- More than 1900 unique visitors /days

An international and collaborative project for life-science community
**Editable tool cards**

This package implements a Poisson mixture model to cluster observations (e.g., genes) in high throughput sequencing data. Parameter estimation is performed using either the EM or CEM algorithm, and the slope heuristics are used for model selection (i.e., to choose the number of clusters).

**HTSCluster**

<table>
<thead>
<tr>
<th>RRID</th>
<th>OMICS_07338</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current release</td>
<td>HTSCluster Version 2.0.4</td>
</tr>
<tr>
<td>Type of tool</td>
<td>Program</td>
</tr>
<tr>
<td>Nature of tool</td>
<td>Package</td>
</tr>
<tr>
<td>Interface</td>
<td>Command line</td>
</tr>
<tr>
<td>Operating system Programming language</td>
<td>Unix/Linux, Mac OS X, Windows R</td>
</tr>
<tr>
<td>License</td>
<td>GPL</td>
</tr>
<tr>
<td>Depends</td>
<td>R (≥ 2.10.0), capushe</td>
</tr>
<tr>
<td>Maintainer</td>
<td>Andrea Rau</td>
</tr>
<tr>
<td>Email</td>
<td>&lt;andrea.rau at jouy.inra.fr&gt;</td>
</tr>
<tr>
<td>Article</td>
<td>(Rau et al., 2015) Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. Bioinformatics.</td>
</tr>
</tbody>
</table>

**Participate**

- [Write a review](#)
- [Report a problem](#)
Editable tool cards

HTSCluster

This package implements a Poisson mixture model to cluster observations (e.g., genes) in high throughput sequencing data. Parameter estimation is performed using either the EM or CEM algorithm, and the slope heuristics are used for model selection (i.e., to choose the number of clusters).

<table>
<thead>
<tr>
<th>RRID</th>
<th>OMICS_07338</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current release</td>
<td>HTSCluster Version 2.0.4</td>
</tr>
<tr>
<td>Type of tool</td>
<td>Program</td>
</tr>
<tr>
<td>Nature of tool</td>
<td>Package</td>
</tr>
<tr>
<td>Interface</td>
<td>Command line</td>
</tr>
<tr>
<td>Operating system</td>
<td>Unix/Linux, Mac OS X, Windows</td>
</tr>
<tr>
<td>Programming language</td>
<td>R</td>
</tr>
<tr>
<td>License</td>
<td>GPL</td>
</tr>
<tr>
<td>Depends</td>
<td>R (≥ 2.10.0), capusho</td>
</tr>
<tr>
<td>Maintainer</td>
<td>Andrea Rau</td>
</tr>
<tr>
<td>Email</td>
<td>&lt;andrea.rau at jouy.inra.fr&gt;</td>
</tr>
<tr>
<td>Article</td>
<td>(Rau et al., 2015) Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. Bioinformatics.</td>
</tr>
</tbody>
</table>
Perspectives: A new platform more open to the community

Editable tool cards

Tool repository

HTSCluster

This package implements a Poisson mixture model to cluster observations (e.g., genes) in high-throughput sequencing data. Parameter estimation is performed using either the EM or CEM algorithm, and the slope heuristics are used for model selection (i.e., to choose the number of clusters).

Visit HTSCluster

RRID: OMICS_07338
Current release: HTSCluster Version 2.0.4
Type of tool: Program
Nature of tool: Package
Interface: Command line
Operating system: Unix/Linux, Mac OS X, Windows
Programming language: R
License: GPL
Depends: R (≥ 2.10.0), capusho
Maintainer: Andrea Rau
Email: andrea.rau at jouy.inra.fr


GitLab

Trunks
Branches
Merges
Tags
Discontinued development branch

Participate
Write a review
Report a problem
Merci

Dr Arnaud Desfeux
Dr Anne-Sophie Pépin

Dr Bruno Gonzalez
Pr. Soumeya Bekri
Michelle Hauchecorne
Céline Lesueur

Pr. Stéfan Darmoni
Dr Lina Soualmia
Dr Julien Grosjean

Dr Hélène Dauchel

Caire Valencien
Christian Fondrat

Dr Didier Pélaprat

http://omictools.com

arnaud.desfeux@omictools.com
vincent.henry@omictools.com