

Analysis of iCLIP data with the iCount python module

A short tutorial

Tomaž Curk
tomaz.curk@fri.uni-lj.si

University of Ljubljana
Faculty of Computer and Information Science

The Crick Institute, July 2017



University of Ljubljana
Faculty of Computer and
Information Science

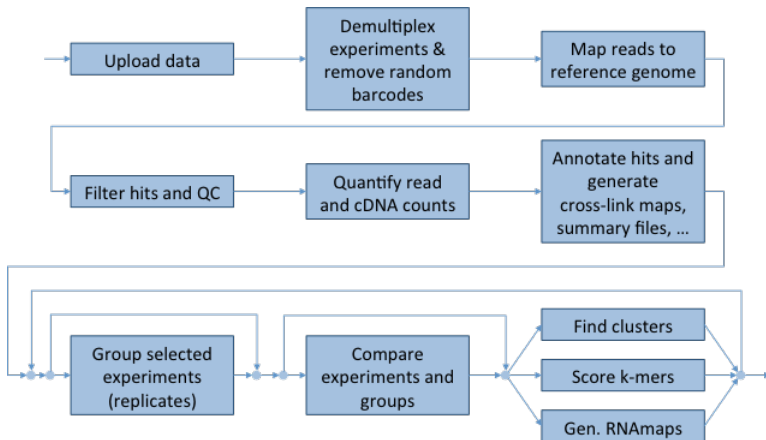
History and acknowledgements for iCount

From its start in late 2008, a great number of people contributed to the development of iCount.

In mid-2016, Jure Zmrzlikar from Genialis helped to improve the code, which is now available on github.

- Tomaž Curk
- Gregor Rot
- Črtomir Gorup
- Julian König
- Yoichiro Sugimoto
- Nejc Haberman
- Goran Bobojević
- Jure Zmrzlikar
- Christian Hauer
- Matthias Hentze
- Blaž Zupan
- Jernej Ule

Pipeline



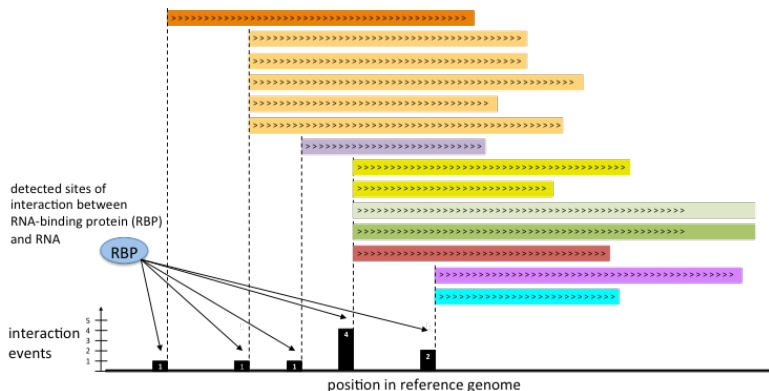
Main steps in the cross-link mapping and annotation pipeline.

Finding cross-linked sites

Mapping and quantification of protein-RNA interactions

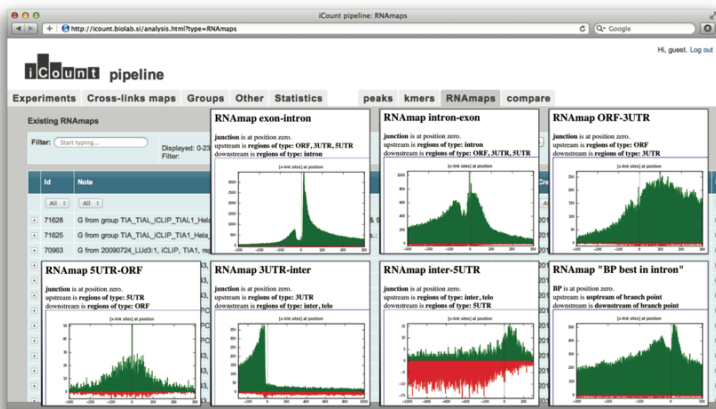
input: >20M sequencing reads (50nt long) / experiment

[>2GB / experiment]



The most crucial step is site identification and quantification.

The old web interface



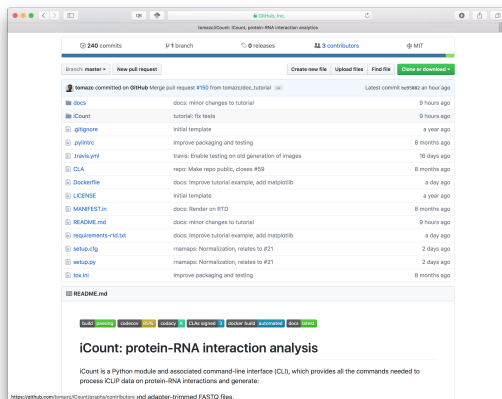
This interface is still in use.

It will be replaced soon by an implementation by Genialis.

The old iCount - impact

- 100+ users
- 40+ institutions: UCL, MRC LMB, EMBL, mpi-cbg.de, ...
- 2M+ EUR cost of sequencing alone
- 80+ proteins (8% of known human RBPs)
- 5000+ experiments (iCLIP, CLIP, pA-Seq, mRNA-seq, ...)
- 800+ user-defined groupings of experiments
- 10 species
- 33000 analyses
- (Co)authors of over 25 publications using results from iCount.

iCount is available on github!



<http://github.com/tomazc/iCount>
Comments and PR are welcome!

Tutorial

The screenshot shows a web browser window displaying the iCount documentation. The page title is "Tutorial" and the URL is "icount.readthedocs.io". The left sidebar contains a navigation menu with links to "iCount", "Navigation", "Tutorial", "Installation", "Reference", "Contributing", "Frequently asked questions", "Version history", "License", and "How to cite". The main content area is titled "Tutorial" and contains the following text:

You will need to install iCount first. Please, follow [these instructions](#).

iCount provides all commands needed to process FASTQ files with iCLIP sequencing data and generate BED files listing identified and quantified cross-linked sites.

iCount uses and generates a number of files. We suggest you run this tutorial in an empty folder:

```
$ mkdir tutorial_example
$ cd tutorial_example
```

Preparing a genome index

iCLIP sequencing reads must be mapped to a reference genome. The user can prepare its own **FASTA genome sequence** and **GTf genome annotation** files.

Another option is to download a release from [ensembl](#). You can use the command **releases** to get a list of available releases supported by iCount:

```
$ iCount releases
```

There are 38 releases available: 88,87,86,85,84,83,82,81,80,79,78,77,76,75,74,73,72,71,70,69,68,67,66,65,64,63,62,61,60,59

You can then use the command **species** to get a list of species available in a release:

```
$ iCount species -r 88
```

There are 87 species available: alluropoda_melanoleuca,anas_platyrhynchos,ancestral_alleles,anolis_carolinensis,astyanax_mexicanus,bos_taurus,...gorilla_gorilla,homo_sapiens,ictidomys_tridecemlineatus,latimeria_chalumnae,

<http://icount.readthedocs.io/en/latest/tutorial.html>

Use and contribute to iCount

In case of problems, post an issue on github or contact:

tomaz.curk@fri.uni-lj.si

Welcome to contribute to <http://github.com/tomazc/iCount>