# Towards Bioinformatics for All: Galaxy at UoM

Peter Briggs*, Ian Donaldson, Sarah Griffiths and Leo Zeef
University of Manchester, Manchester, UK
** Contact author: peter.briggs@manchester.ac.uk*

## Introduction

Within the Bioinformatics Core Facility (BCF) at the University of Manchester (UoM) we have developed a number of bespoke Galaxy tools to support local researchers conducting next generation sequencing (NGS) analyses. Our aim is to enable non-bioinformaticians to analyse their own data, circumventing installation and use of command line programs, by creating and distributing new and easy to install Galaxy tools to fill current gaps, and ultimately bring "bioinformatics to all".

## Who we are

The BCF provides university-wide support to researchers, but primarily in the Faculty of Life Sciences and School of Medicine, by performing routine and custom analyses of NGS data including RNA-seq, ChIP-seq and metagenomics. We also provide software support, which includes the maintenance of a local Galaxy instance and the development of bespoke Galaxy tools.


BCF group at UoM:
(front row: Leo Zeef, Ian Donaldson, Ping Wang; back row: Tim Burgis, Peter Briggs)

## Trimmomatic: preparing NGS data

**Trimmomatic** [1] is a flexible "pair-aware" utility for preparing Illumina sequence data for analysis by trimming or removing poor quality reads, and is an essential first step common to many NGS analyses.

- Our tool provides access to all **Trimmomatic**'s functions within Galaxy
- Installation from the toolshed automatically sets up the program and the reference data for immediate use:
  **https://toolshed.g2.bx.psu.edu/view/pjbriggs/trimmomatic**

## PALFinder: microsatellite detection

**PALFinder** [2] detects microsatellite repeats and designs PCR primers to amplify them. It is quicker and more reliable than designing primers by hand, but can require substantial compute power to run on large NGS datasets.

- Incorporating **PALFinder** within Galaxy allowed researchers unfamiliar with cluster computing to run the program easily and to build reproducible workflows
- Allowed additional filtering scripts to be incorporated
- **PALFinder** within Galaxy is now an essential part of local primer design protocols

Installation from the toolshed ensures correct versions of dependencies (e.g. Perl) are used automatically: **https://toolshed.g2.bx.psu.edu/view/pjbriggs/pal_finder**

## MACS2, CEAS, Weeder2: ChIP-seq analysis

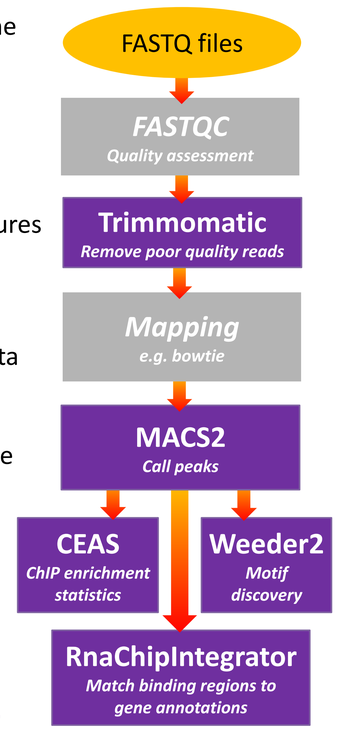
**ChIP-seq analysis** is one of the standard protocols offered by the BCF, and makes use of the following software:

- **MACS2**[3] for peak calling
- **CEAS** [4] for statistics on ChIP enrichment at important genome features
- **Weeder2** [5] for motif discovery
- **RnaChipIntegrator** [6] to associate binding regions with gene annotation or other features

We have developed Galaxy tools for each of these programs, to fill gaps in the workflow (**Weeder2**, **RnaChipIntegrator**) and extend existing tools to add new or novel features (**MACS2**, **CEAS**). Providing a full Galaxy toolset allows our users to further explore their data after it has left the BCF. Novel features include:

- Enabling implicit generation of bigWig files from **MACS2** that can be used directly in the **CEAS** tool via the Cistrome-modified version of CEAS ("ceasBW") [7]
- Adding data managers [8] into **CEAS** and **RnaChipIntegrator** tools to make installation and management of reference data easier

- **MACS2: https://toolshed.g2.bx.psu.edu/view/pjbriggs/macs21**
- **CEAS: https://toolshed.g2.bx.psu.edu/view/pjbriggs/ceas**
- **Weeder2: https://toolshed.g2.bx.psu.edu/view/pjbriggs/weeder2**
- **RnaChipIntegrator: https://toolshed.g2.bx.psu.edu/view/pjbriggs/rnachipintegrator**




## Availability

UoM researchers can access the tools via a private Galaxy instance maintained by the BCF. We have also published the tools on the public toolshed so that they are available to the wider Galaxy community, and so they can be cited in academic publications. Installation from the Galaxy toolshed ensures that the underlying dependencies are also installed, with data managers to assist with setting up reference data, so they can be used "out of the box" with minimal additional configuration.

## What's next

Our aim is to provide UoM researchers with access to bioinformatics programs that they would not otherwise be able to use, by developing new tools and delivering them via our local Galaxy instance. We also intend to provide these tools to non-experts in the wider community via the toolshed, making them as easy as possible to install and configure, and to promote good practice by using them as a framework for generally-available tutorials that could be run on desktops via Biolinux [9] or Galaxy-Docker [10].

## Acknowledgements

Peter Briggs, Ian Donaldson and Leo Zeef are funded by the University of Manchester; Sarah Griffiths is funded by NERC. The PALfinder work was initiated with Nathan Truelove (now at the Smithsonian Marine Station in Fort Pierce, USA).

## Resources

Tutorial videos available via BCF YouTube channel:
**https://www.youtube.com/user/ManchesterBCF/**
Blog on administering Galaxy deployments:
**http://galacticengineer.blogspot.co.uk/**

## Web links

[1] – Trimmomatic **http://goo.gl/5BlRyn**
[2] – PALfinder **http://goo.gl/s8tswT**
[3] – MACS2 **https://goo.gl/iGiHPE**
[4] – CEAS **http://goo.gl/gv0tFZ**
[5] – Weeder2 **http://159.149.160.51/modtools/**
[6] – RnaChipIntegrator **https://goo.gl/jpyYpG**
[7] – Cistome CEAS **https://goo.gl/D2QpsF**
[8] – Galaxy data managers **https://goo.gl/XK9leq**
[9] – Galaxy in Biolinux **http://goo.gl/AjWxti**
[10] – Galaxy-Docker **https://goo.gl/dx6ETZ**