Yassine Souilmi

yassine@souilmi.me

Ph.D. Student

Scientific workflow

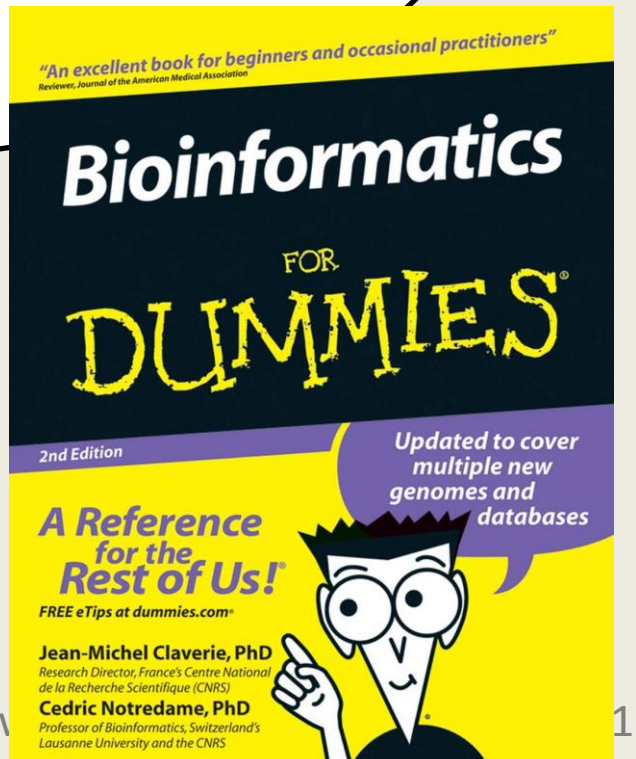
Data analysis

Data publishing

Galaxy Team, T. (2010)

Galaxy

Scientists with no computer programming experience



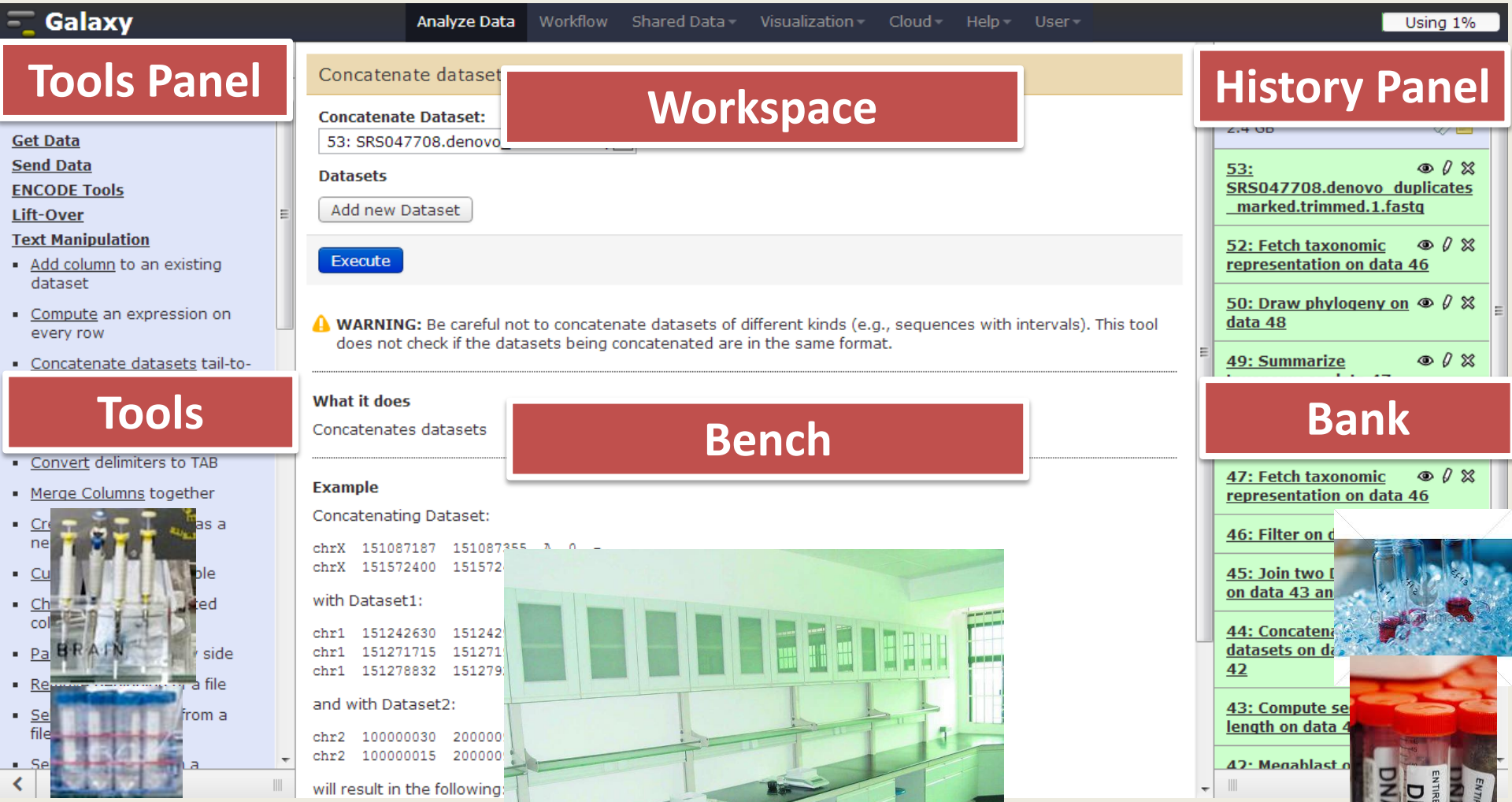
Galaxy Team

- **The Project is developed by**
  - BX, Center for Comparative Genomics and Bioinformatics at Penn State, Pennsylvania
  - Biology and Mathematics and Computer Science departments at Emory University, Georgia

- **The Project is supported in part by**
  - NHGRI, National Human Genome Research Institute
  - NSF, National Science Foundation
  - The Huck Institutes of the Life Sciences at Penn State, Pennsylvania
  - The Institute for Cyber Science at Penn State, Pennsylvania
  - Emory University, Georgia

```r
each <- function(.column,.data,.lambda){

  # Find the column index from it's name
  column_index <- which(
    names(.data) == .column)

  # Find the unique values in the column
  column_levels <- unique(.data[,column_index])

  # Loop over these values
  for(i in 1:length(column_levels)){

    # Subset the data and call the function
    .lambda(.data[
      .data[,column_index] == column_levels[i],
    ])
  }
}
```

Graphical User Interface



Why Galaxy ?

- Accessibility
- Reproducibility
- Transparency
- Availability

UCSC Genome Bioinformatics

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the ENCODE and Neandertal projects.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list.

News

News Archives ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

`20 December 2012 - 28 New Vertebrate Assemblies!`

Over the past several weeks, we have released 28 new vertebrate assemblies on the public Genome Browser website, featuring 22 new species and 6 assembly updates. These assemblies were added to support the 60-species Conservation track on the latest mouse assembly (mm10/GRCm38). Several of these species were originally sequenced and assembled for the Mammalian Genome Project (Lindblad-Toh *et al.*, *Nature* 2011)*.

Primates:

- **Baboon** (*Papio hamadryas*) **papHam1** – Pham_1.0 (Nov. 2008) from the Baylor College of Medicine HGSC
- **Bushbaby** (*Otolemur garnettii*) **otoGar3** – OtoGar3 (Mar. 2011) from the Broad Institute

Primer3 (v. 0.4.0) Pick primers from a DNA sequence.	Checks for mispriming in template.	disclaimer	Primer3 Home
	Primer3plus interface	cautions	FAQ/WIKI

There is a newer version of Primer3 available at http://primer3.wi.mit.edu/

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a Mispriming Library (repeat library): NONE

☑ Pick left primer, or use left primer below:	☐ Pick hybridization probe (internal oligo), or use oligo below:	☑ Pick right primer, or use right primer below (5' to 3' on opposite strand):

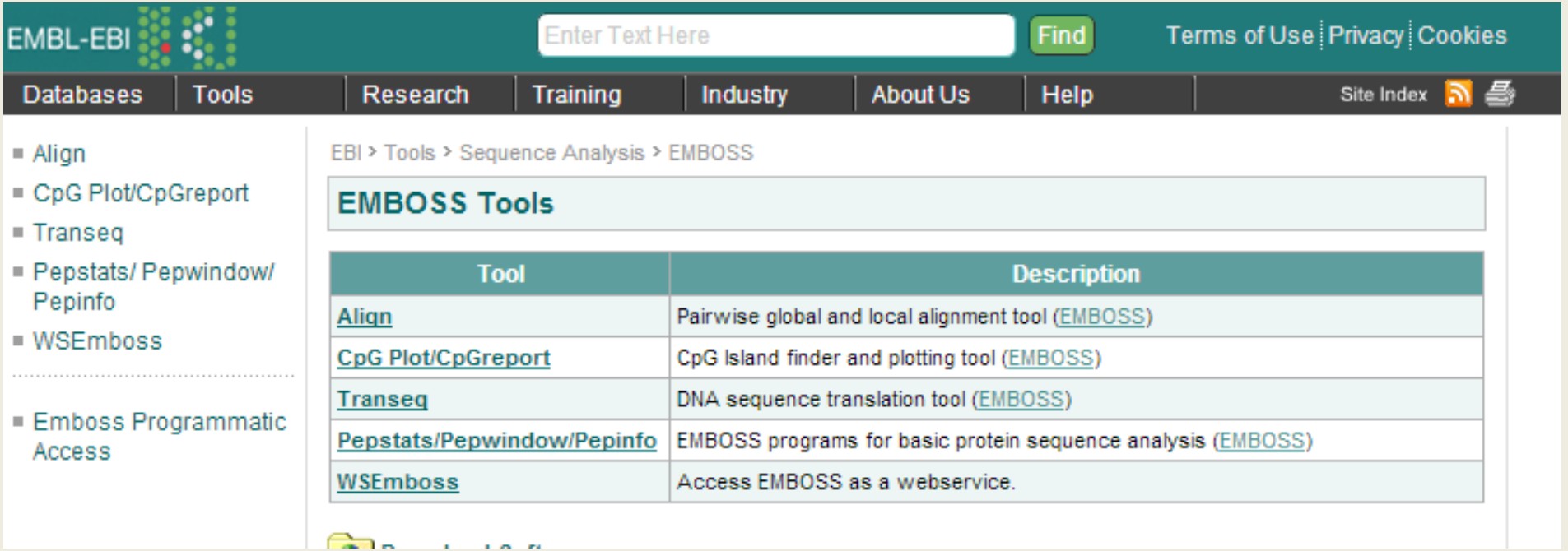
Pick Primers Reset Form

gatk Home About Guide Community Downloads Search

The Genome Analysis Toolkit or GATK is a software package developed at the Broad Institute to analyse next-generation resequencing data. The toolkit offers a wide variety of tools, with a primary focus on variant discovery and genotyping as well as strong emphasis on data quality assurance. Its robust architecture, powerful processing engine and high-performance computing features make it capable of taking on projects of any size.

Learn more »

Galaxy


EMBL-EBI
Enter Text Here
Find
Terms of Use
Privacy
Cookies
Databases
Tools
Research
Training
Industry
About Us
Help
Site Index
Align
CpG Plot/CpGreport
Transeq
Pepstats/ Pepwindow/ Pepinfo
WSEmboss
Emboss Programmatic Access
EBI > Tools > Sequence Analysis > EMBOSS
EMBOSS Tools
Tool
Description
Align
Pairwise global and local alignment tool (EMBOSS)
CpG Plot/CpGreport
CpG Island finder and plotting tool (EMBOSS)
Transeq
DNA sequence translation tool (EMBOSS)
Pepstats/Pepwindow/Pepinfo
EMBOSS programs for basic protein sequence analysis (EMBOSS)
WSEmboss
Access EMBOSS as a webservice.

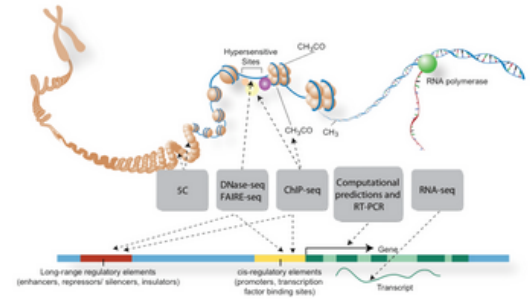
Galaxy

Encyclopedia of DNA Elements

- **Human**
- Integrative Analysis
- Experiment Matrix
- Experiment List
- Search
- Downloads
- Genome Browser (hg19)
- Session Gallery

About ENCODE Data

The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.



Click to enlarge

ENCODE data are now available for the entire human genome. ***All ENCODE data are free and available for immediate use via*** :

- Search for displayable tracks and downloadable files
- Download of data files
- Visualization in the UCSC Genome Browser (ENCODE data marked with the NHGRI logo)
- Data mining with the UCSC Table Browser and other UCSC Genome Bioinformatics tools

Galaxy

400 function !

Hands On exercise