Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L - 2022

Pierre-Luc Germain

ETH Zürich

Plan for today

- About your lecturer
- About your fellow students
- What’s epigenetics/epigenomics?
- Structure of the course
- Expectations and evaluations

- Introduction to the practical tools for the course:
  - R notebooks, bioconductor, git & github

About your lecturer

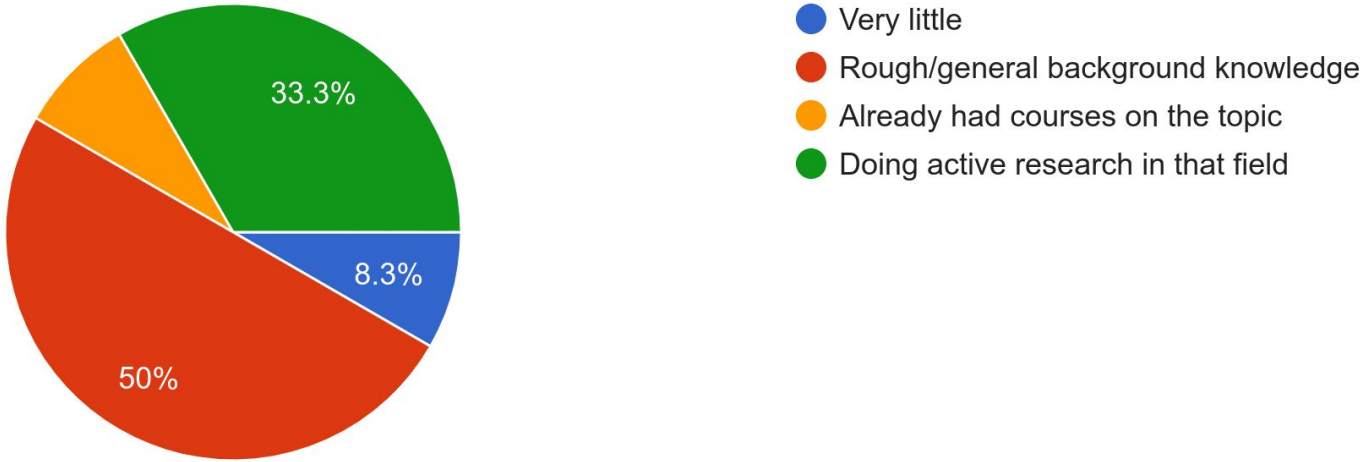
- Oberassistent at the D-HEST Institute for Neurosciences (ETH) & Lab of Statistical Bioinformatics (UZH)
- From Quebec, Canada 
- Background in the humanities (history and philosophy of the life sciences):
    - biological functions & evolutionary explanations, animal experimentation, modeling & extrapolation, molecular oncology, etc.
- Biological research:
    - gene expression regulation in stem cells, germline and the brain; brain and cellular responses to stress; genetics of neurodevelopmental disorders
- Bioinformatic research:
    - methods for bulk and single-cell transcriptomics and epigenomics; miRNA target prediction; TF binding & activity inference, reconstruction of transcriptional networks

About your fellow students

Chiefly master students, a few doctoral students

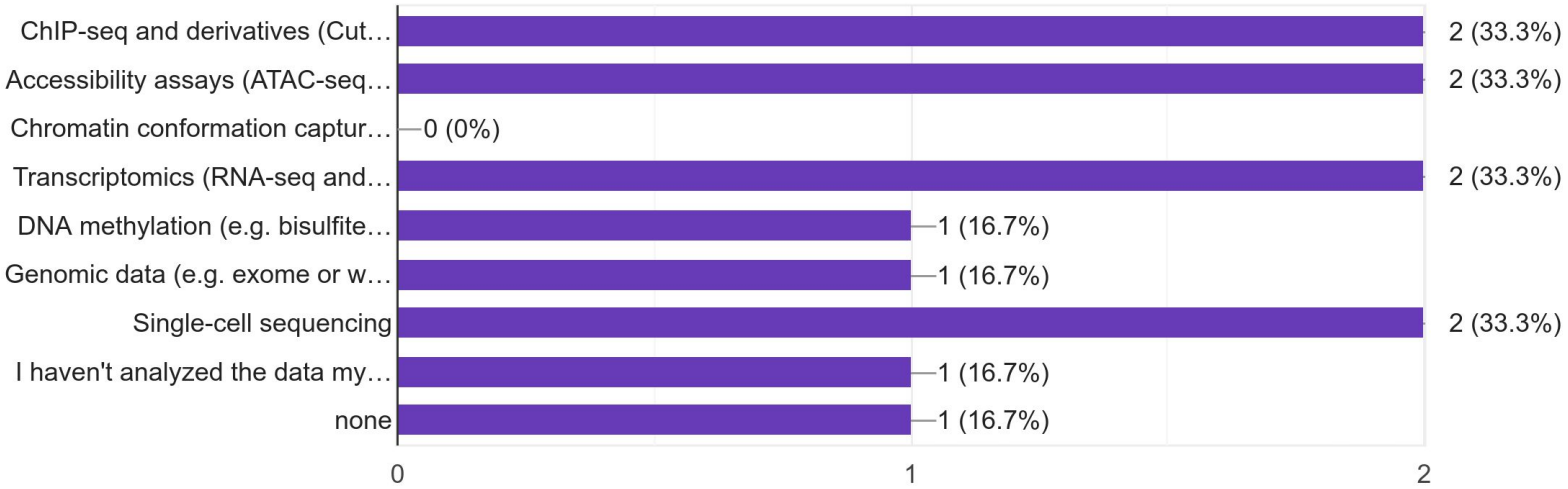
Prior knowledge about epigenetics & regulatory genetics

12 responses



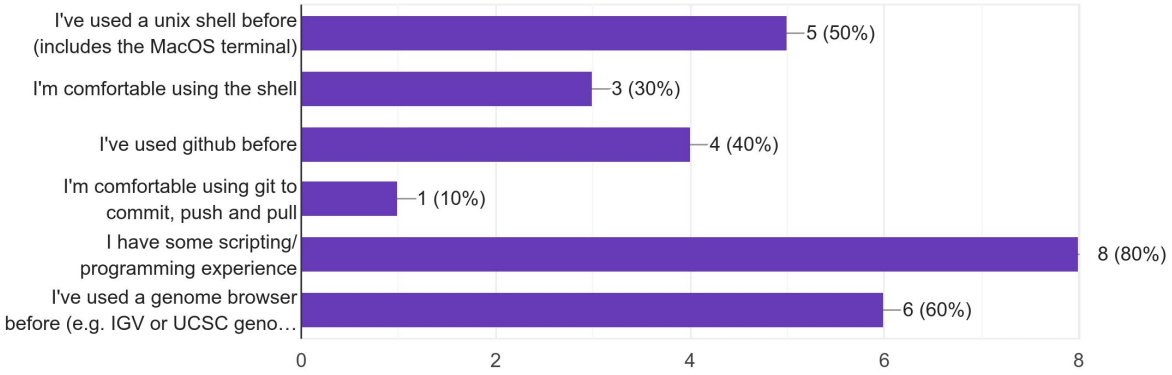
If you have worked with Next Generation Sequencing data before, what kind?

6 responses



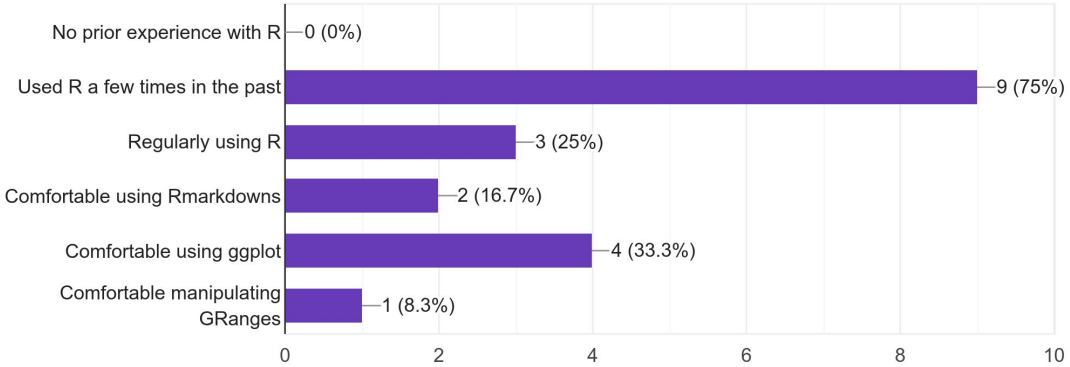
General computer skills; please check all that are relevant

10 responses



Prior experience with R/Bioconductor; please check all that are relevant

12 responses



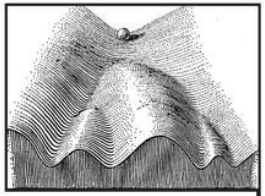
Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L - 2022

Pierre-Luc Germain

ETH Zürich

ep•i•genet•ics: (*n*) ...



1957

1942

WADDINGTON

"...to discover the causal mechanisms [by which the genes of the genotype bring about phenotypic effects], and to relate them as far as possible to what experimental embryology has already revealed of the mechanics of development. We might use the name 'epigenetics' for such studies..."

1958

NANNEY

"The term "epigenetic" is chosen to emphasize the reliance of [the supplementary regulatory] systems on the genetic systems and to underscore their significance in developmental processes."

1993

HERRING

"... is the entire series of interactions among cells and cell products which leads to morphogenesis and differentiation"

2007

PTASHNE

"... corresponds to a change in the state of expression of a gene that does not involve a mutation, but that is nevertheless inherited in the absence of the signal that initiated the change"

1987

HOLLIDAY

"... the strategy of the genes in unfolding the genetic program for development"

DNA methylation HOLLIDAY, RIGGS

normalized fraction of articles with 'epigen - '

1 in:
100
1,000
10,000

"epigen-" in publications

rec DNA - genetic reductionism

1940 1950 1960 1970 1980 1990 2000 2010

DNA = genetic material AVERY

Biochemical bistability DELBRÜCK

DNA structure WATSON, CRICK

HUXLEY

"...denote[s] the analytic study of individual development (ontogeny) with its central problem of differentiation"

1956

EPHRUSSI

"We must admit that not everything that is inherited is genetic"

1958

Gene regulation MONOD, JACOB

Attractors = cell type KAUFFMAN

HOLLIDAY

"... the programme for development depends on a temporal sequence of epigenetic changes..."
"... epigenetic or non-mutational origin of cancer"

1979

RIGGS

"... is the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence"
"...mechanisms include DNA methylation and histone modification"

1996

Transcriptome attractor HUANG

BIRD

"... depends on a structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states"

2007

(Oliveira Pisco, Fouquier d'Hérouël and Huang, 2016)

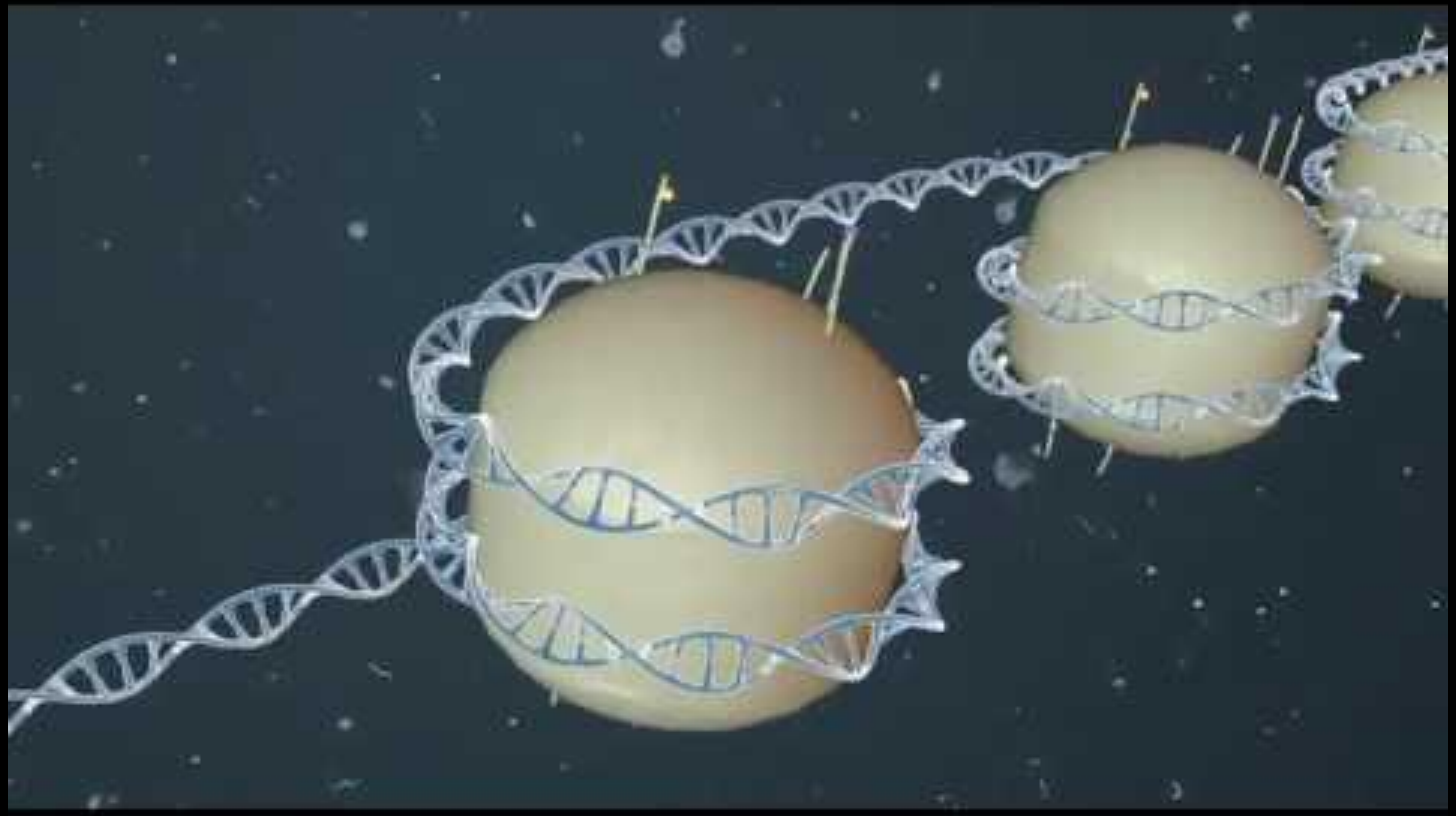
Epigenetic(s)

- Waddington (1942) :
  “the causal interactions between genes and their products which bring the phenotype into being”

- Ptashne (2007) (also Nanay 1958):
  "A change in the state of expression of a gene that does not involve a mutation, but that is nevertheless **inherited in the absence of the signal** (or event) that initiated the change."

- Bird (2007) and common contemporary usage in molecular biology:
  "molecular **modifications of DNA and chromatin** that do not alter the sequence"

Basic primer on epigenetics



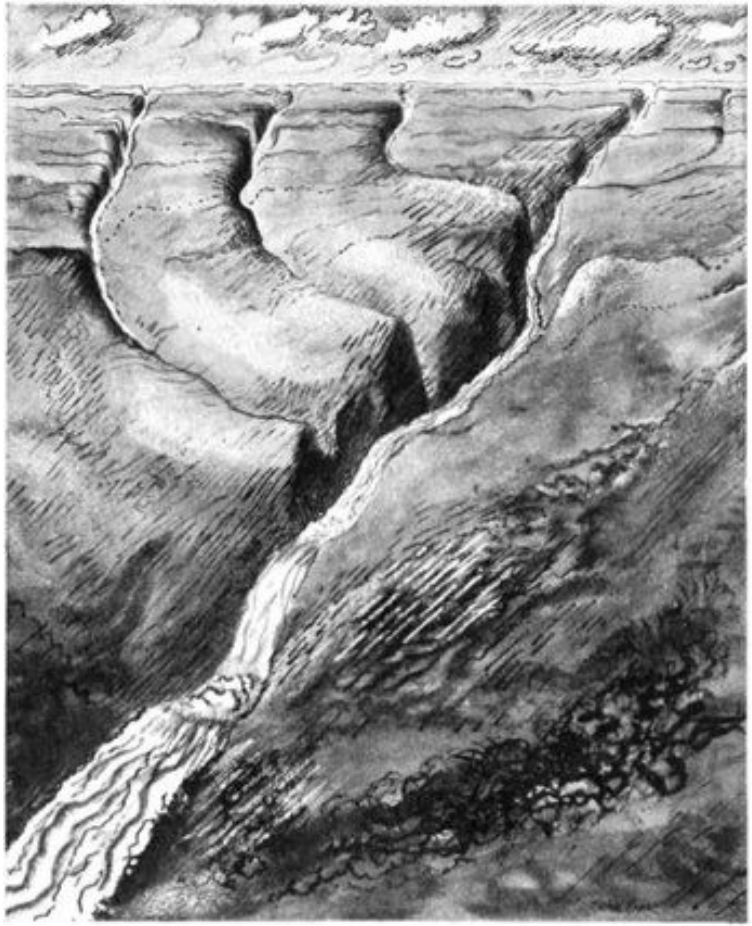
Epigenetic(s)

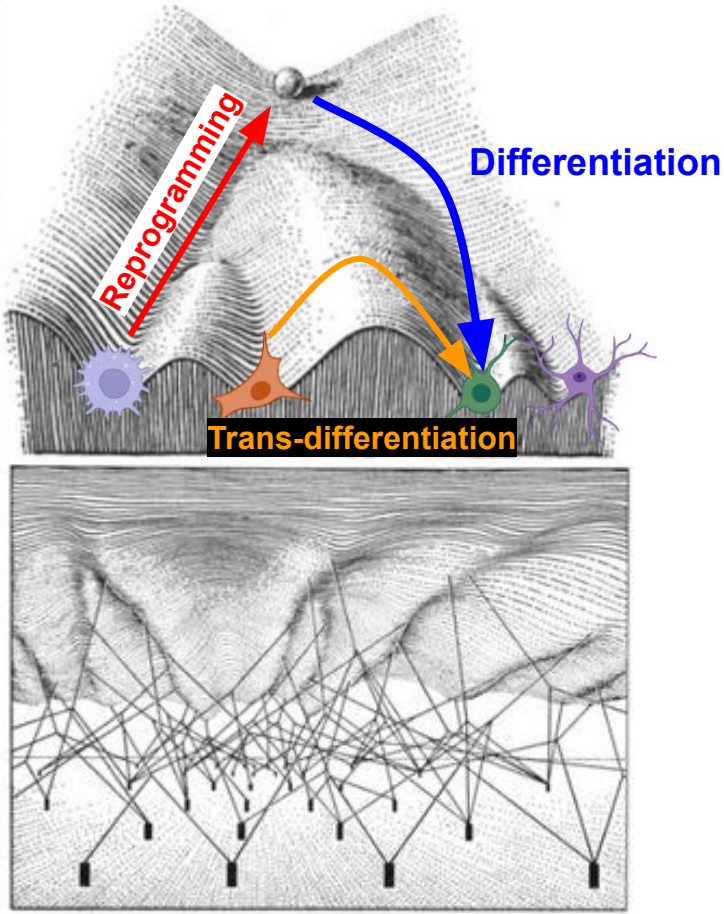
- Waddington (1942) :
  “the causal interactions between genes and their products which bring the phenotype into being”

- Ptashne (2007) (also Nanay 1958):
  "A change in the state of expression of a gene that does not involve a mutation, but that is nevertheless **inherited in the absence of the signal** (or event) that initiated the change."

- Bird (2007) and common contemporary usage in molecular biology:
  "molecular **modifications of DNA and chromatin** that do not alter the sequence"

Waddington's epigenetic landscape

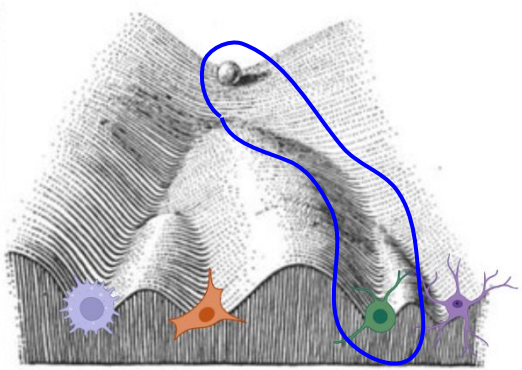


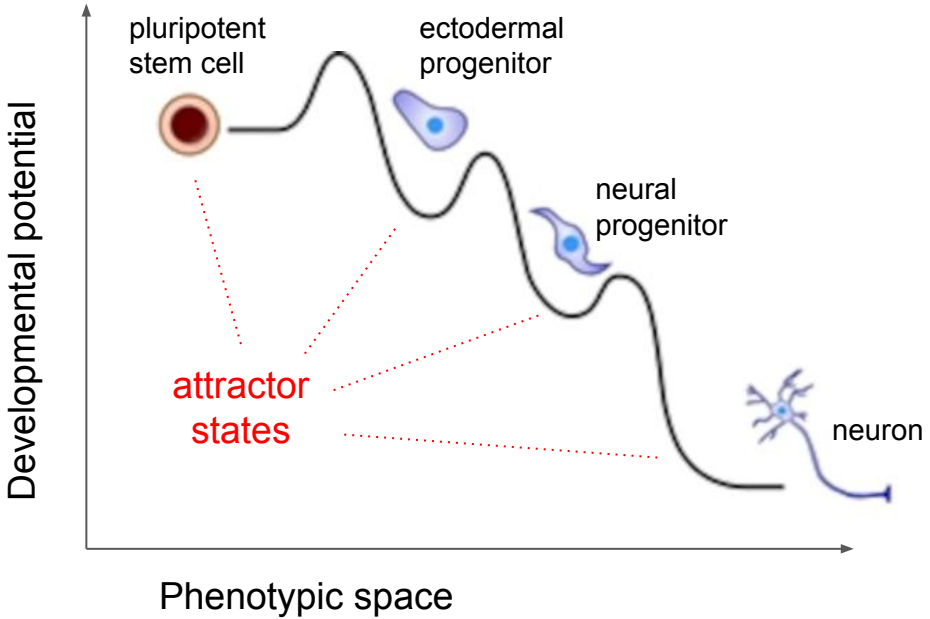


Left from Waddington (1940) "Organisers & Genes"

Right from Waddington (1957) "The strategy of the genes"

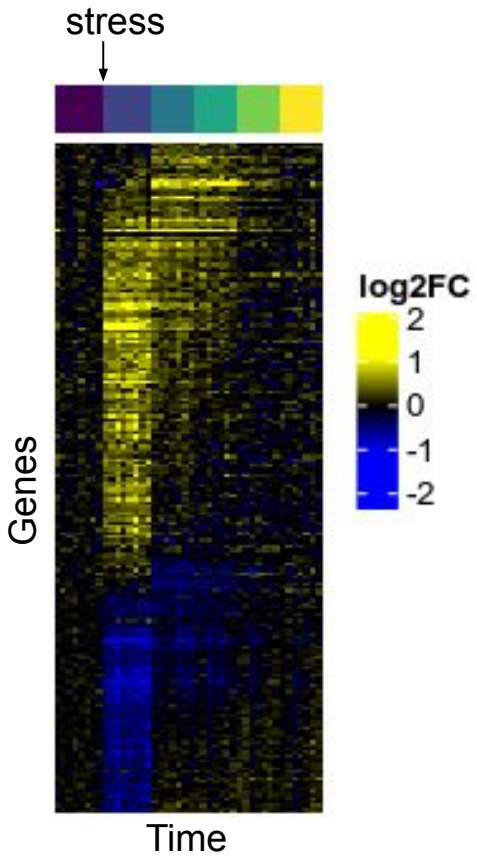
Maintaining and changing cell identity

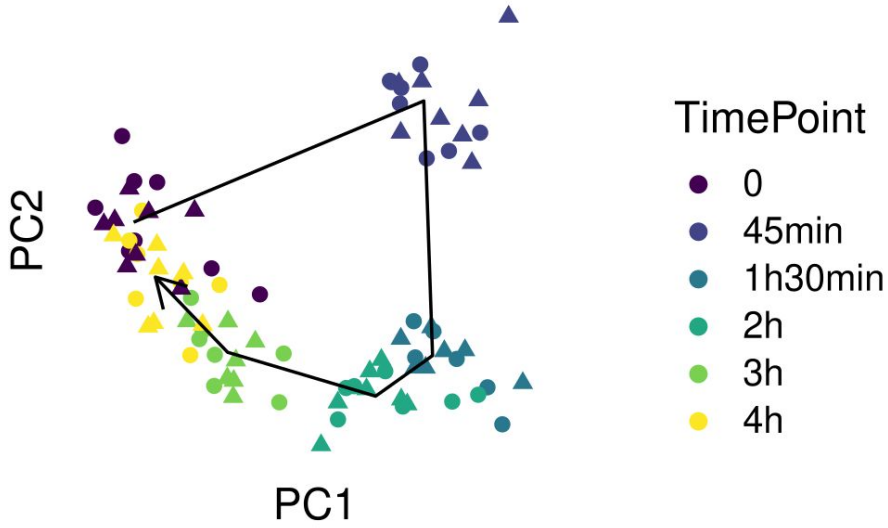




(Adapted from an original figure by KT Rodolfa 2008)

Regulating cell-type-specific responses to stimuli



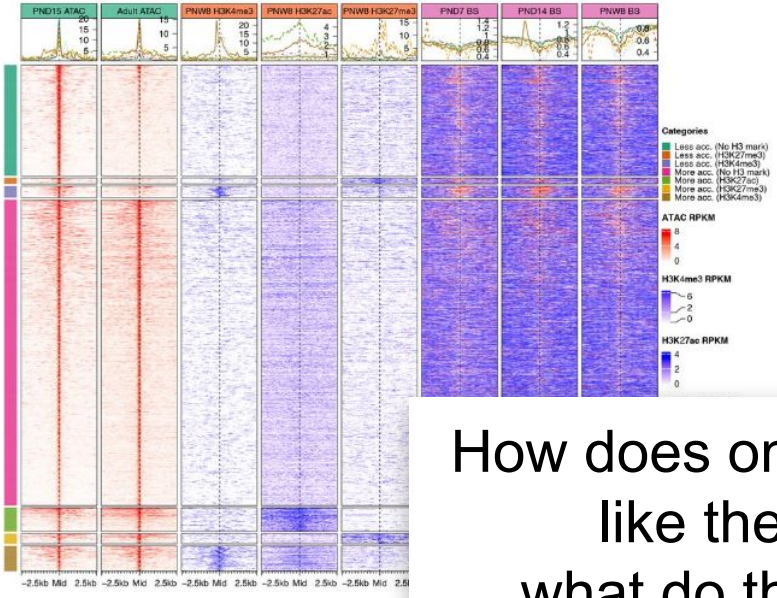


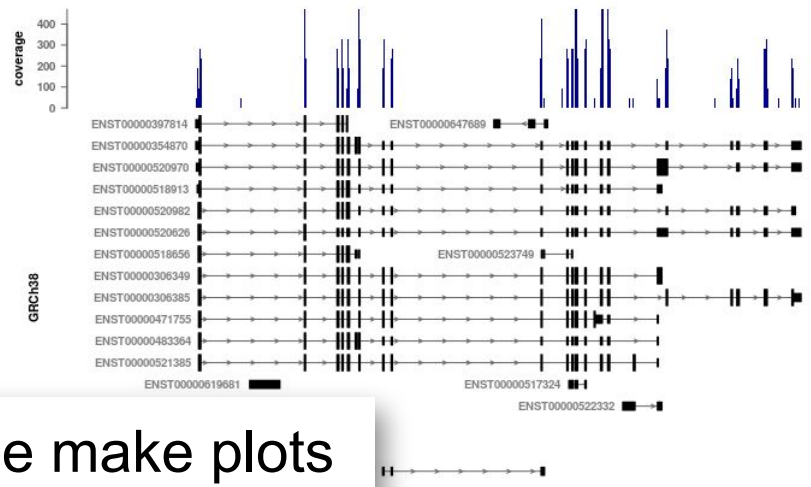
(from a study of the gene expression response, in the hippocampus, to acute stress ; von Ziegler et al., in print)

Aims of the course

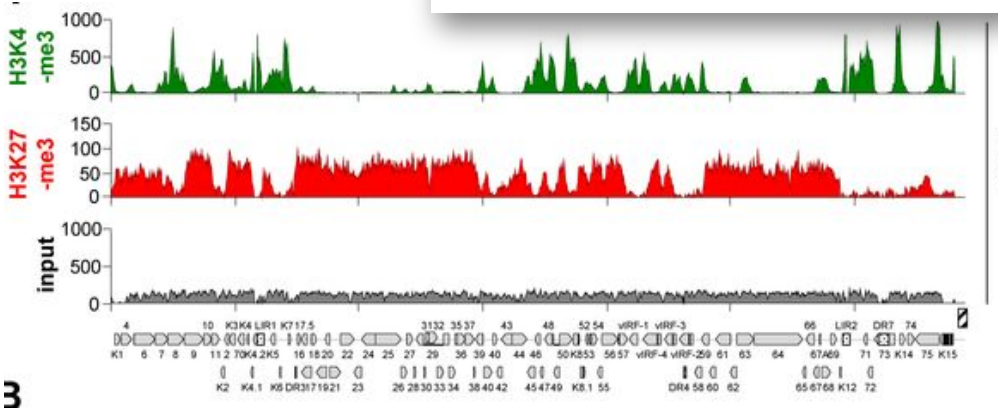
1. to enable students to be, if not wholly independent with respect to epigenomics data analysis, at least able to **autonomously explore, visualize and interpret** such data;

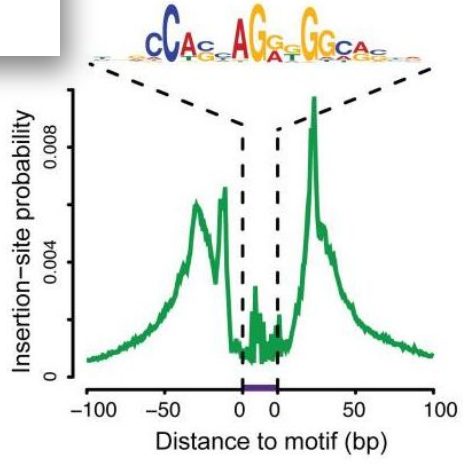
2. to understand and critically appraise, from a genomics perspective and through **hands-on data exploration**, the key **concepts underlying chromatin regulation** of transcription and its impact on various biological phenomena.

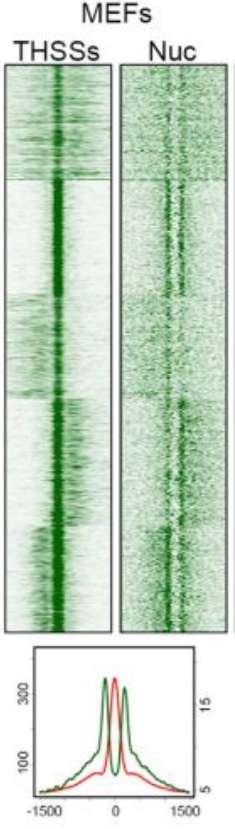




How does one make plots like these, and what do they tell us?







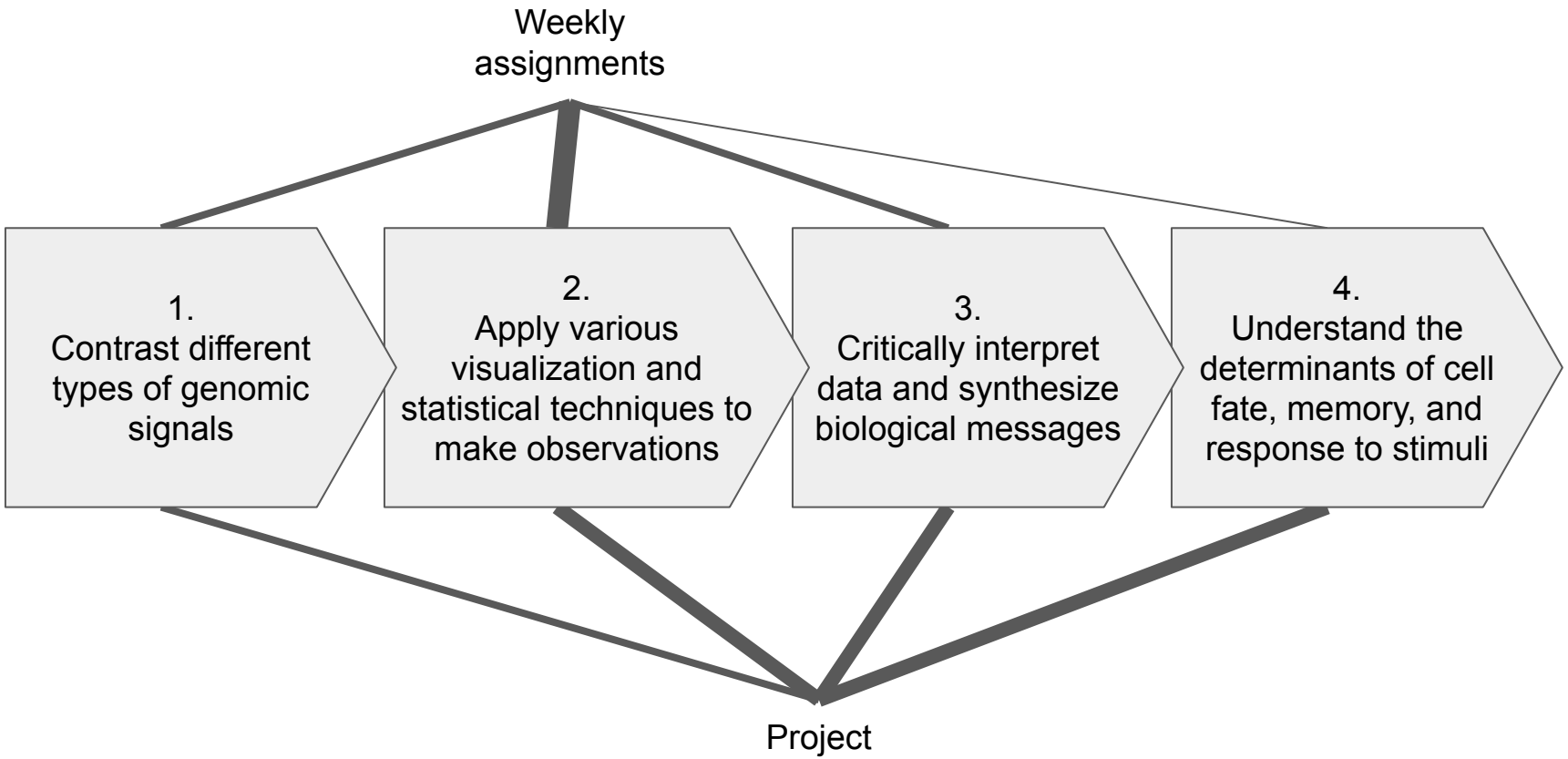
3

Grading and expectations

- 50% of the grade is based on **weekly exercices**
  - Exercices should be **submitted via github**, by thursday the following week
  - The best 8 exercises will make up the grade

- 50% of the grade is based on the **project** (alone or in groups of 2-3)
  - The project can be either:
    - Re-producing the analyses from a publication (in a critical fashion)
    - Analyzing new data (e.g. yours or in collaboration with a group)
  - The project *must be discussed and approved in advance*
  - The expected outputs of the project are:
    - a report (e.g. ~10-15 pages) with embedded full code and figures, and including an introduction and discussion of the results (40% of the grade)
    - a short presentation (10%)

Learning objectives



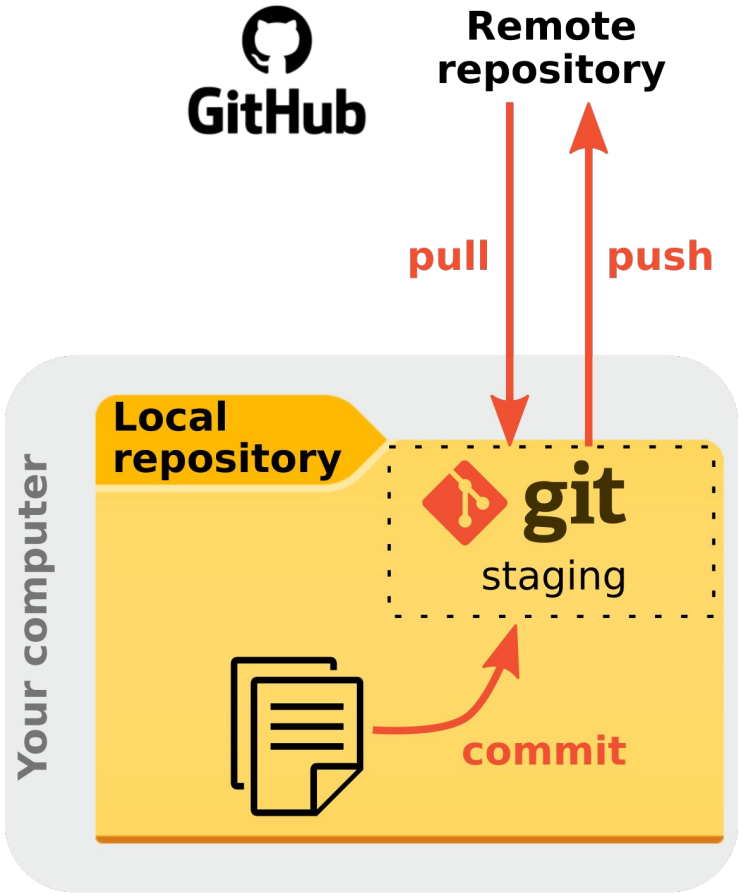
Tentative schedule of the course

Week	Theory	Practice
1	Introduction to the course	Introduction to the practical tools
2	Genome builds, assemblies and annotations	Annotation and feature manipulation in R
3	Overview of NGS chromatin assays and their analysis	Formats & NGS analysis pipeline
4	Types of transcriptional regulators, modes and dynamics of their binding	Analysis and exploration of TF ChIP-seq
5	Functional elements & the histone code	Analysis and exploration of histone ChIP-seq
6	DNA motifs: underpinnings and limitations	Motif scanning and discovery
7	From repression to expression and back	Comparing datasets
8	DNA accessibility, nucleosome positioning, and TF footprints	Analysis and exploration of ATAC-seq data
9	Normalization and differential analysis	Normalization and differential analysis
10	DNA methylation and CpG islands	DNAme visualization; enrichment analysis
11	Chromatin conformation, domains and looping	Identify putative targets of distal regulatory elements
12	Single-cell chromatin assays	Student presentations
13	Chromatin and disease; open questions	Student presentations

The tools we'll be using… and some documentation

- Unix shell environment : short primer / long intro
- R & Bioconductor : primer
- R markdowns : primer
- git & github : tutorials

Introduction to git and github



Task:

- If you don’t already have one, create a github account
- Post your github username on slack
  (I will add you to the course’s github organization)
- Fork the course’s repository, and send me the link to your forked repo

This week's assignment

1. If you haven't done it yet, install R, Rstudio, and Bioconductor

2. Install the packages we'll need (see the `install.R` file on the repo)

3. Create a R markdown in which you:
   a. load the epiwraps library (using `library(epiwraps)`)
   b. print the session info (using `sessionInfo()`)

4. Render your markdown, and push both the Rmd and html files to your github repo.