Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L

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

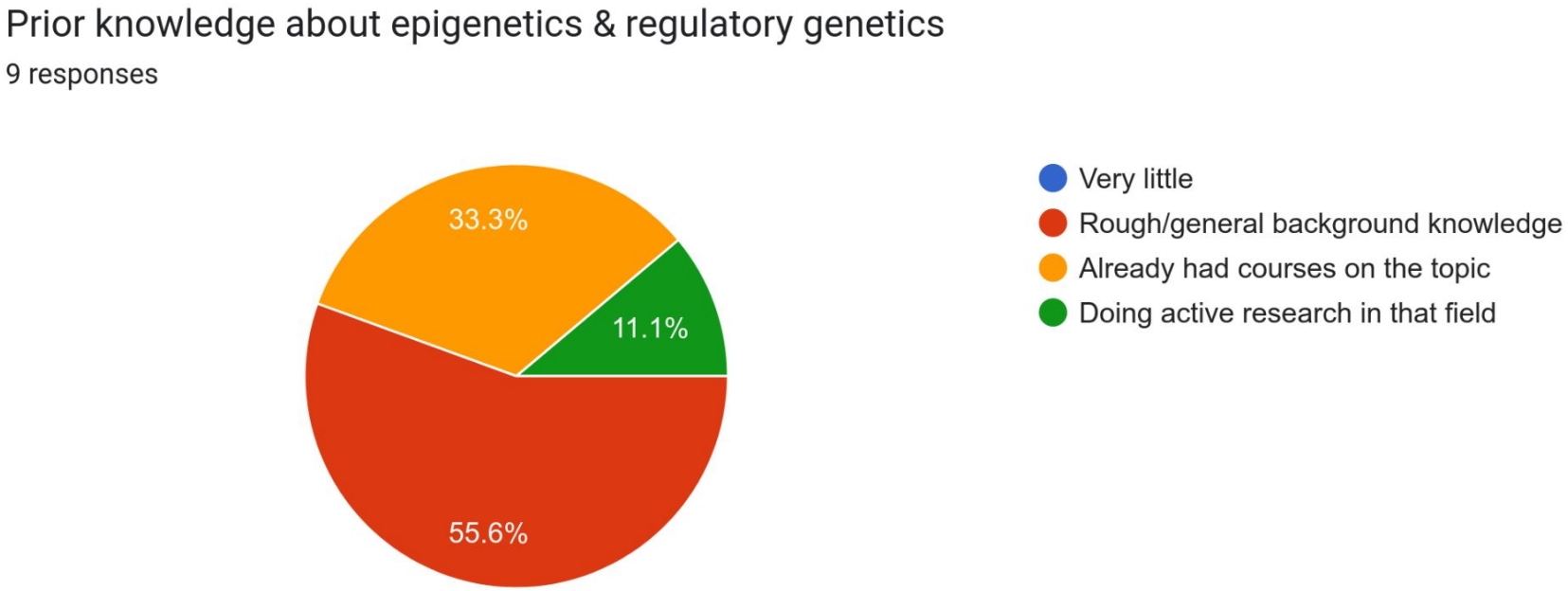
- Introduction to the practical tools for the course:
  - R notebooks, bioconductor, git & github

About your lecturer

- Senior scientist & Junior PI 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, 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 analyzing bulk and single-cell transcriptomics and epigenomics; TF binding & activity inference, reconstruction of transcriptional networks

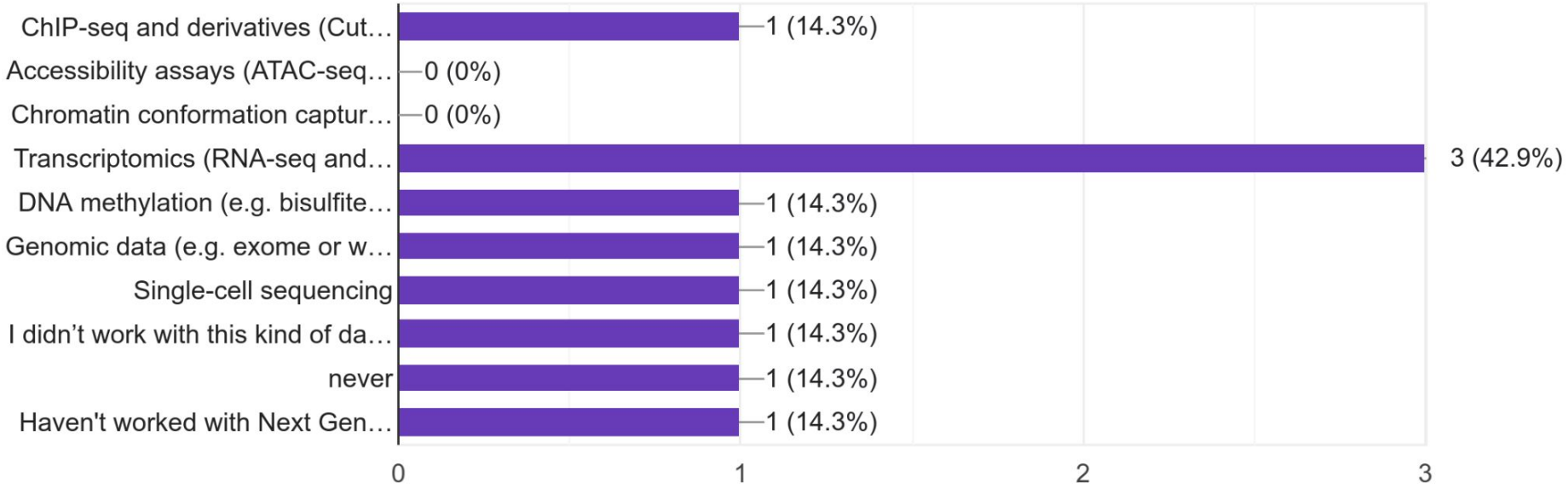
About your fellow students

Chiefly master students, a few doctoral students and others



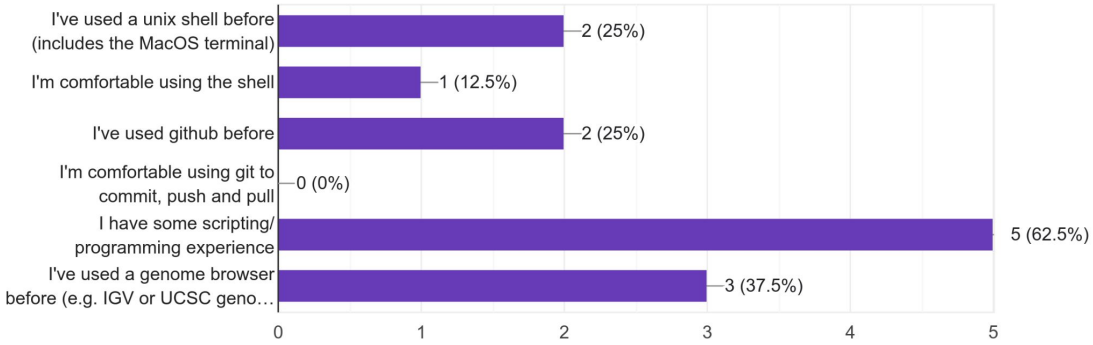
If you have worked with Next Generation Sequencing data before, what kind?

7 responses



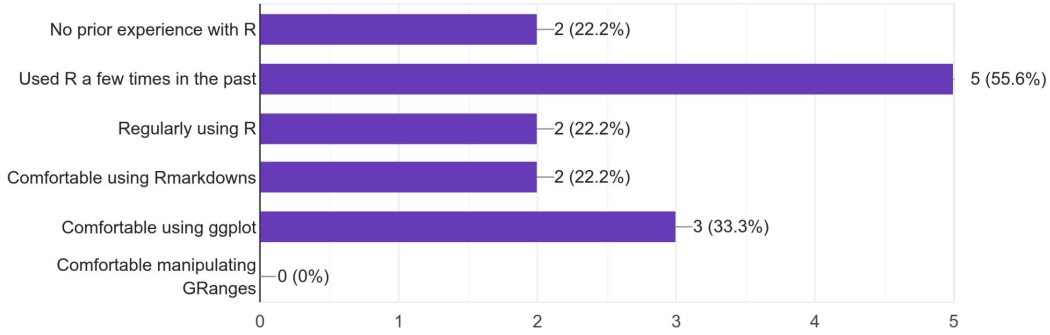
General computer skills; please check all that are relevant

8 responses



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

9 responses



Some expectations given by your fellow students

- “To get a broad and basic introduction to R.”
- “getting better in programming skills”
- “I expect to learn how to use R and how it can be used regarding genetics”

Content	- Introduction to the Bioconductor ecosystem for genomic data analysis, with a focus on flexible and re-usable tools (e.g. standard data structures and visualization techniques) - How to find and re-use data from the literature - Basic analysis pipelines for ATAC-seq, ChIP-seq, and related assays, with a focus on data manipulation, exploration and visualization - Introduction to regulatory genomics, with critical discussions of some of its conceptual issues: types of functional elements, the histone code, sequence recognition and binding specificity, transcriptor factor (TF) activation and binding, DNA accessibility, topological domains and chromatin loops, etc. - Importance of chromatin regulation in differentiation and neurological conditions
Prerequisites / Notice	Familiarity with R (as can be obtained from statistics courses, dedicated R courses, or some bioinformatic courses)

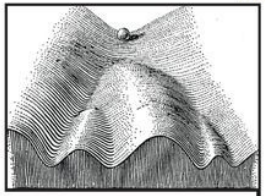
Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L

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)

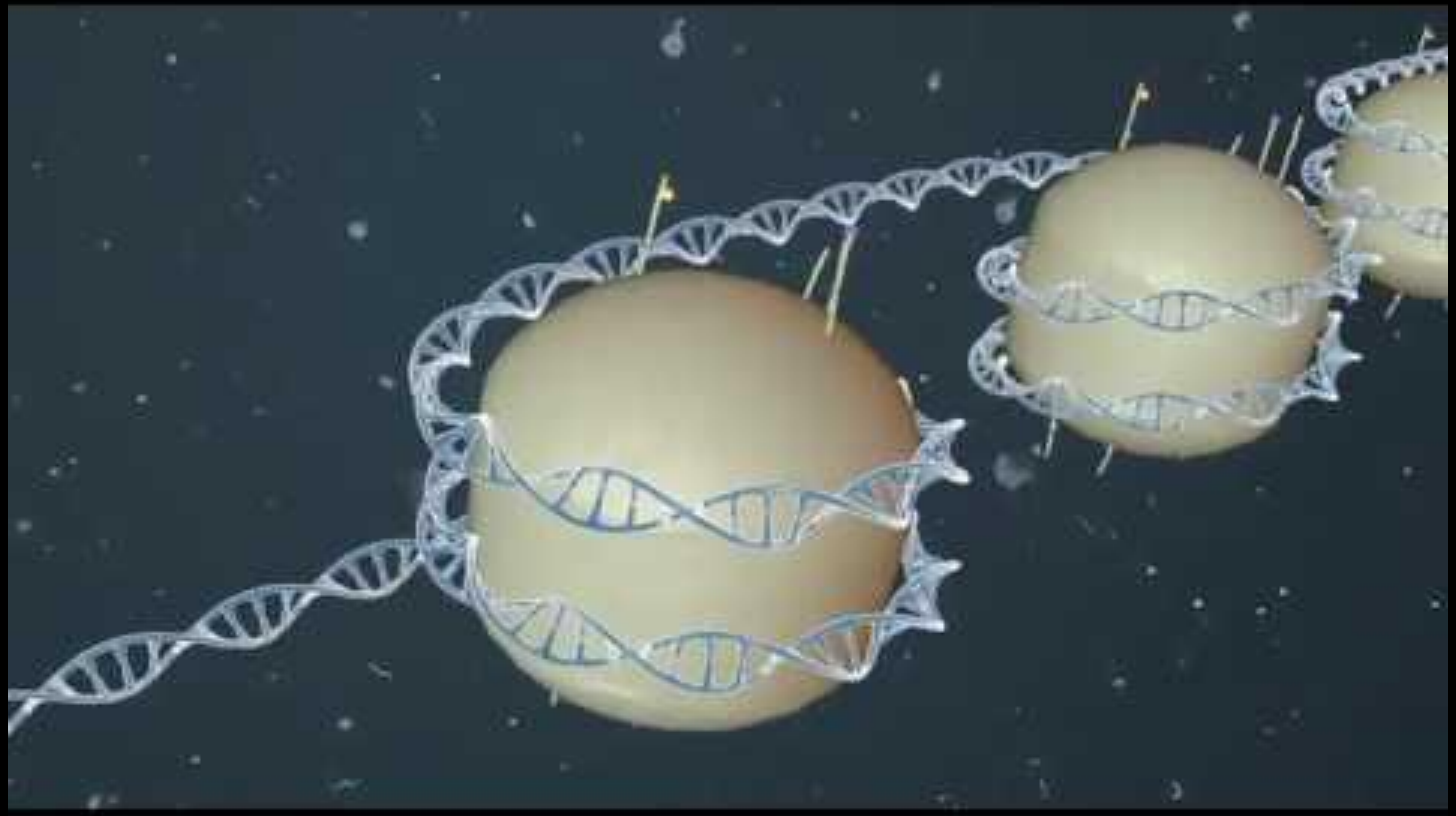
Epigenetic(s)

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

- Ptashne (2007) (also Nanney 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



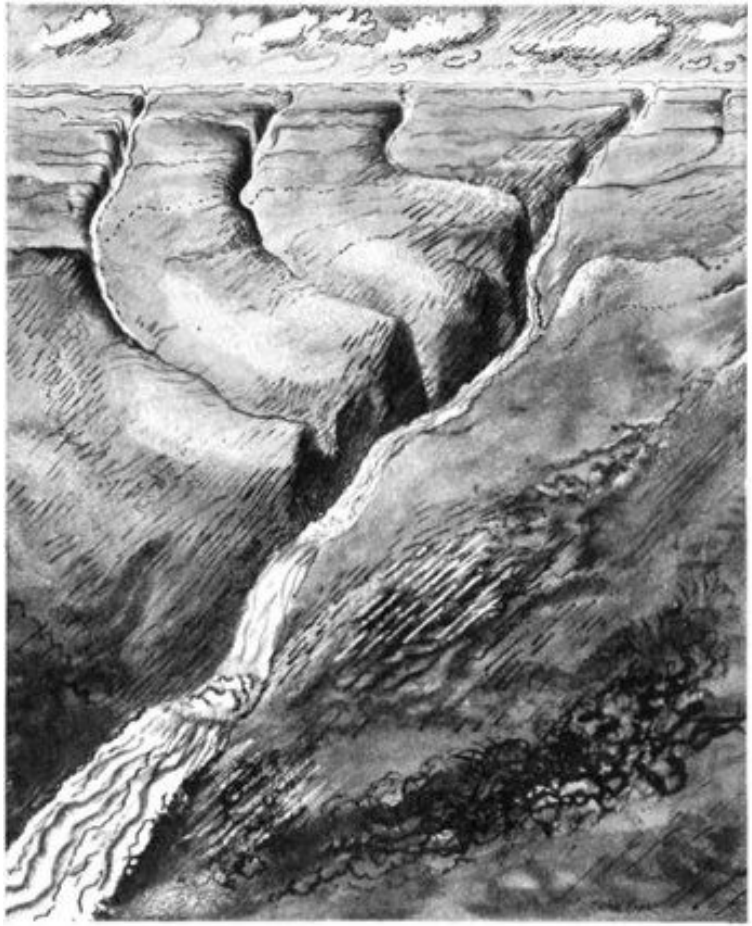
Epigenetic(s)

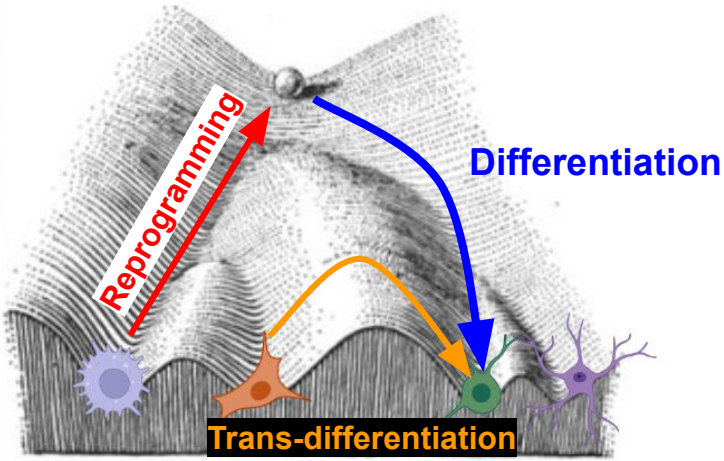
- Waddington (1942) :
  “the causal interactions between genes and their products which bring the phenotype into being”

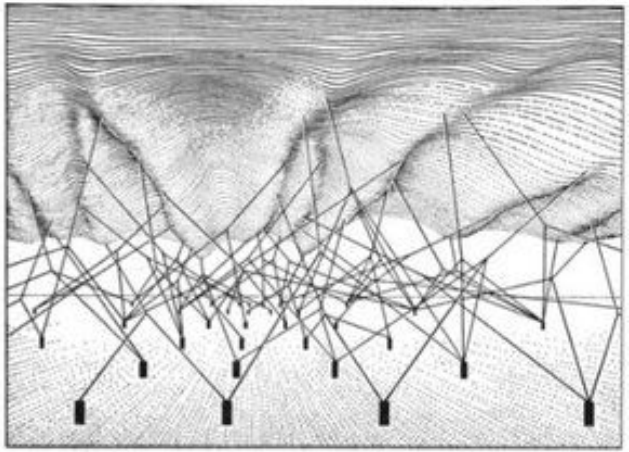
- Ptashne (2007) (also Nanney 1958):
  "A change in the state of expression of a gene that does not involve a mutation, but that is nevertheless **inherited in the absence of the signal** (or event) that initiated the change."

- 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



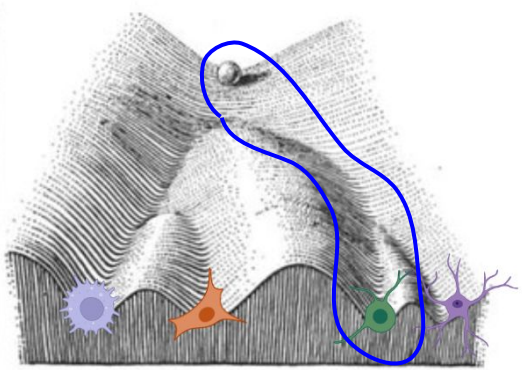


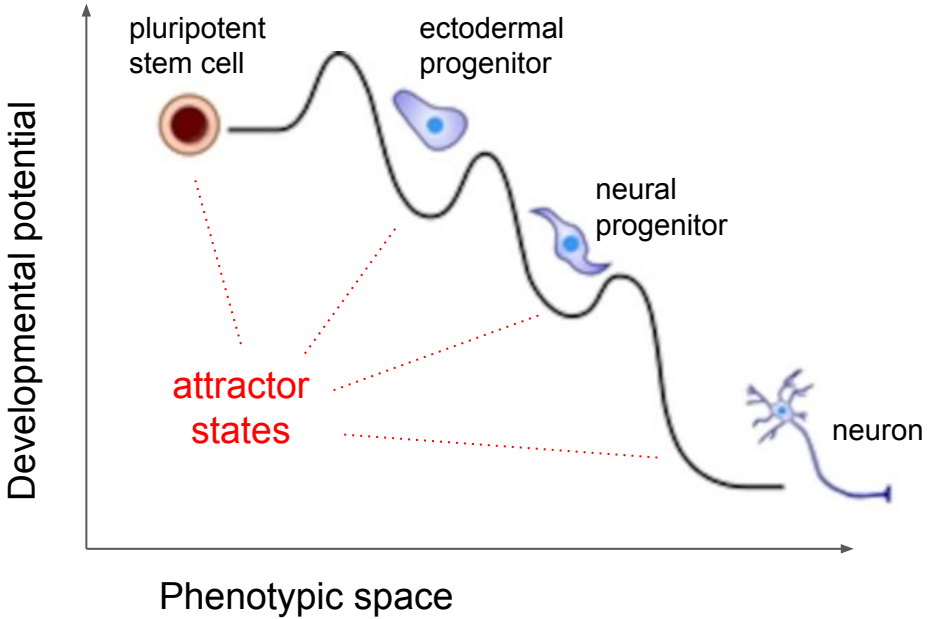


Left from Waddington (1940) "Organisers & Genes"

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

Maintaining and changing cell identity

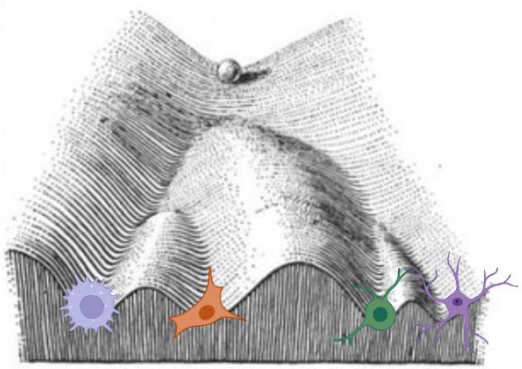


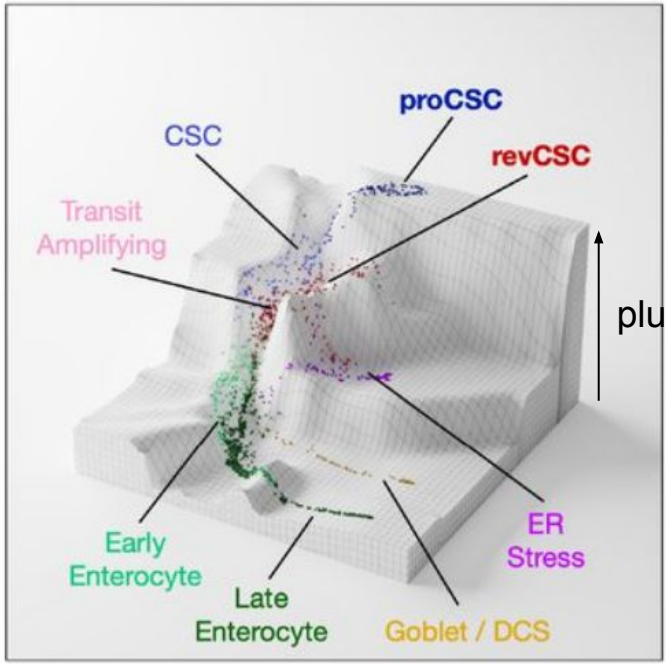


(Adapted from an original figure by KT Rodolfa 2008)

Maintaining and changing cell identity

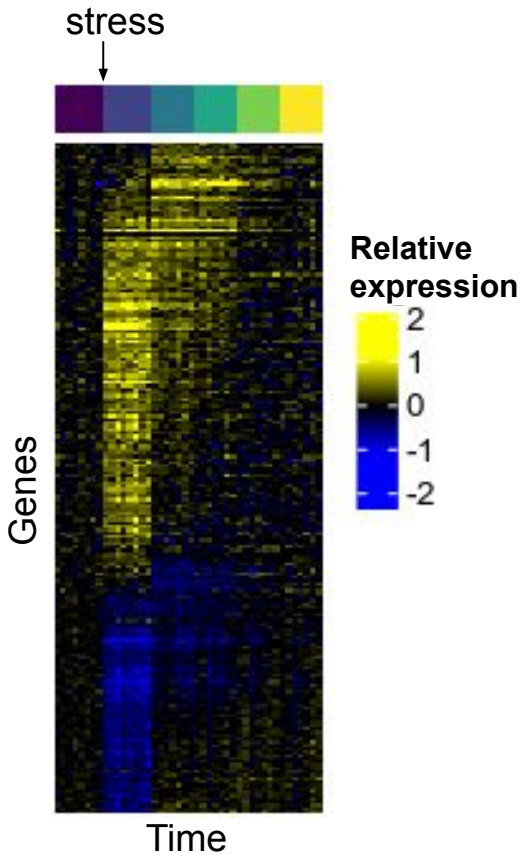
From concept to quantitative measurement

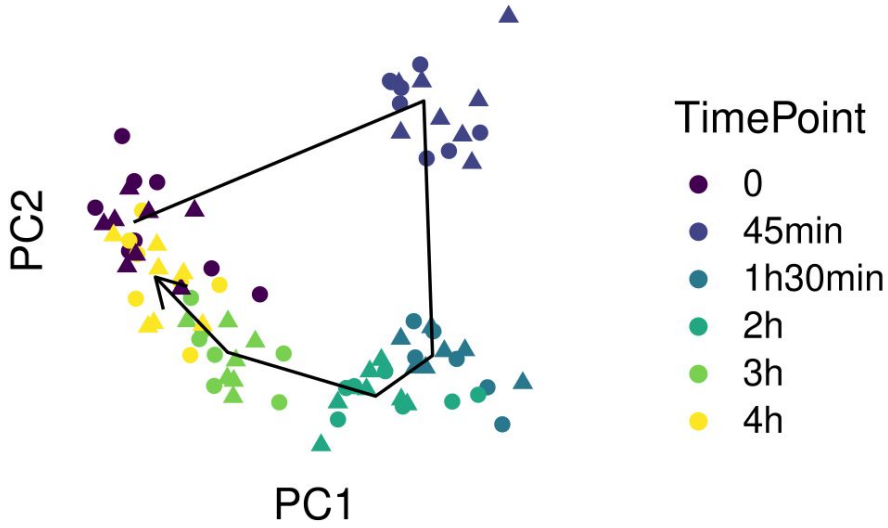




(adapted from
Qin et al., bioRxiv 2023)

Regulating cell-type-specific responses to stimuli

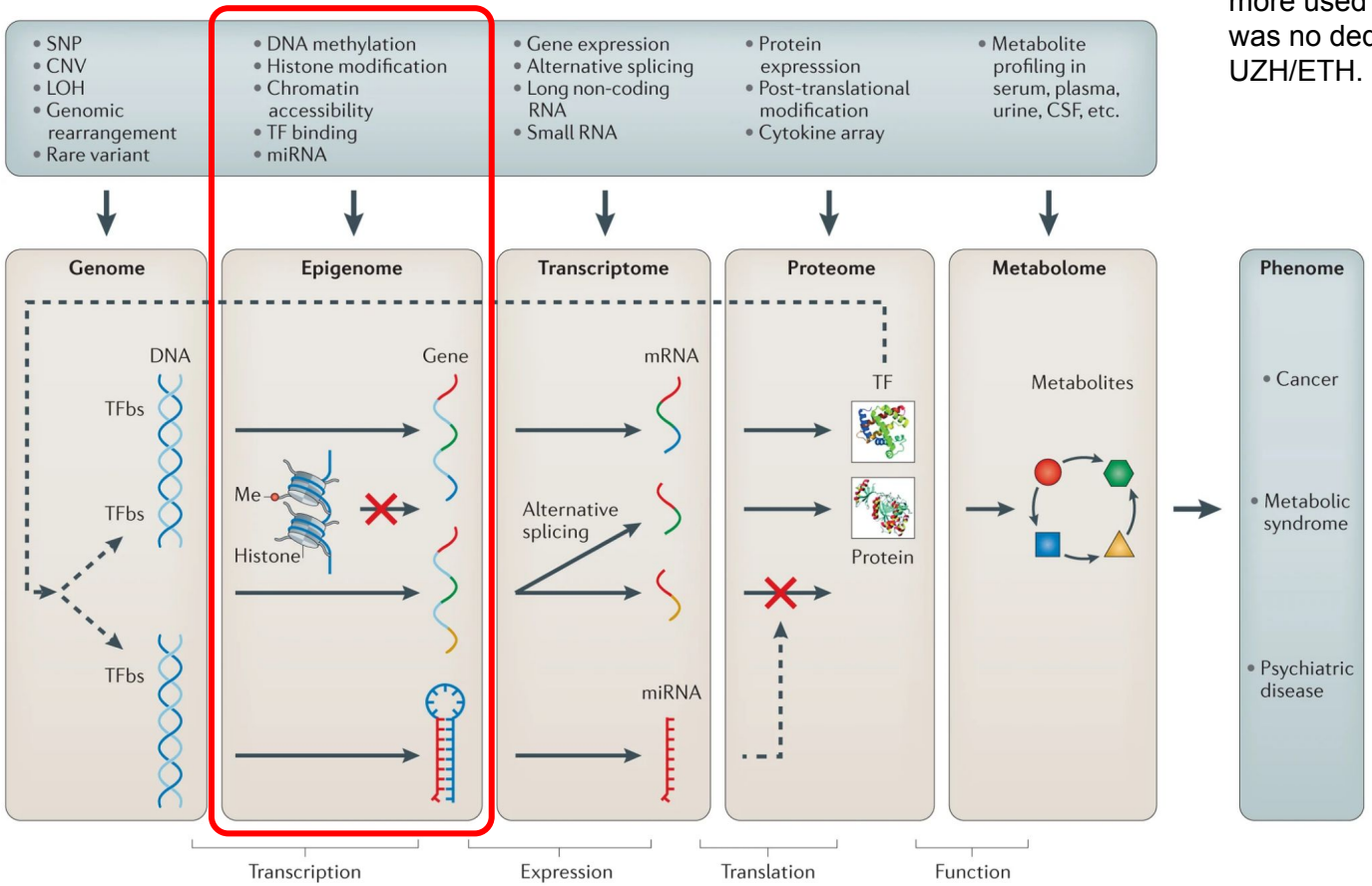




(from a study of the gene expression response,
in the hippocampus, to acute stress ; von Ziegler et al. 2022)

Situating the course

While genomics and transcriptomics are much more used in the clinics, until this course there was no dedicated teaching on epigenomics at UZH/ETH.



(Ritchie et al., Nature Reviews Genetics 2015)

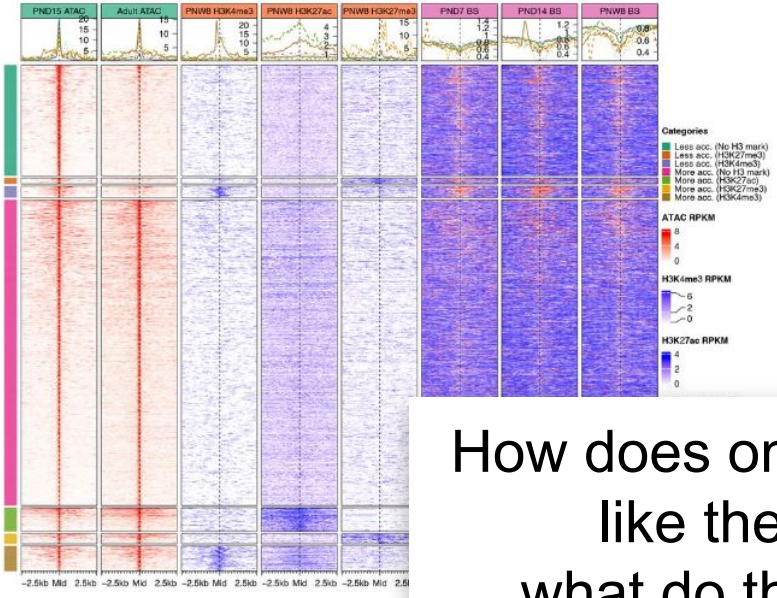
Nature Reviews | Genetics

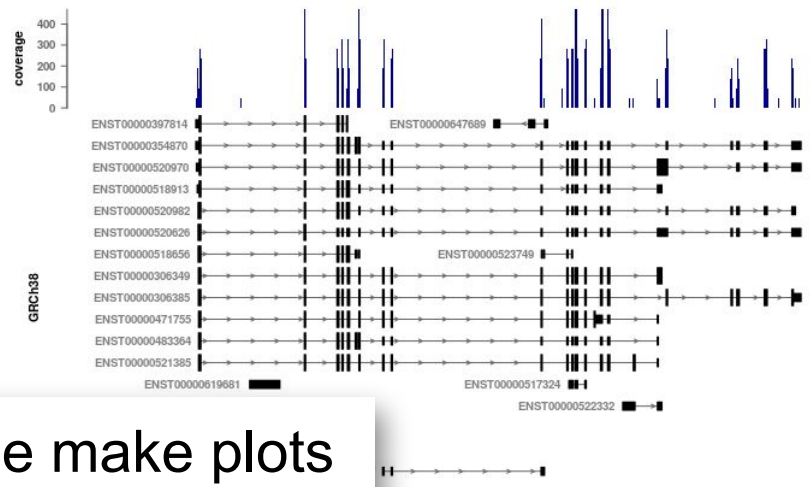
Students interested in transcriptomic analysis may want to follow the excellent 401-6282-00L (UZH Sta426) course (emphasis on stats)

Aims of the course

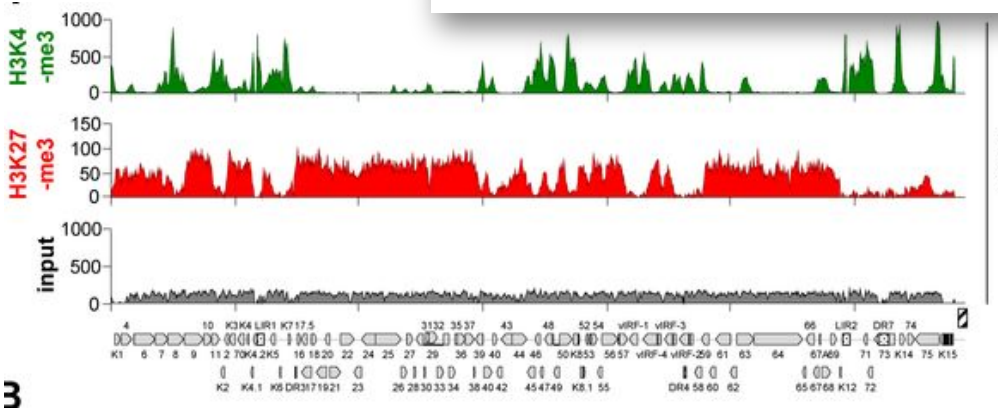
1. to enable students to be, if not wholly independent with respect to epigenomics data analysis, at least able to **autonomously explore, visualize and interpret** such data;

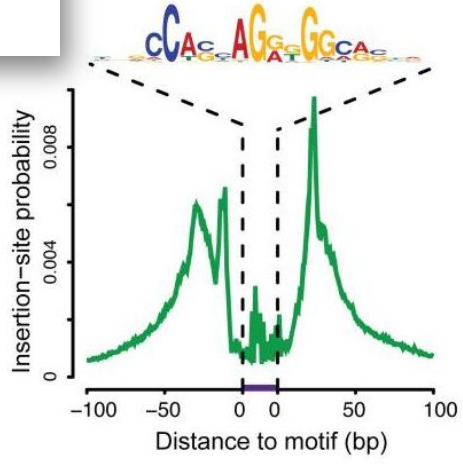
2. to understand and critically appraise, from a genomics perspective and through **hands-on data exploration**, the key **concepts underlying chromatin regulation** of transcription and its impact on various biological phenomena.

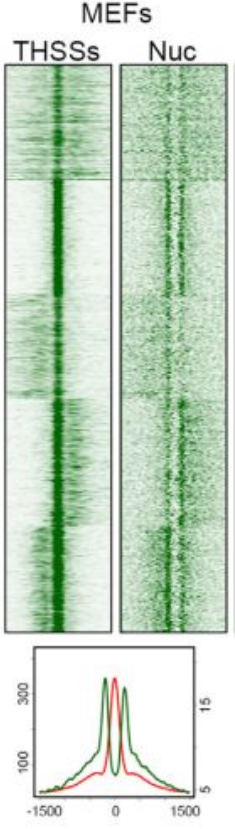




How does one make plots like these, and what do they tell us?







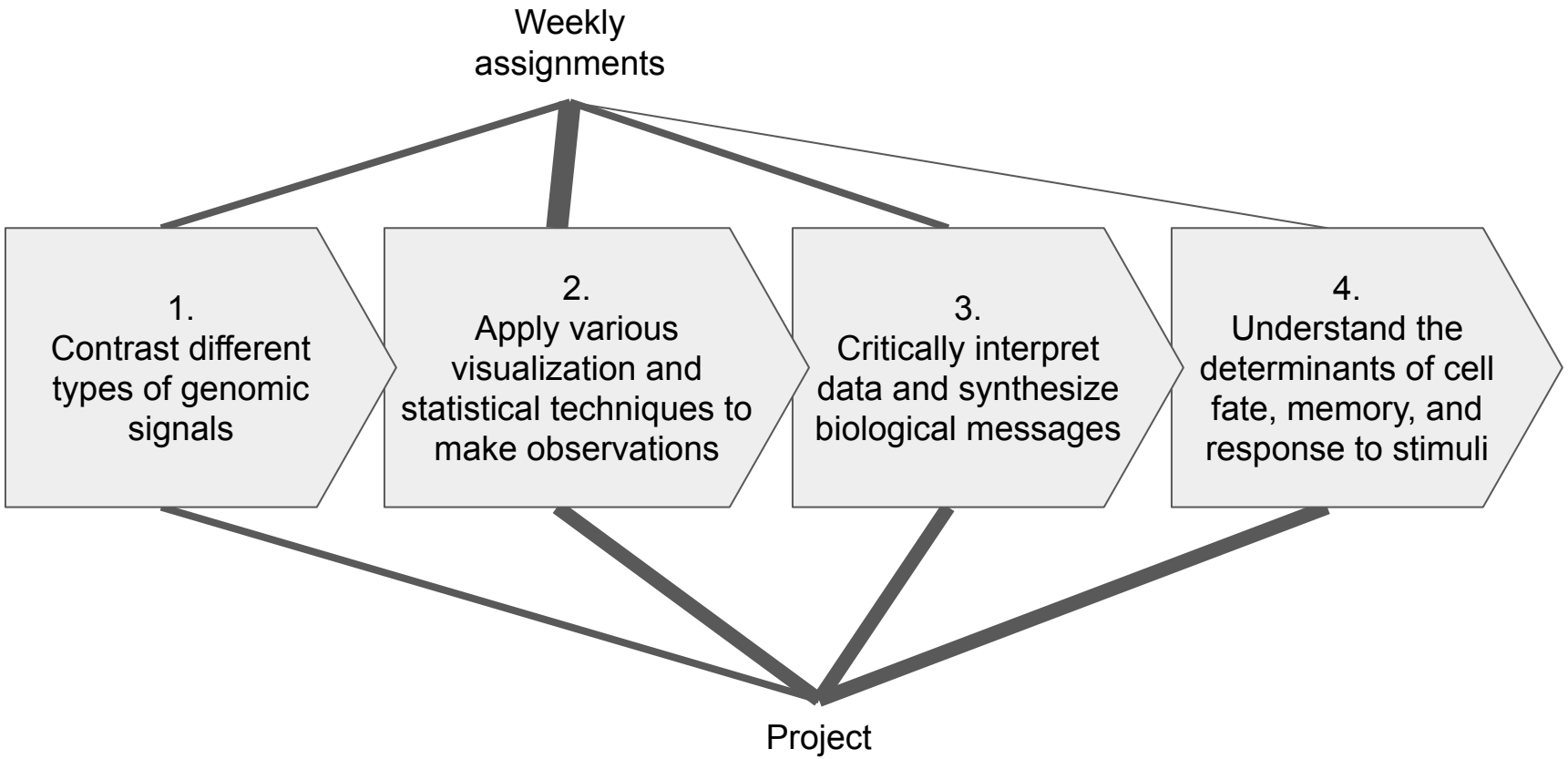
3

Grading and expectations

- 50% of the grade is based on **weekly exercices**
  - Exercices should be **submitted via github**, by thursday noon the following week
  - The best half of the 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

Learning objectives



Tentative schedule of the course

Each week:

- **~5-10min of debriefing** on the exercises from the previous week
- **~25-30min of "lecture"** to provide background and theory
- **~60min of hands-on** demonstration (live coding)

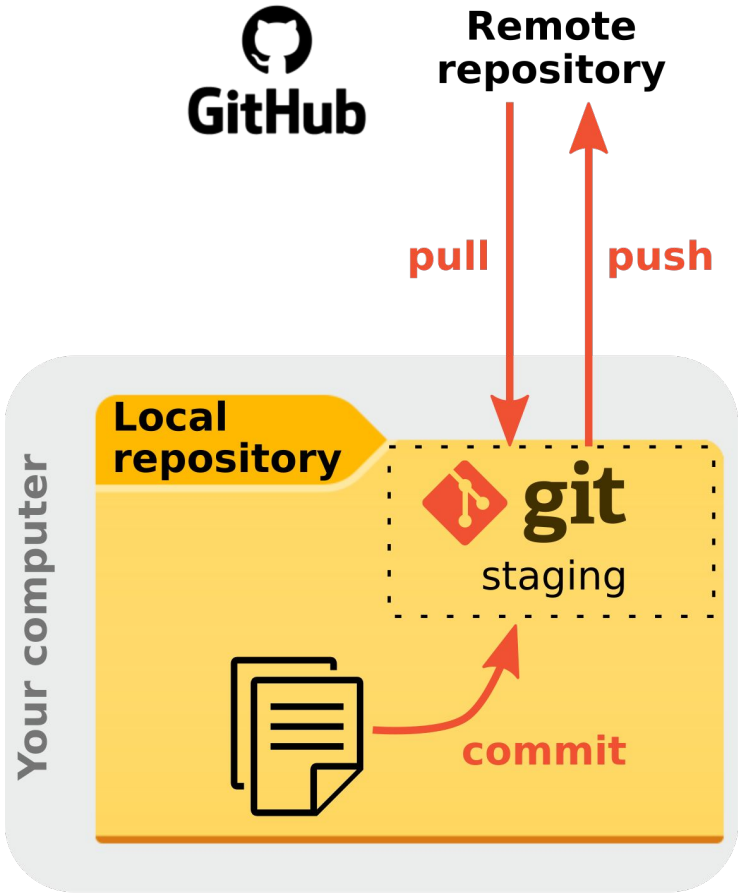
W	Date	Lecture (~30min)	Mixed/hands-on (~60min)
1	23/02	Introduction to the course	Introduction to the practical tools for the course
2	01/03	Genome builds, transcriptome assemblies and annotations	AnnotationHub, Genomic Ranges, Features and their manipulating in R
3	08/03	Overview of NGS chromatin assays and their analysis	Primary ChIPseq analysis pipeline
4	15/03	Specificities and QCs on primary analysis	Finding data from the literature
5	22/03	Functional elements & the histone code	Visualization and exploration of histone modifications
*	29/03	###### Easter - No Course ######	
*	05/04	###### Easter - No Course ######	
6	12/04	Types of transcriptional regulators, modes and dynamics of their binding	Motif analysis
7	19/04	DNA accessibility, nucleosome positioning, and TF footprints	Analysis and exploration of ATAC-seq data
8	26/04	Differential analysis	Differential analysis
9	03/05	Normalization and enrichment analysis	Normalization and enrichment analysis
10	10/05	From repression to expression and back	Clustering genomic signals and characterizing the clusters
11	17/05	DNA methylation and CpG islands	DNAme visualization
12	24/05	Chromatin conformation, domains and looping	Working with distal regulatory elements
13	31/05	Single-cell chromatin assays; chromatin and disease; open questions	

The tools we'll be using… and some documentation

- Unix shell environment : short tutorials / long intro
- R & Bioconductor : primer
  - (installation instructions for R, rstudio and bioconductor)
- R markdowns : primer
- git & github : tutorials

Particularly important for the course

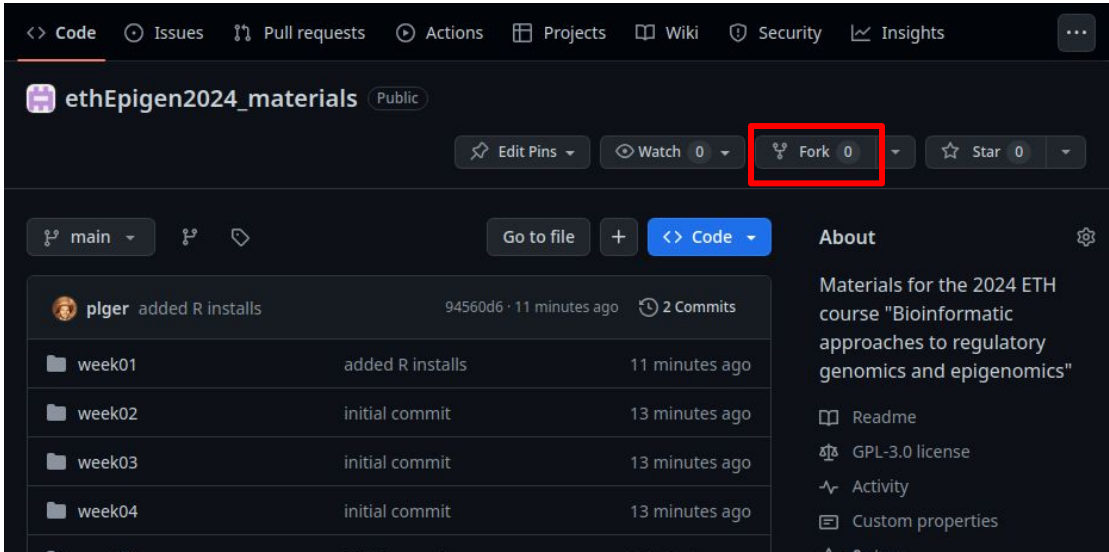
Introduction to git and github



Setting up your github repository for the course:

- If you don't already have one, create a github account
- Send us your github username on slack
- Fork the course's repository, and send me the link to your forked repo

To fork, go to https://github.com/ETHZ-INS/ethEpigen2024_materials , make sure you are logged into github, and click the "Fork" button



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 named "`assignment.Rmd`" 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, in the directory `week01/`