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 & student expectations
- 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; genetics of neurodevelopmental disorders; brain and cellular responses to stress

- 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 (HEST) master students, a few doctoral students and others

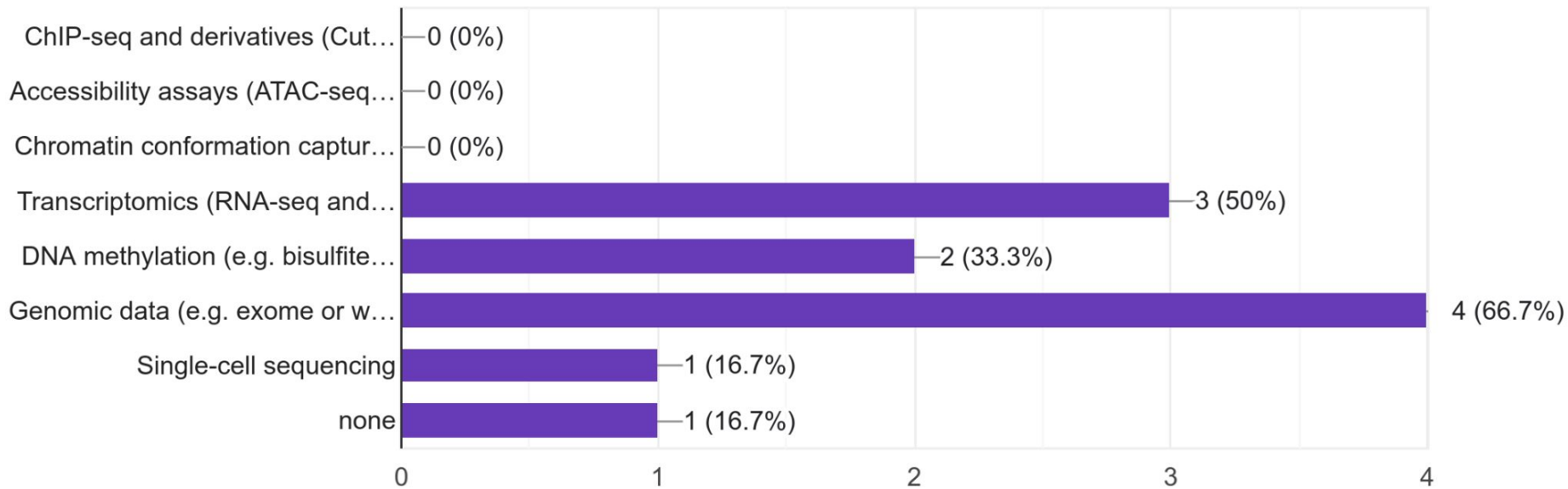Prior knowledge about epigenetics & regulatory genetics

17 responses
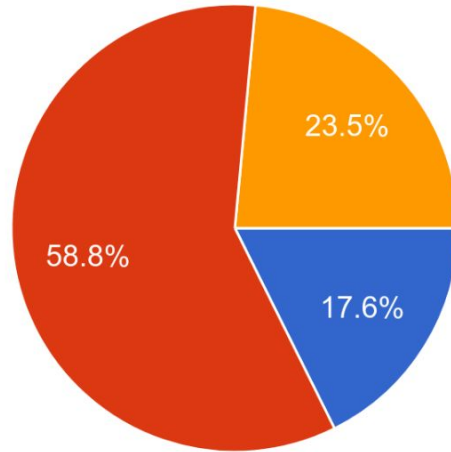
- Very little
- Rough/general background knowledge
- Already had courses on the topic
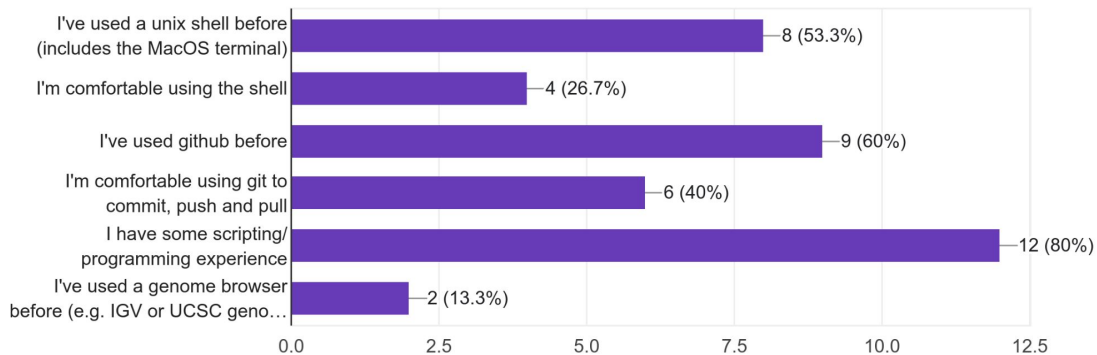- Doing active research in that field

23.5%

58.8%

17.6%

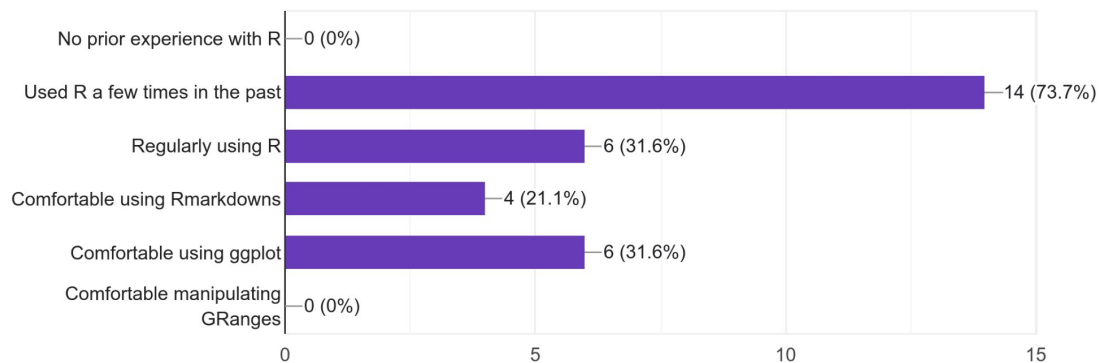# If you have worked with Next Generation Sequencing data before, what kind?

6 responses

General computer skills; please check all that are relevant

15 responses

Category	Count
I've used a unix shell before (includes the MacOS terminal)	8 (53.3%)
I'm comfortable using the shell	4 (26.7%)
I've used github before	9 (60%)
I'm comfortable using git to commit, push and pull	6 (40%)
I have some scripting/ programming experience	12 (80%)
I've used a genome browser before (e.g. IGV or UCSC geno…)	2 (13.3%)

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

19 responses

Category	Count
No prior experience with R	0 (0%)
Used R a few times in the past	14 (73.7%)
Regularly using R	6 (31.6%)
Comfortable using Rmarkdowns	4 (21.1%)
Comfortable using ggplot	6 (31.6%)
Comfortable manipulating GRanges	0 (0%)

Some expectations given by your fellow students

- "Familiarization with data formats. Multi-omics Integration. Data analysis…"

- "...understand what type of analysis is suited for what specific type of data and why…"

- "Improve my proficiency with R coding language and being able to conduct data analysis of epigenetic data from start to finish."

- "I took the course Introduction to Data Science (for HST students) which I really enjoyed. So I would like to learn more about how to combine data science and biology."

| 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)
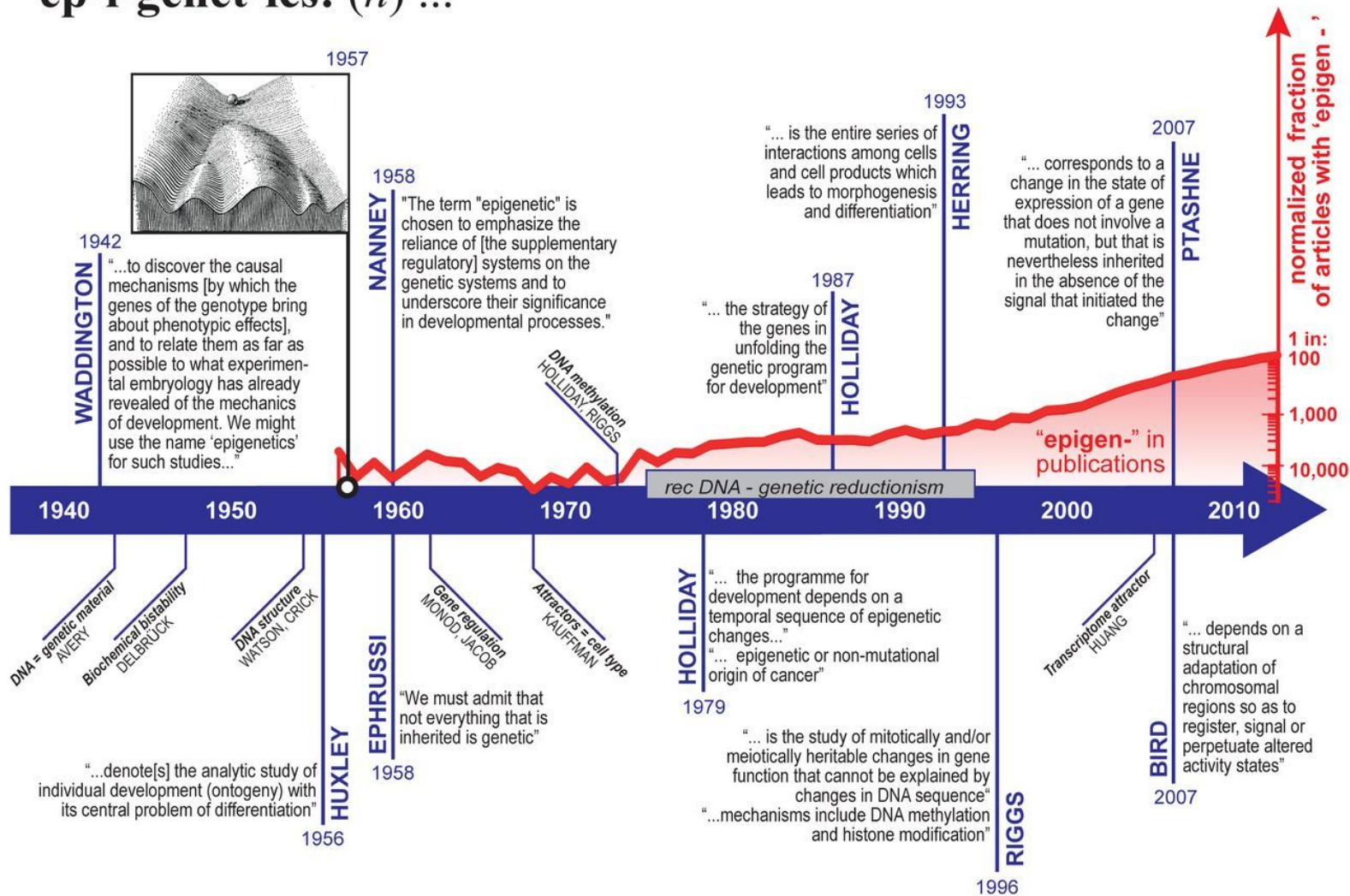
ep•i•genet•ics: (n) ...

(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"
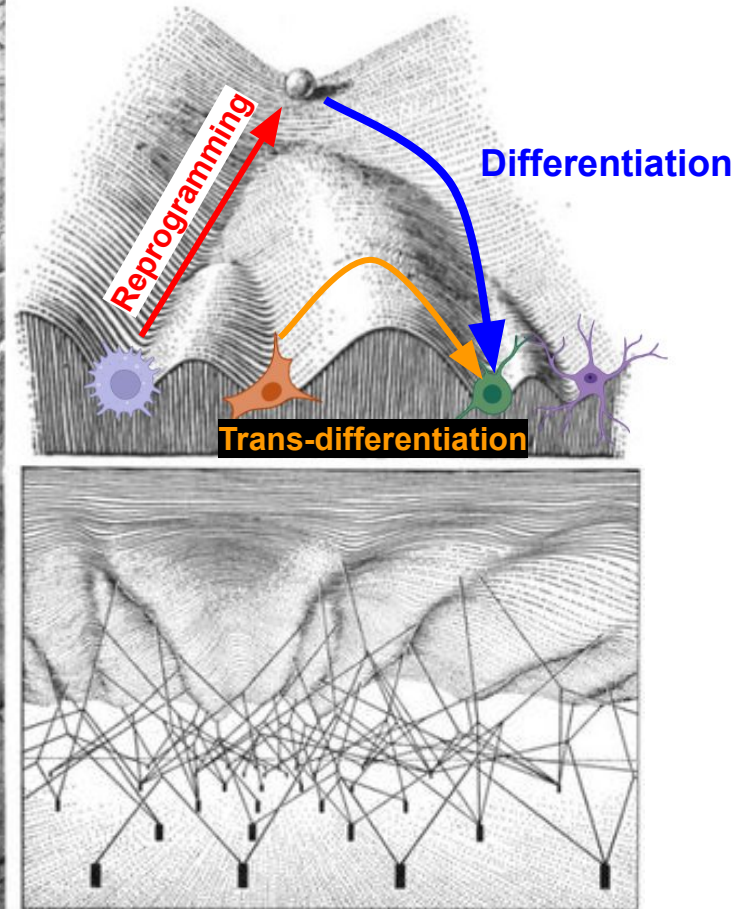
Basic primer on epigenetics
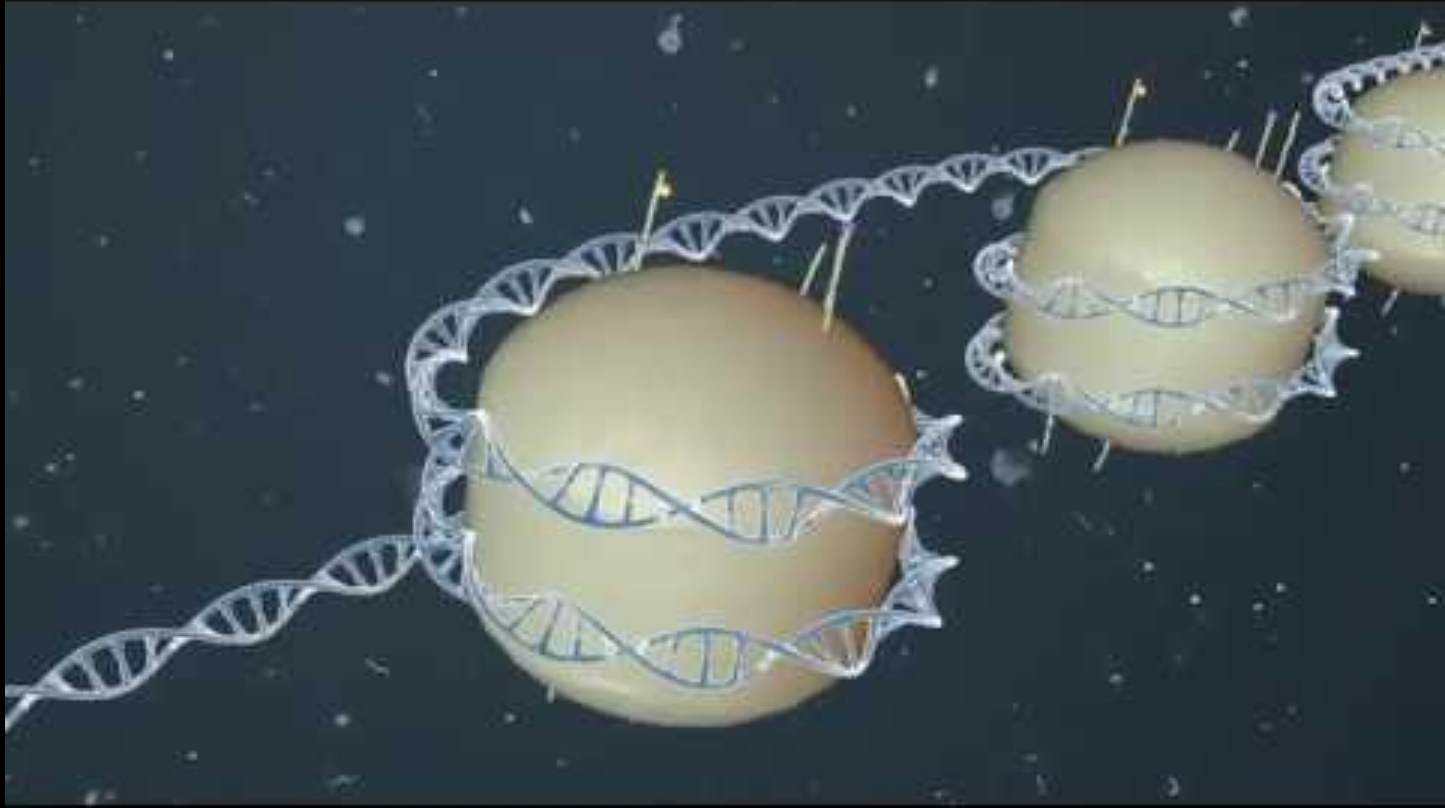
Epigenetic(s)

- 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"
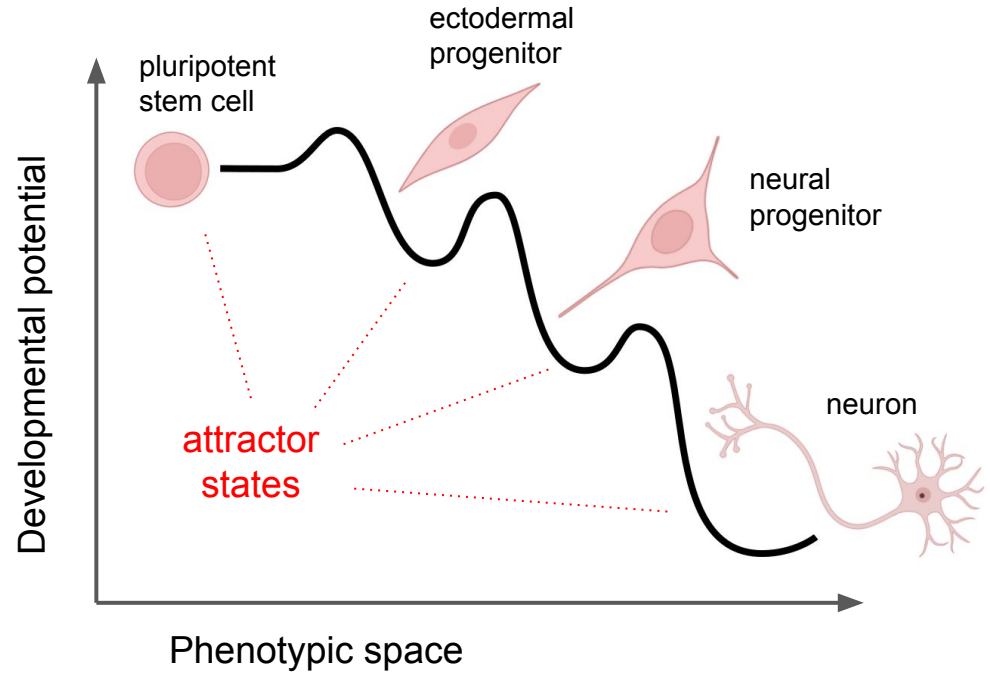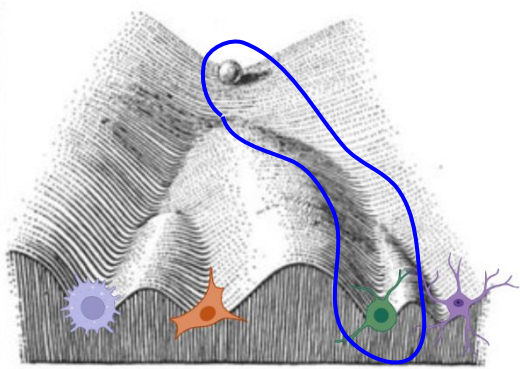
Waddington's epigenetic landscape

Left: from Waddington (1940) "Organisers & Genes"   Right: from Waddington (1957) "The strategy of the genes"
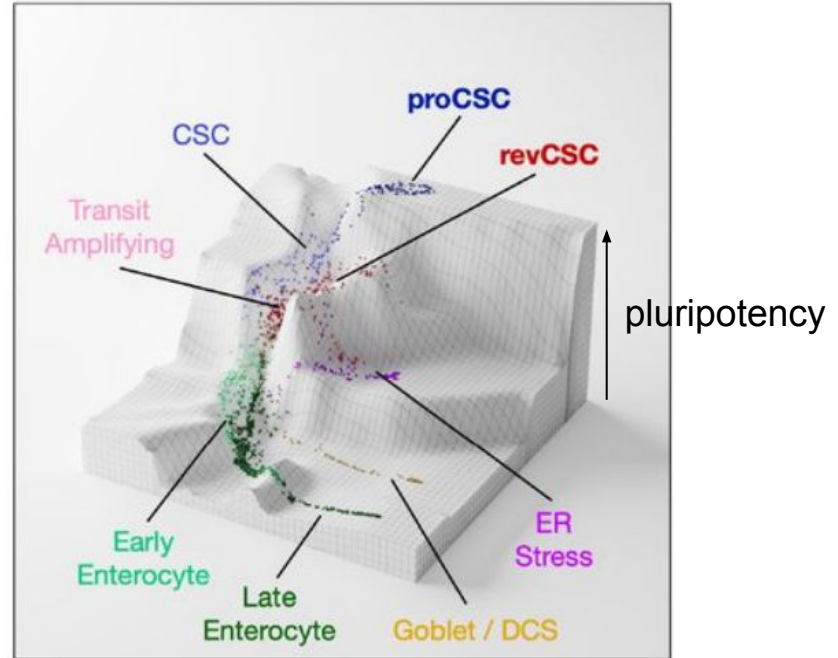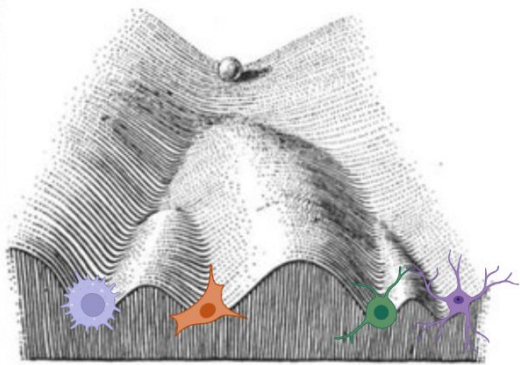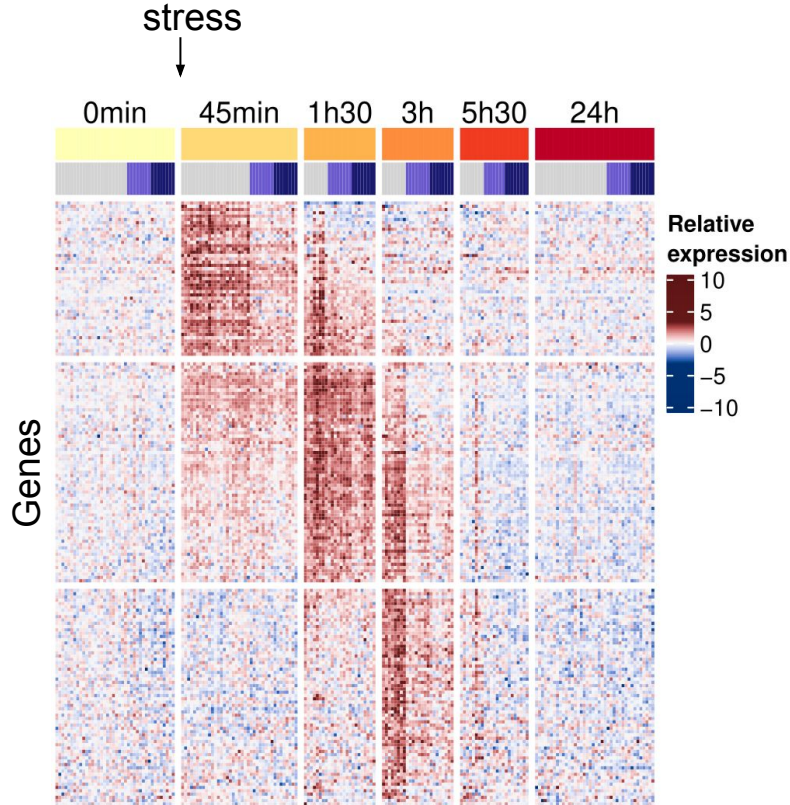
Maintaining and changing cell identity

Maintaining and changing cell identity

From concept to quantitative measurement

pluripotency

(adapted from
Qin et al., bioRxiv 2023)

# Regulating cell-type-specific responses to stimuli

What are the mechanisms behind the cascades of transcriptional changes following stimulation?

(from a study of the hippocampal gene expression response to acute and chronic stress)

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.
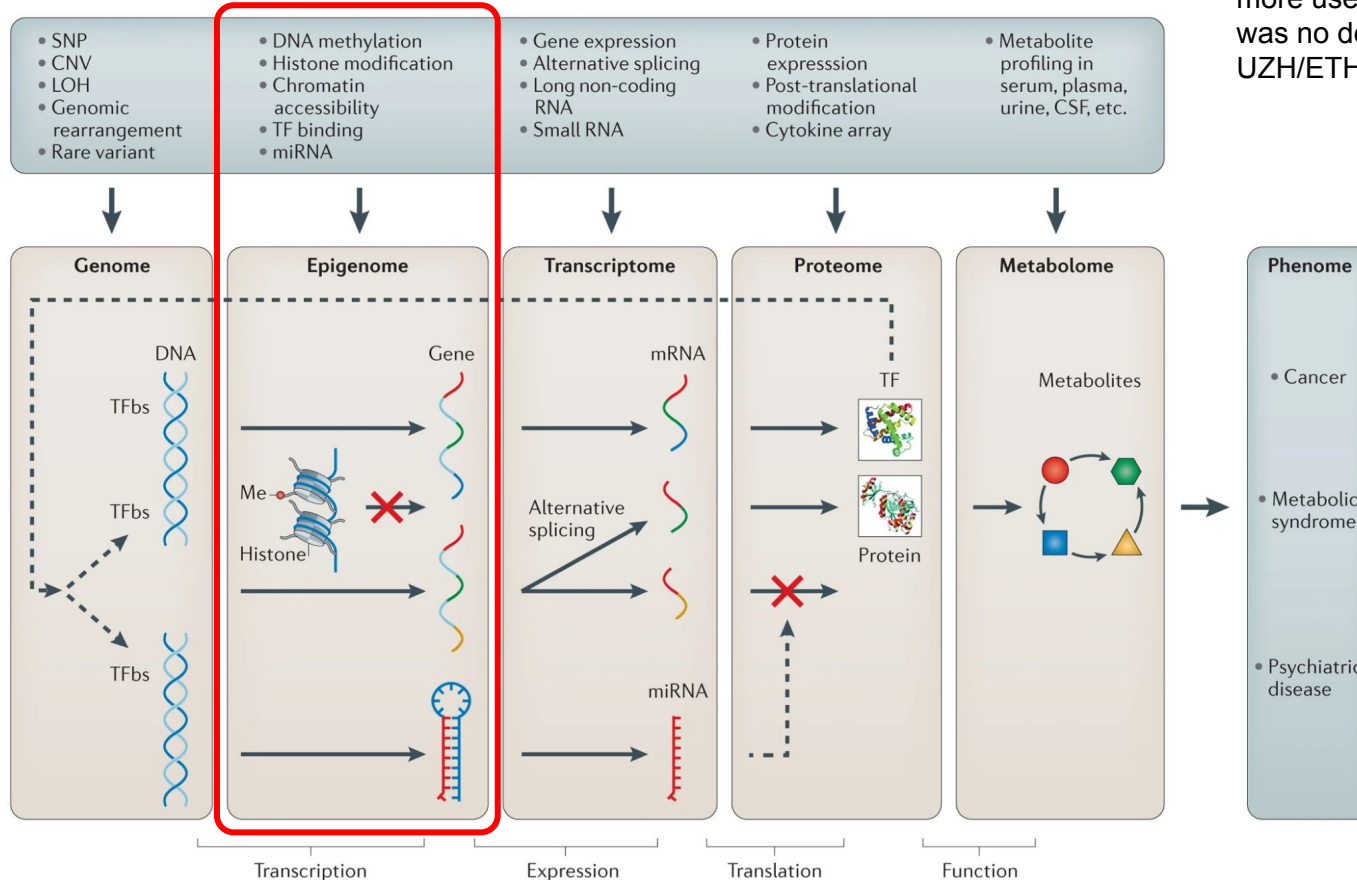
Students interested in transcriptomic analysis may want to follow the excellent 401-6282-00L (UZH Sta426) course (emphasis on stats); for single-cell data analysis, see the 636-0121-00G course.

(Ritchie et al., Nature Reviews Genetics 2015)

Nature Reviews | Genetics

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?
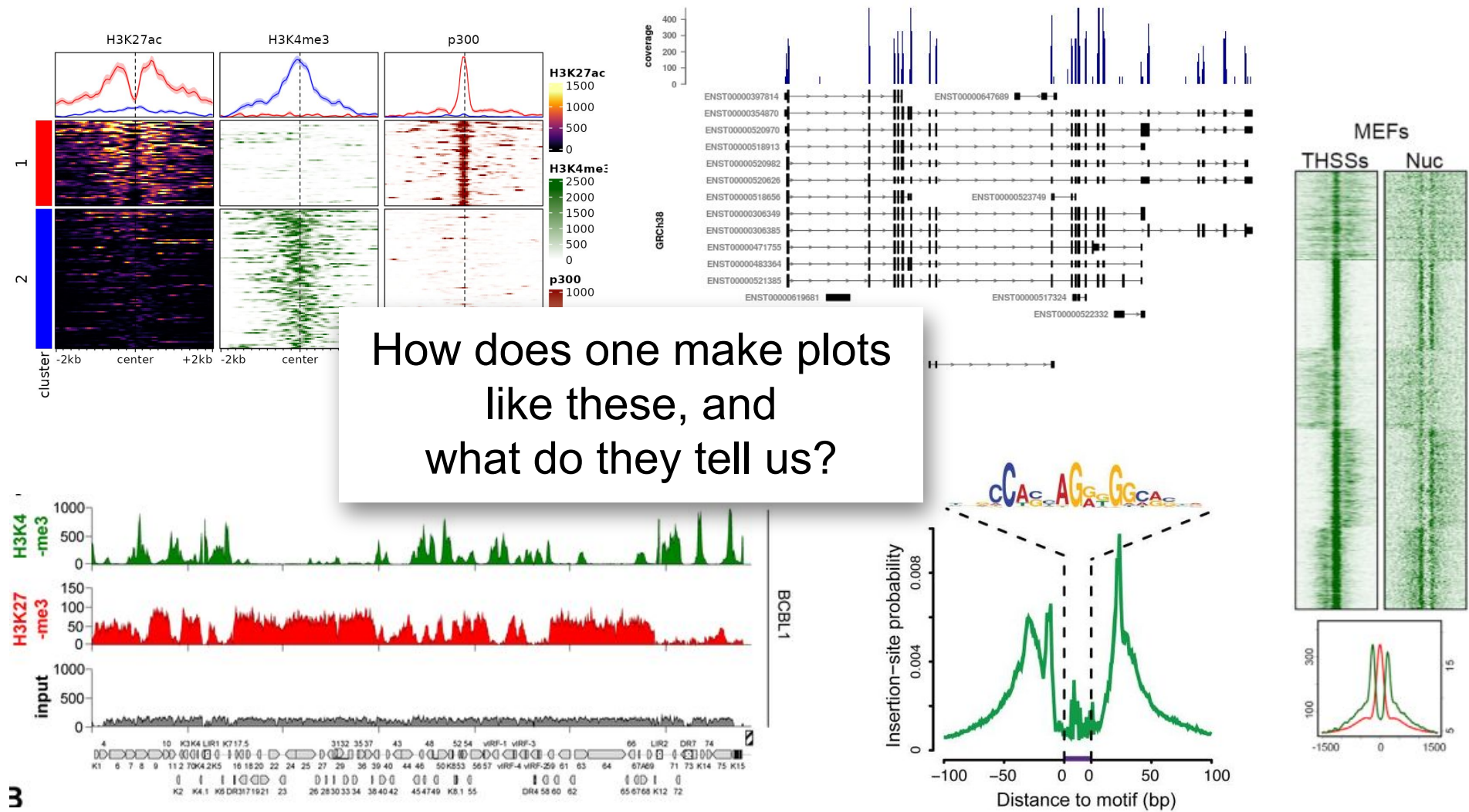
Grading and expectations

- 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 2-4)
  - 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

Weekly assignments

1.
Contrast different types of genomic signals

2.
Apply various visualization and statistical techniques to make observations

3.
Critically interpret data and synthesize biological messages

4.
Understand the determinants of cell fate, memory, and response to stimuli

Project
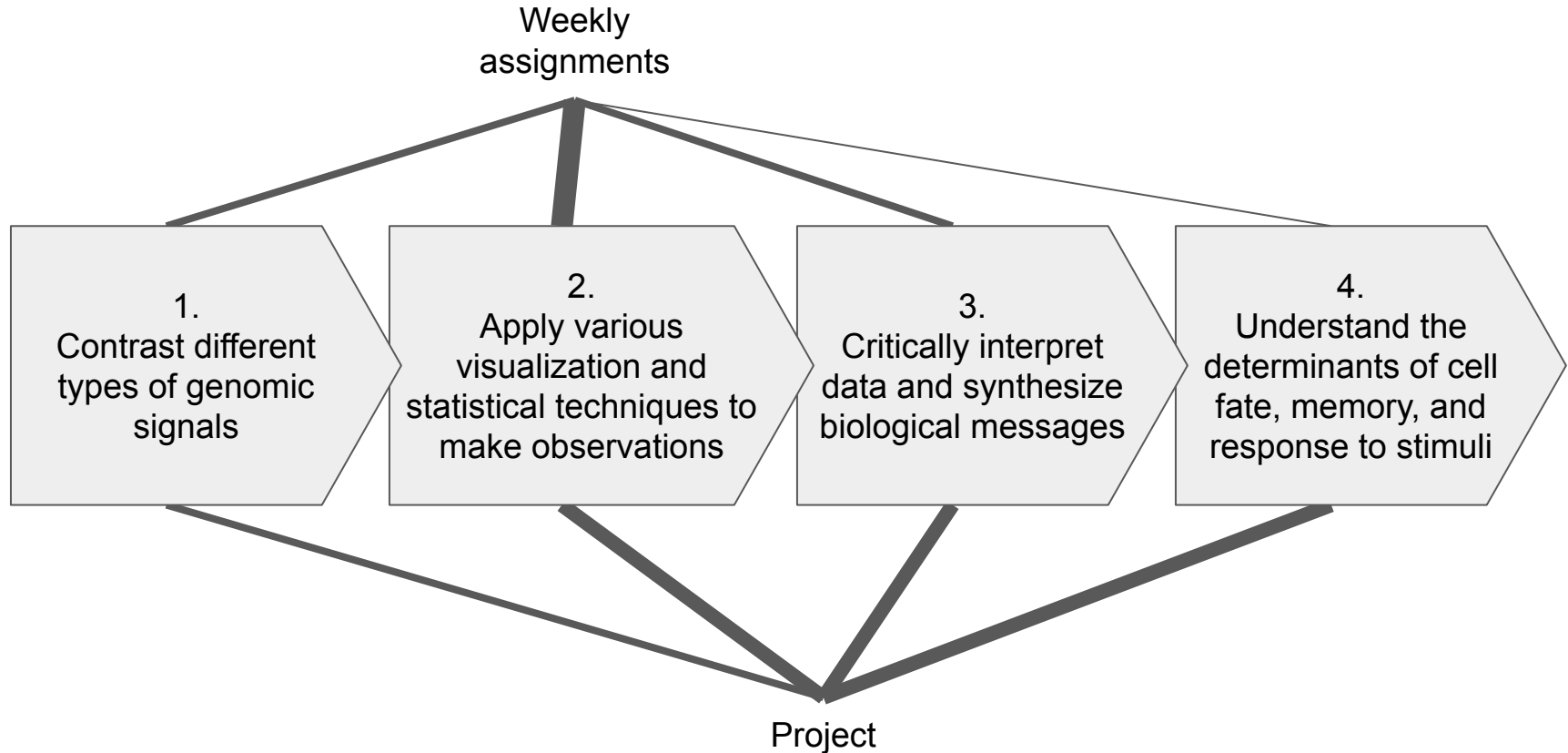
Tentative schedule of the course

Each week:

- **~5-10min of debriefing** on the exercises from the previous week

- **~30min of "lecture"** to provide background and theory

  (15min break)

- **~50-55min of hands-on** demonstration (live coding)

| W	Date	Lecture (~30min)	Mixed/hands-on (~60min)
1	21/02	Introduction to the course	Introduction to the practical tools for the course
2	28/02	Genome builds, transcriptome assemblies and annotations	AnnotationHub, Genomic Ranges, Features and their manipulating in R
3	07/03	Overview of NGS chromatin assays and their analysis	Primary ChIPseq analysis pipeline
4	14/03	Specificities and QCs on primary analysis	Finding data from the literature
5	21/03	Functional elements & the histone code	Visualization and exploration of histone modifications
6	28/03	Types of transcriptional regulators, modes and dynamics of their binding	Motif analysis
7	04/04	DNA accessibility, nucleosome positioning, and TF footprints	Analysis and exploration of ATAC-seq data
8	11/04	Differential analysis	Differential analysis
*	18/04	###### Easter - No Course ######	
*	25/04	###### Easter - No Course ######	
9	02/05	Normalization and enrichment analysis	Normalization and enrichment analysis
10	09/05	From repression to expression and back	Clustering genomic signals and characterizing the clusters
11	16/05	DNA methylation and CpG islands	DNAme visualization
12	23/05	Chromatin conformation, domains and looping	Working with distal regulatory elements
13	30/05	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 (newer quarto)

- git & github : tutorials

Particularly important for the course

Introduction to git and github
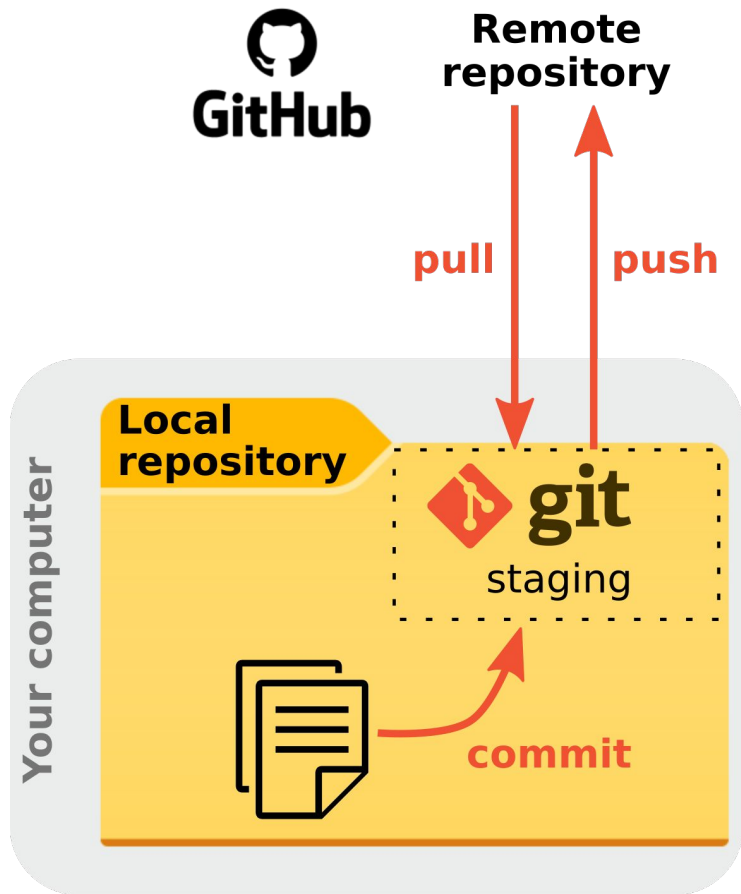
Setting up your github repository for the course:

- 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/ethEpigen2025_materials , make sure you are logged into github, and click the "Fork" button
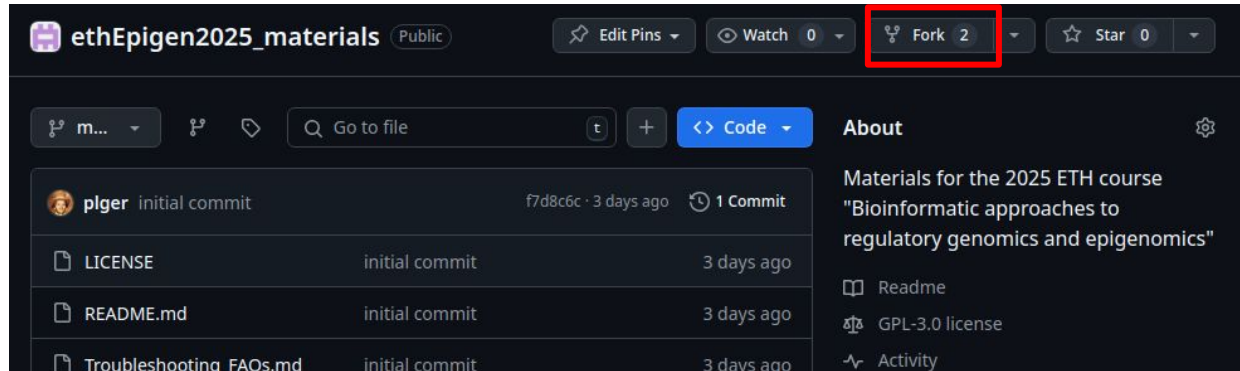
This week's assignment

1. If you don't have one already, create a [github](#) account. Send your github username to me (Pierre-Luc) via slack.

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

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

4. 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()`)

5. Render your markdown, and push **both the Rmd and html files** to your (forked) github repo, in the directory `week01/`