



Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L

Pierre-Luc Germain



Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L
Pierre-Luc Germain

Plan for today

- About your lecturer and 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 & head of the Computational Neurogenomics group at the D-HEST Institute for Neurosciences, ETH
- Originally 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 data; TF binding & activity inference, reconstruction of transcriptional networks

About your fellow students

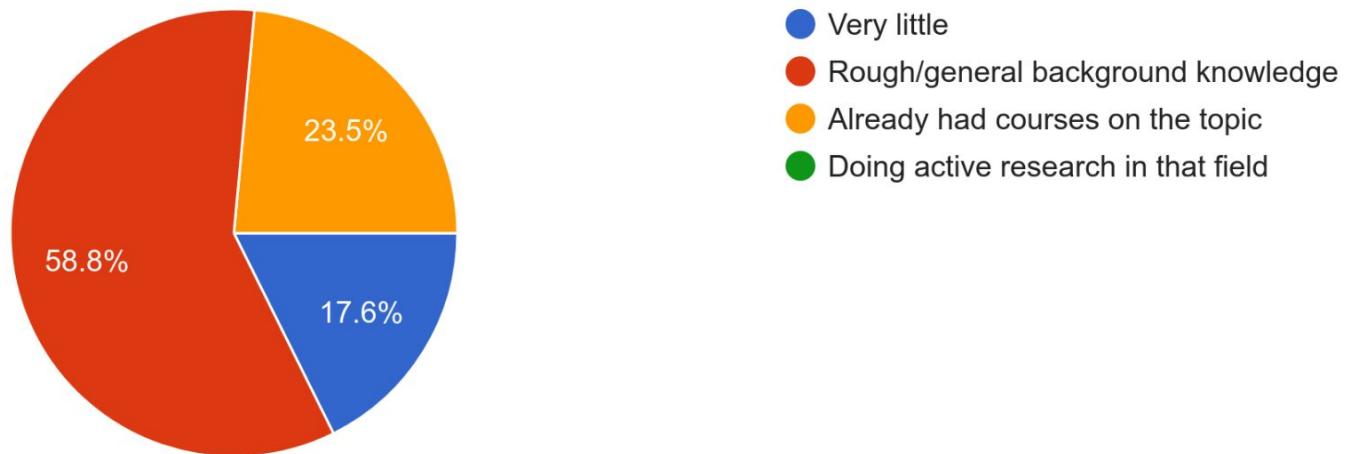
Chiefly HEST and CBB master students, a few PhD students...

About your fellow students

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

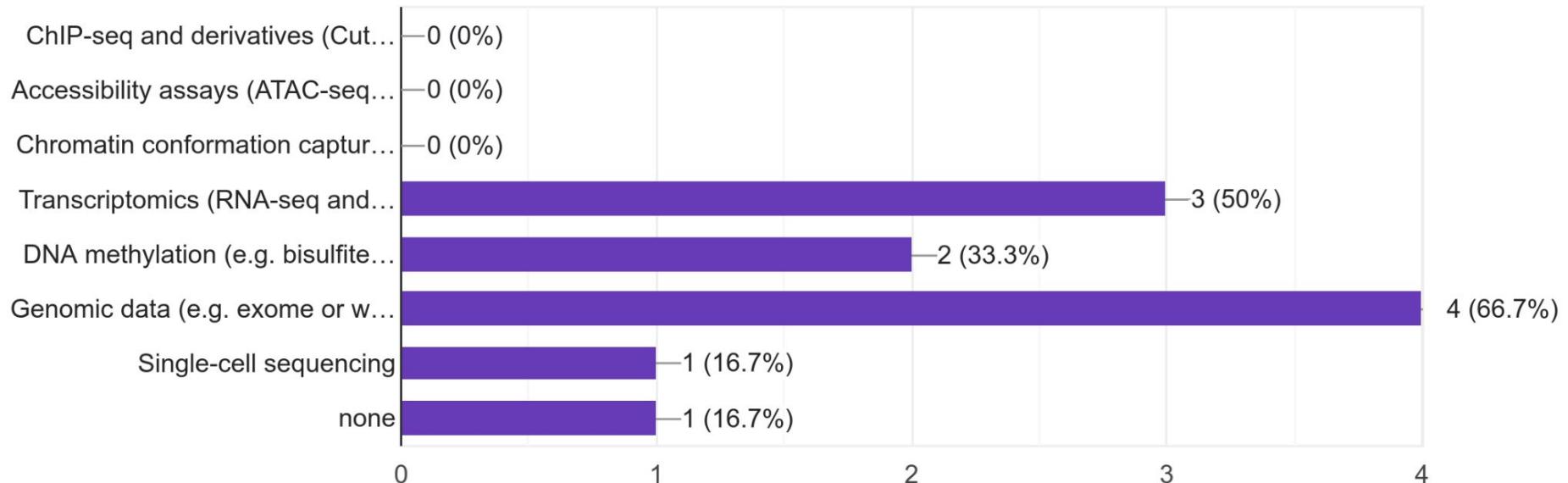
Prior knowledge about epigenetics & regulatory genetics

17 responses



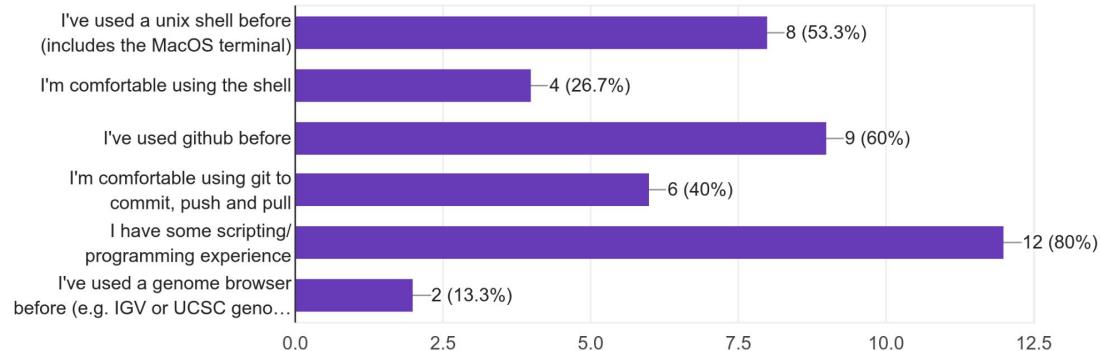
If you have worked with Next Generation Sequencing data before, what kind?

6 responses



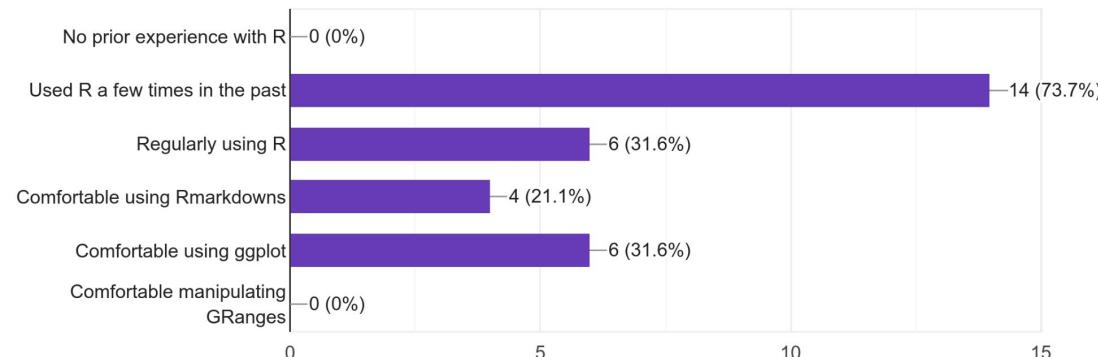
General computer skills; please check all that are relevant

15 responses



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

19 responses



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

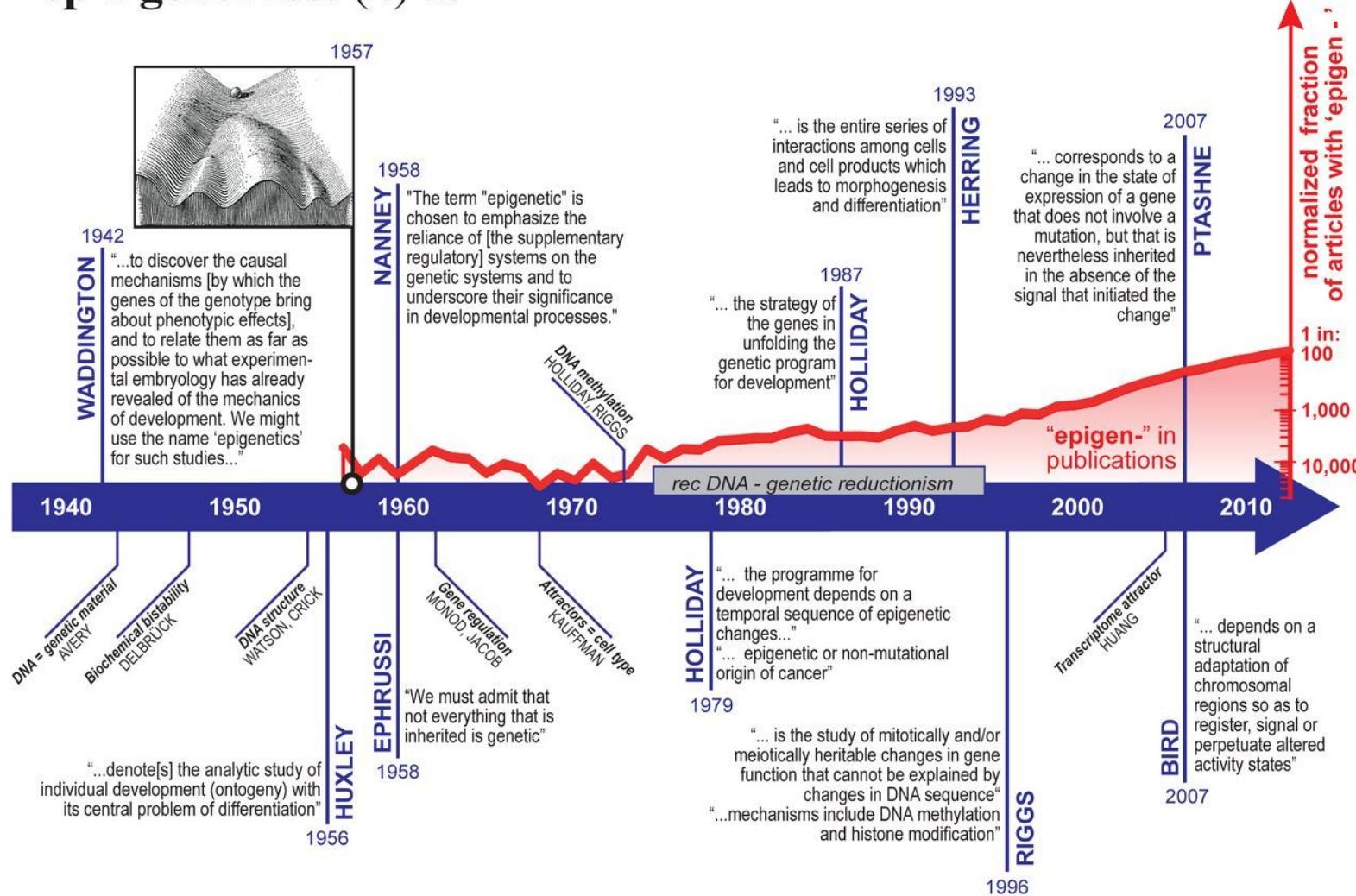
Familiarity with R (as can be obtained from statistics courses, dedicated R courses, or some bioinformatic courses)

Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L

Pierre-Luc Germain

epi•genet•ics: (n) ...

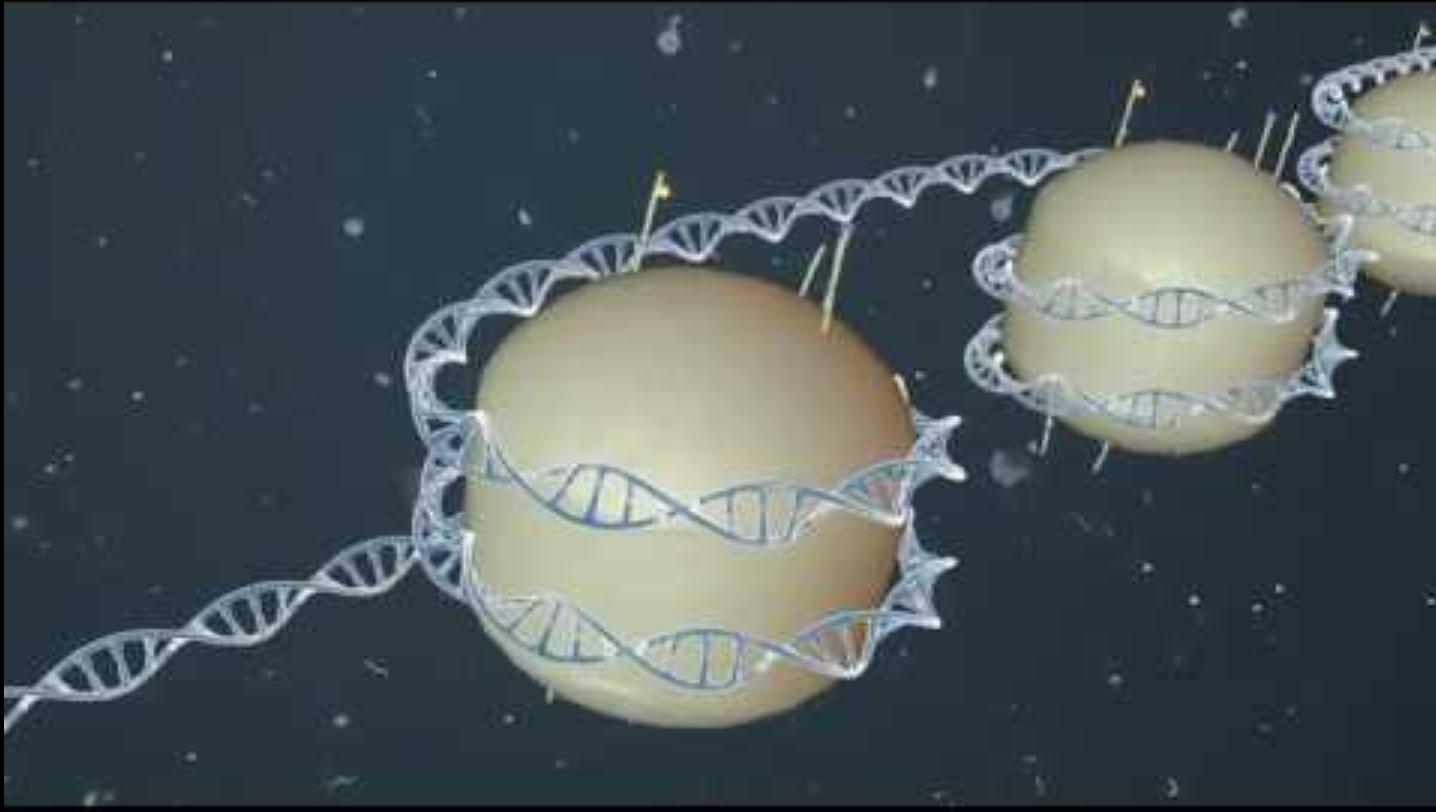


(Oliveira Pisco,
Fouquier d'Hérouël
and Huang, 2016)

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"

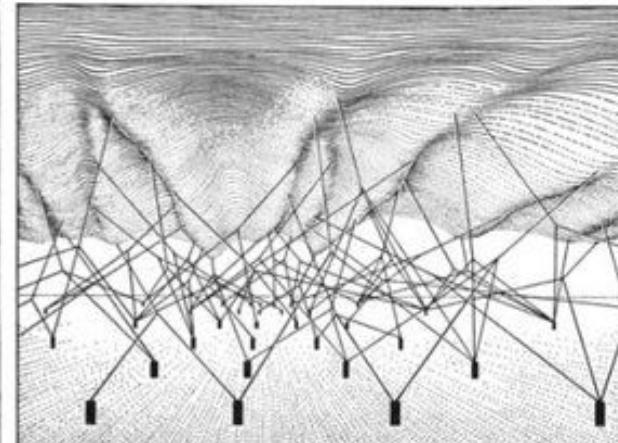
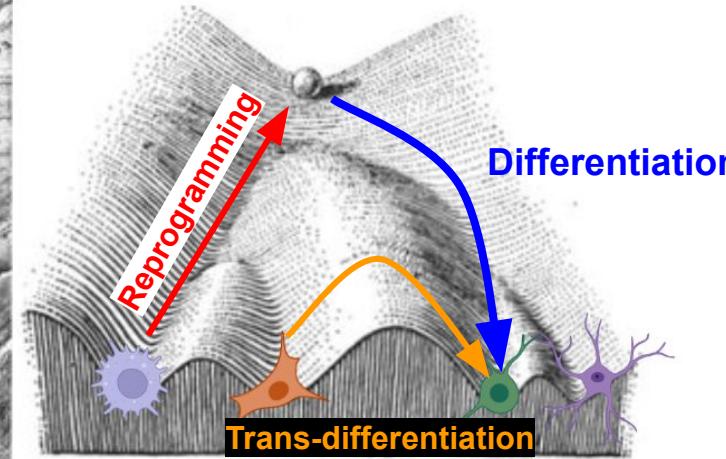
Basic primer on epigenetics



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"

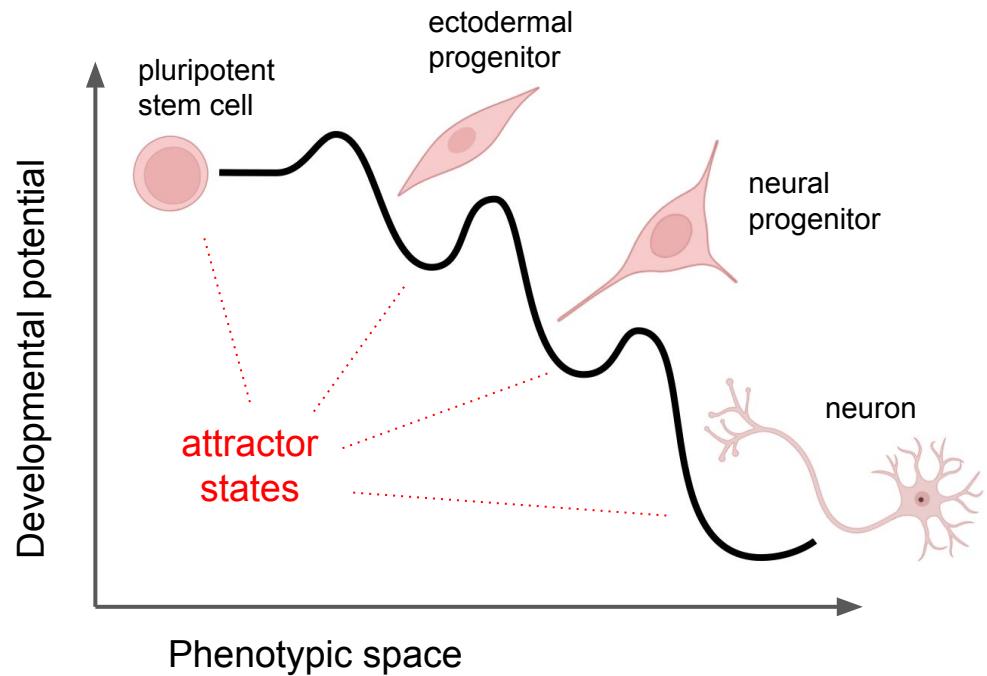
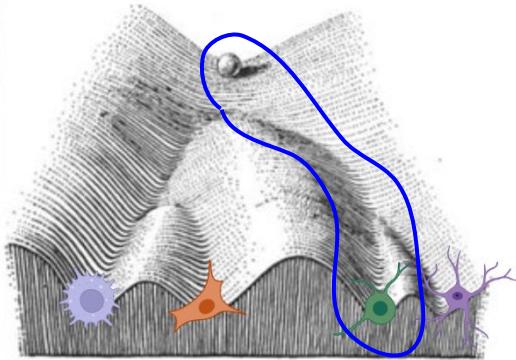
Waddington's epigenetic landscape



Left: from Waddington (1940) "Organisers & Genes"

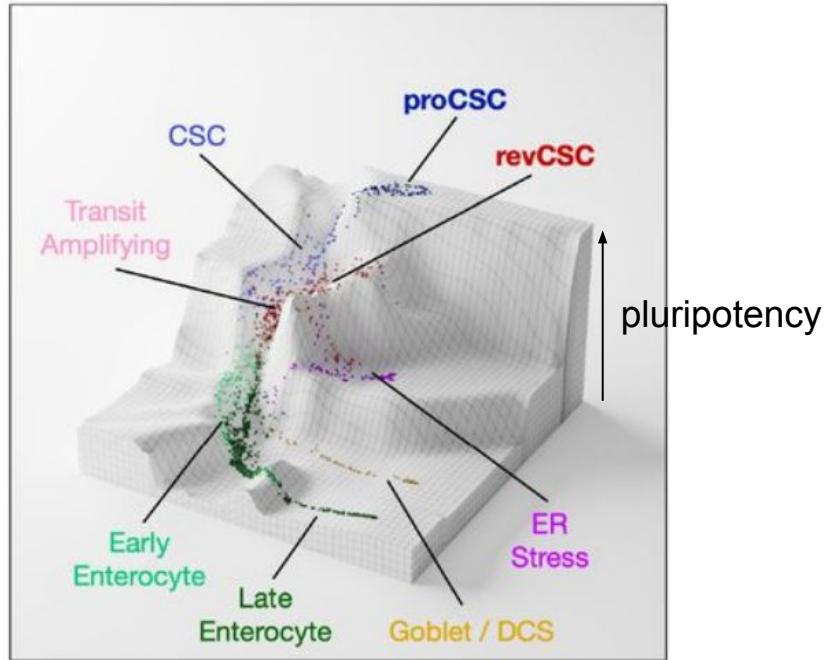
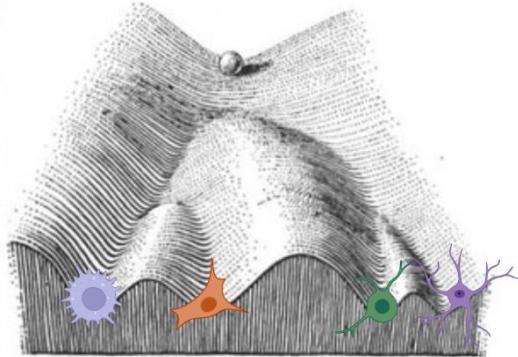
Right: from Waddington (1957) "The strategy of the genes"

Maintaining and changing cell identity



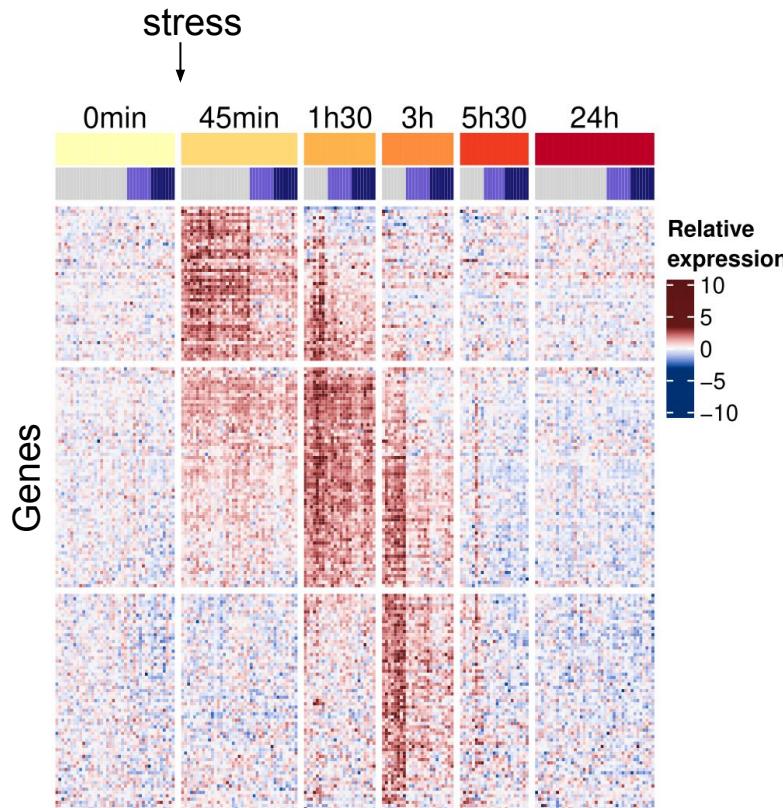
Maintaining and changing cell identity

From concept to quantitative measurement



(adapted from
Qin et al., Cell 2023)

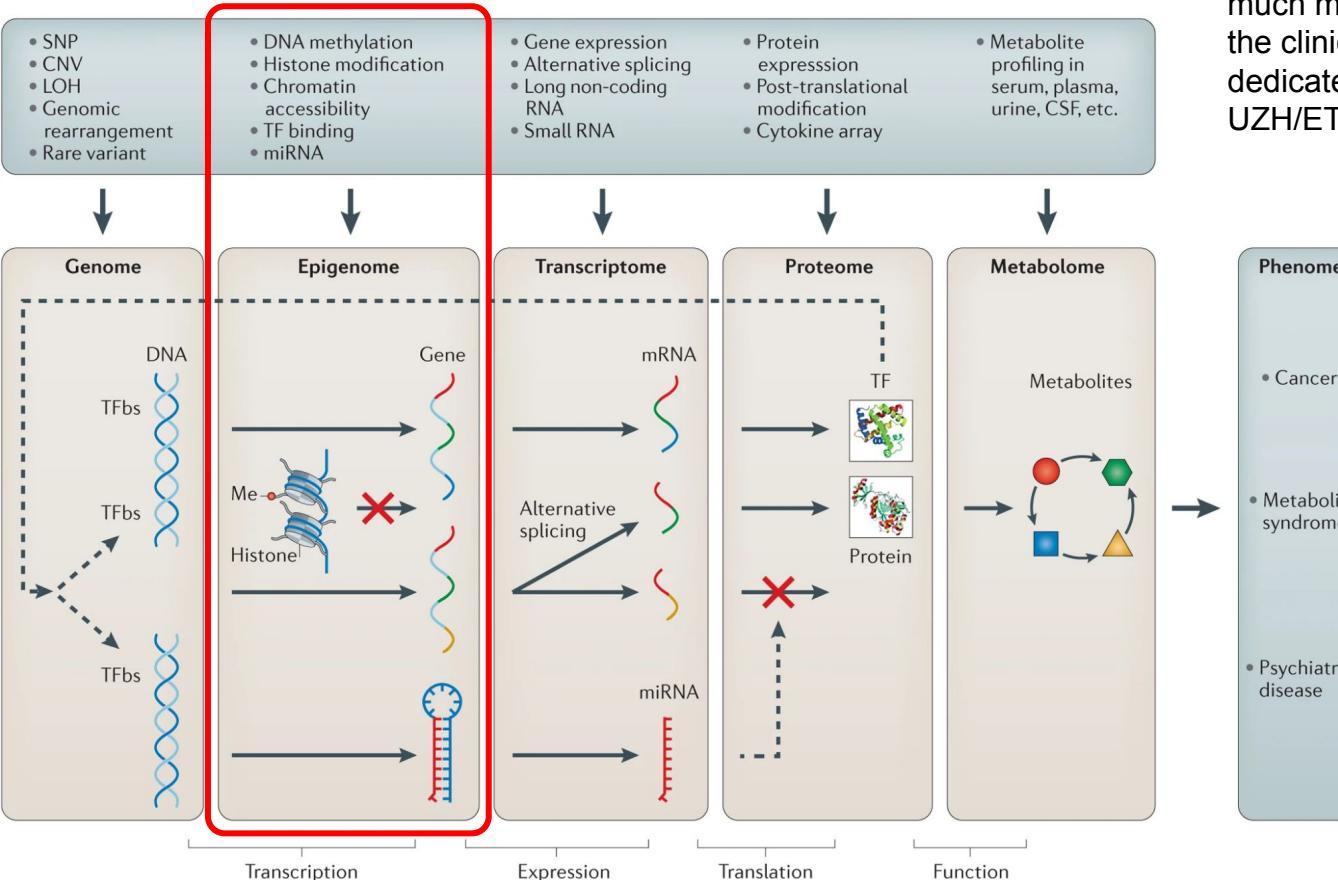
Regulating cell-type-specific responses to stimuli



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, Waag et al., *Nat Comm* 2025)

Situating the course



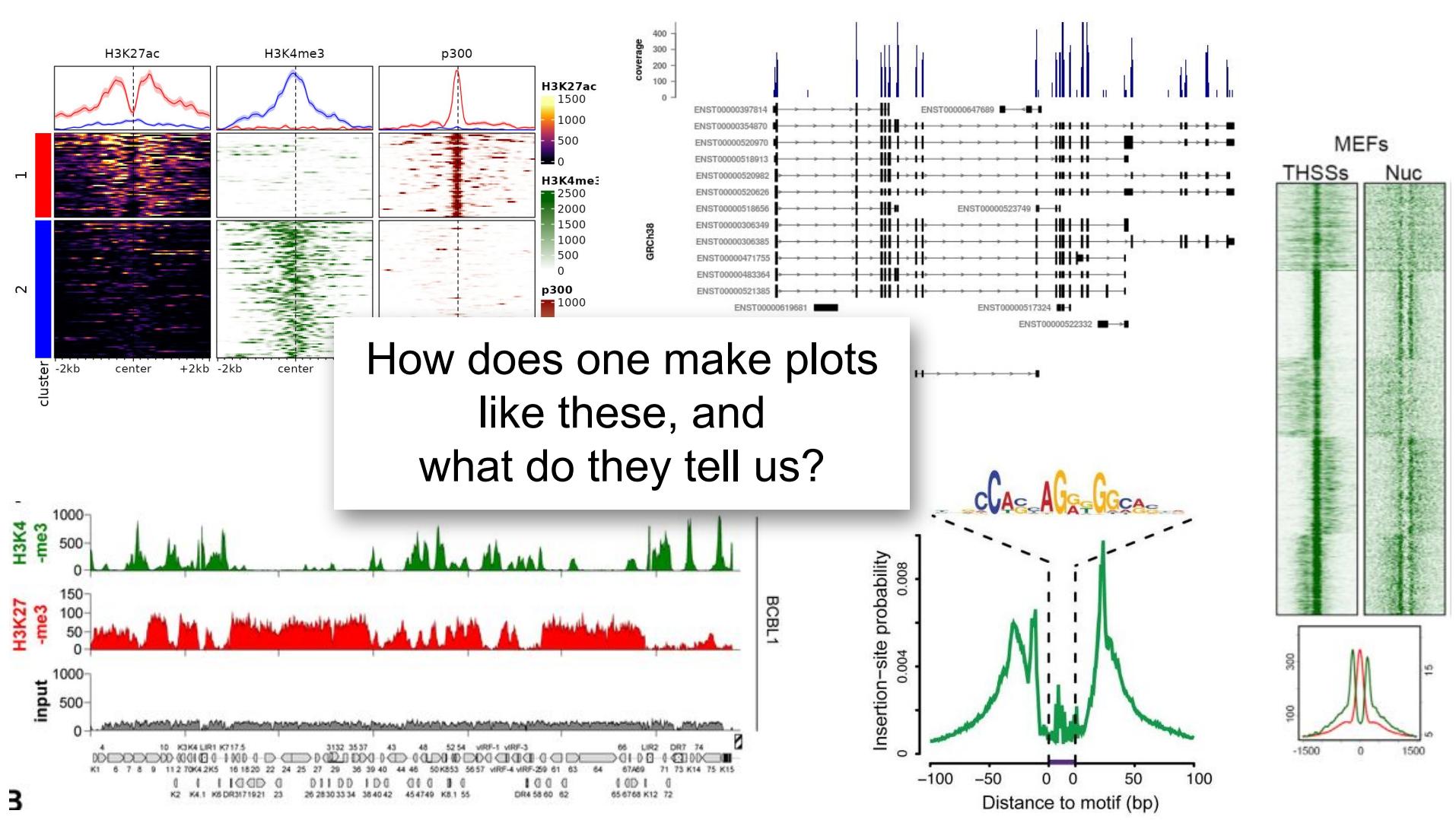
While genomics and transcriptomics are much more widespread (in research and in the clinic), until this course there was no dedicated teaching on epigenomics at UZH/ETH.

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.

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;

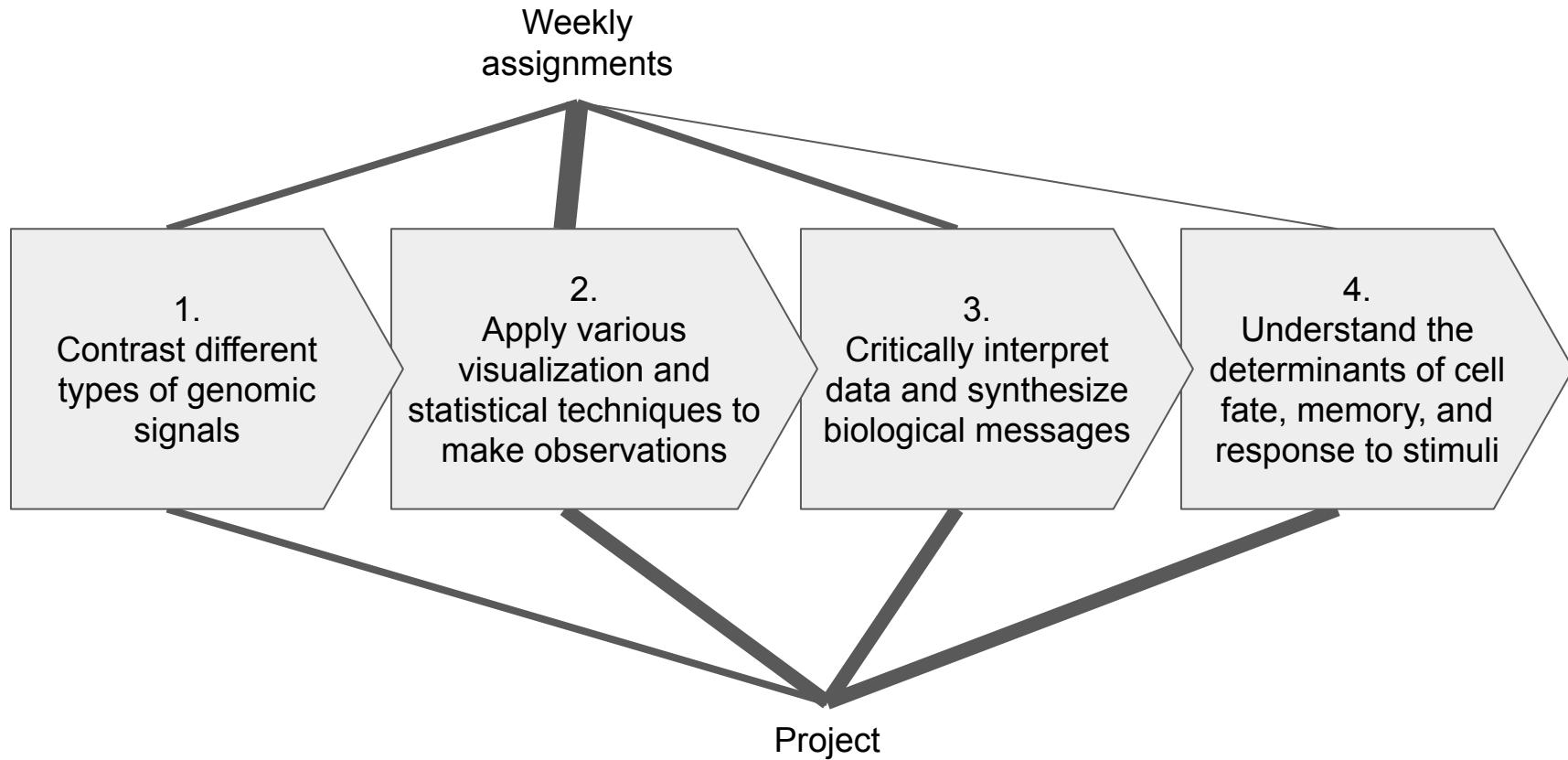
2. to understand and critically appraise, from a genomics perspective and through **hands-on data exploration**, the key **concepts underlying chromatin regulation** of transcription and its impact on various biological phenomena.



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



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	02.20	Introduction to the course	Introduction to the practical tools for the course
2	02.27	Genome builds, transcriptome assemblies and annotations	AnnotationHub, Genomic Ranges, Features and their manipulating in R
3	03.06*	Overview of NGS chromatin assays and their analysis	Primary ChIPseq analysis pipeline
4	03.13	QC, functional elements & the histone code	Finding data, data visualization and exploration
5	03.20	Differential analysis and normalization	Differential analysis and normalization
6	03.27	Types of transcriptional regulators, modes and dynamics of their binding	Motif analysis
*	04.03	##### Spring break - No Course #####	
*	04.10	##### Spring break - No Course #####	
7	04.17	DNA accessibility, nucleosome positioning, and TF footprints	Analysis and exploration of ATAC-seq data
8	04.24	DNA methylation and CpG islands	DNAm analysis and visualization
*	05.01	##### Labour day - No Course	
9	05.08	From repression to expression and back	Clustering genomic signals and characterizing the clusters
10	05.15#	Chromatin conformation, domains and looping	Working with distal regulatory elements
11	05.22	Multi-modal integration	Multi-modal integration
12	05.29	Single-cell chromatin assays; chromatin and disease; open questions	Projects

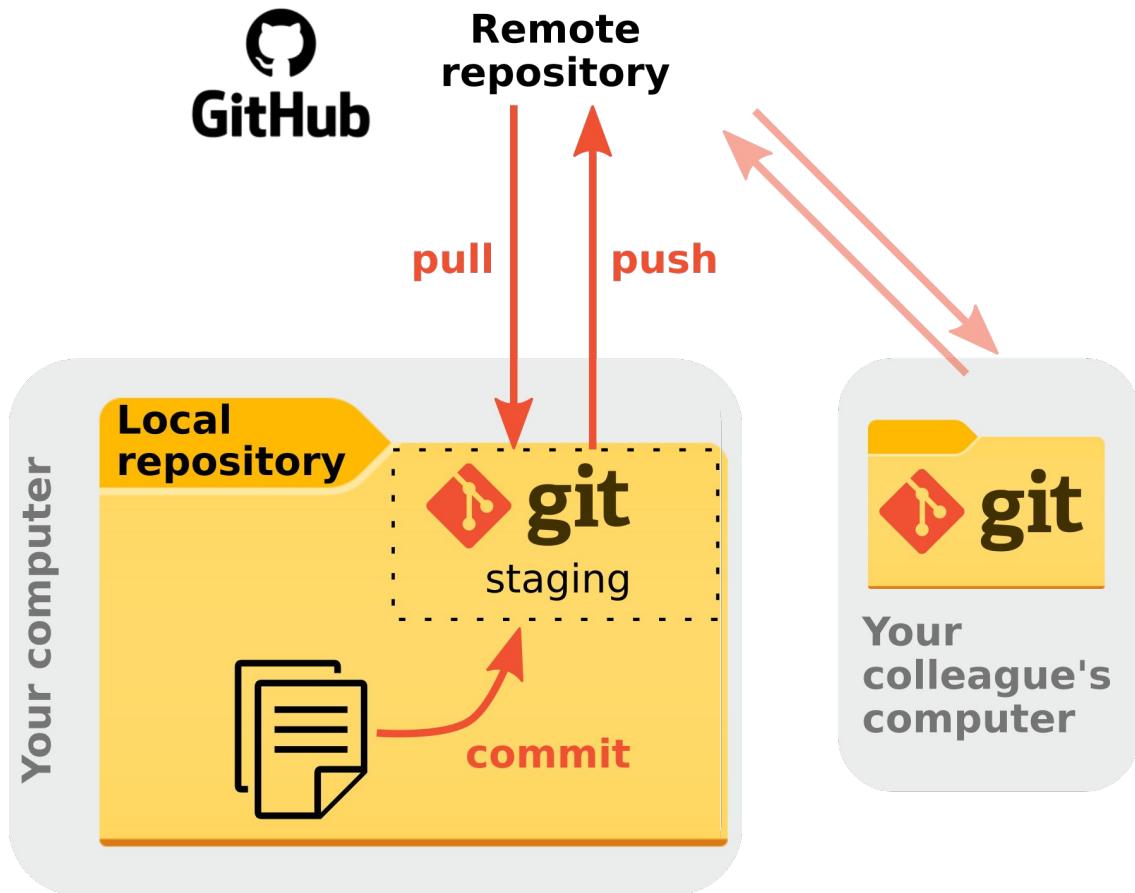
The tools we'll be using... and some documentation

- 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

(For people used to the vscode IDE, a nice alternative to Rstudio is [positron](#))

Introduction to git and github

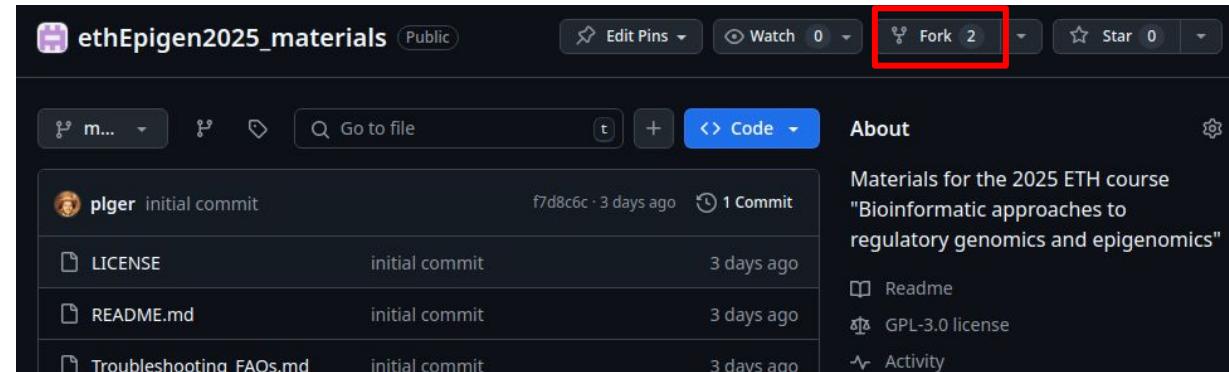


Setting up your github repository for the course:

- 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/ethEpigen2026_material , 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 Emanuel 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 /`