

Central European Institute of Technology BRNO | CZECH REPUBLIC

# Introduction to Bioinformatics (LF:DSIB01)

# Week 1 : Introduction to the course



### Introductions

Let's take 1 minute to introduce ourselves, our name, our background the main project we work on, and our motivation for taking this class

CAAATTGTACAITIOOC ACAAATTGTACATTTGGG Computer Science CAAATTGTACATTTGG Engineering Chemistry TGACAAATTGTACATT **Bioinformatics** 

Mathematics

Biochemistry

ACAAATTGTACATTTG

GACAAATTGTACATTT

CAAATTGTACATTTGG

CAAATTGTACATTTG

TIGACAAATTGTACATT

CAAATTGTACATTTGG

TTGACAAATTGTACATT

CAAATTGTACATTTGGGG

TGACAAATTCT

AAATTGTACATTTGG

AAATTGTACATTT

WATTERIAL ATT

**Statistics** 

Biology

# What is Bioinformatics?

A science that uses Computational and Statistical tools in order to address Biological questions.

Key Skills

Scientific Rigor Engineering Mentality Collaborative Spirit Interdisciplinary Understanding Simplification and Presentation

#### **Key Pitfalls**

Hacker Mentality Bad Code > Reproducibility Uninterpretable Models Jargon and Gatekeeping Losing track of goal: <u>Biology</u>



## Goals of this course

Someone finishing this course should:

- a) Have a <u>basic understanding</u> of the theory and practice of Bioinformatic Research
- b) Be able to <u>practically function</u> in a Bioinformatics research group setting or other collaborative setting
- c) Be able to produce and present analysis results



# Syllabus / week

Week	Date	Material	
Week 1	17-Sep-21	Introduction to the course	
Week 2	24-Sep-21	Intro to Molecular Biology / Genomics	Organizing data, Git
Week 3	1-Oct-21	Exploring the human genome	Sequence Databases, Preprocessing NGS data
Week 4	8-Oct-21	Sequence Alignment	NGS data Alignment, Peak Calling
Week 5	15-Oct-21	Sequence (over)Representation	Motif Finding Tools
-	22-Oct-21	Week Off	
-	29-Oct-21	Week Off	
Week 6	5-Nov-21	Evaluation Metrics and Data Visualization	
Week 7	12-Nov-21	ML techniques I (log regresion, SVM)	
Week 8	19-Nov-21	ML techniques II (decision trees, random forest)	
Week 9	16-Nov-21	ML techniques III (ANN, perceptrons)	
Week 10	3-Dec-21	ML techniques IV (CNN) + Advances	
-	10-Dec-21	Wrap up	
-	17-Dec-21 (?)	Colloquium	



# **Resources / Tools**

#### <u>Tools</u>

- Python
- Github
- Google Colaboratory
- Jupyter Notebook



#### AN INTRODUCTION TO BIOINFORMATICS ALGORITHMS

NEIL C. JONES AND PAVEL A. PEVZNER



# The Clean Coder A Code of Conduct for Professional Programmers

Robert C. Martir

**Robert C. Martin Series** 

#### Resources

https://www.kaggle.com/learn/overview https://wiki.python.org/moin/BeginnersGuide/NonProgrammers https://wiki.python.org/moin/BeginnersGuide/Programmers https://guides.github.com/activities/hello-world/ https://research.google.com/colaboratory/faq.html https://www.dataquest.io/blog/jupyter-notebook-tutorial/ http://www.biostathandbook.com/index.html

# <section-header>



#### 

## **Evaluation**

Ongoing project through the semester

Small task after each practical – can be finished at home

Colloquium: discussion of questions, results etc



