

Computational Science: Bioinformatics

Scientific Computing and Machine Learning (SCML)

12 August 2025

Torbjørn Rognes
Department of Informatics, UiO
torognes@ifi.uio.no



UiO : **University of Oslo**

Computational Science at IFI

Study programme (master):

- Computational Science (CS)

Study programme options at IFI:

- **CS: Bioinformatics** - in the Scientific Computing and Machine Learning (SCML) group (Vitenskapelige beregninger og maskinlæring) – focus on genetics and the molecular level
- **CS: Imaging and Biomedical Computing** – in the Digital Signal Processing and Image Analysis (DSB) group (Digital signalbehandling og bildeanalyse) – focus on organs and non-molecular aspects

What is bioinformatics?

An interdisciplinary subject:

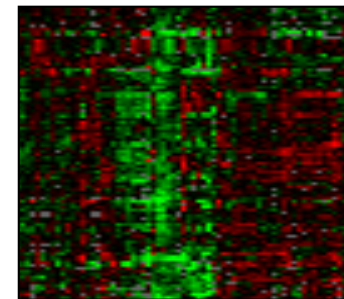
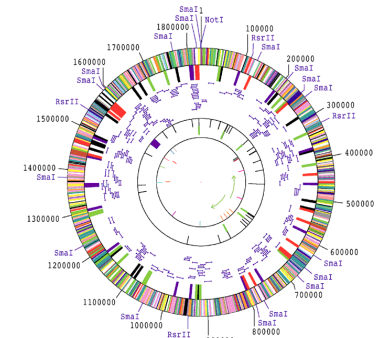
Computer science/maths/statistics + biology/medicine

Definition: Bioinformatics is the development and use of computational and mathematical methods to gather, process and interpret molecular biological data.

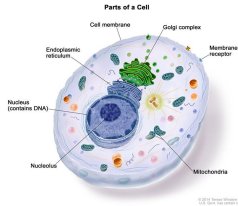
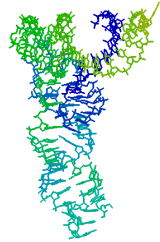
Aim of research: To increase our understanding of biological mechanisms while developing better theories and methods in computer science and statistics.

Methods: Machine learning, statistics, graphs, algorithms, visualisation, parallelisation, efficient data structures, ...

Data: Genomes, genes, DNA, RNA, proteins, immune receptors, gene expression levels, 3D molecular structures, cells, organisms, populations, gene variants and mutations, patient info, ...



Bioinformatics at many levels



DNA

RNA

Protein

Cell

Organ

Individual

Population

Biosphere

Genomics

Transcript-
omics

Proteomics

System
biology

Neuro-
informatics

Precision
medicine

Population
genetics

Metagenomics

Genome
assembly

RNomics

Structural
biology

Cell
simulation

Organ
modelling/
simulation

Variant
detection

Epidemiology

Evolutionary
biology

Genefinding

Microarrays

Drug design

Metabolism
studies

Meta-
genomics

Phylogenomics

Annotation

RNA-folding

MS analysis

Mapping

RNA-seq

Binding site
analysis

Cancer
genomics

ChIP-Seq

Structure
biology

Interaction
networks

Scientific staff at SCML

Please contact any of the scientific staff if you are interested in a master thesis with us



**Geir Kjetil
Sandve**
Professor
Group leader



**Xing
Cai**
Professor



**Torbjørn
Rognes**
Professor



**Pooya
Zakeri**
Assoc. prof.



**Anne-Marie
George**
Assoc. prof.



**Joakim
Sundnes**
Assoc. prof.



**Rolf I.
Skotheim**
Professor II



**Chakravarthi
Kanduri**
Researcher



**Milena
Pavlovic**
Postdoc



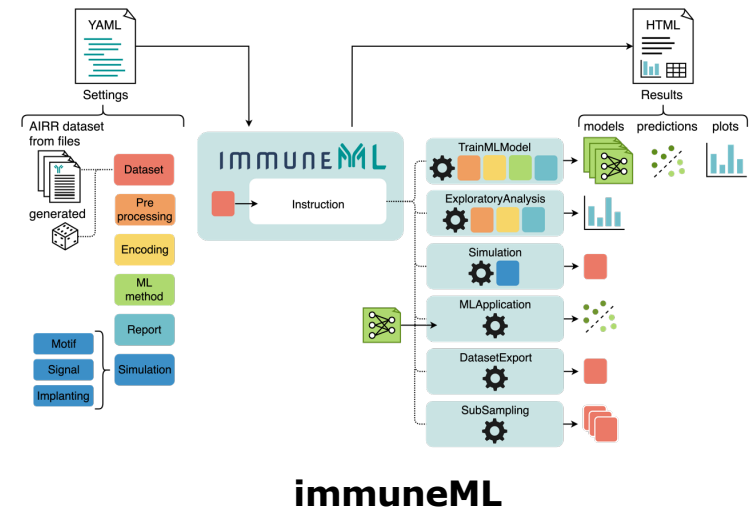
**Ole Christian
Lingjærde**
Professor
Head of Department

Main research areas

Advanced use and development of new methods and tools in computer science and statistics to analyse molecular data

Research projects

- Cancer genomics (OC Lingjærde)
- ML for Immune receptor repertoire analysis + Genome statistics and visualisation (GK Sandve)
- Physics-informed ML and high-performance computing for biomedical applications (X Cai)
- Genome heterogeneity in cancer (RI Skotheim)
- Machine learning, reinforcement learning, user preference learning (AM George)
- Machine learning (P Zakeri)
- Microbiome bioinformatics and efficient sequence comparison (T Rognes)



CS Programme structure

4. semester	Master's thesis	Master's thesis	Master's thesis
3. semester	Master's thesis	Master's thesis	Master's thesis
2. semester	IN4200 – High-Performance Computing and Numerical Projects / Mandatory course	Elective course	Elective course
1. semester	FYS-STK4155 – Applied Data Analysis and Machine Learning / Mandatory course + HSE courses	MAT4110 – Introduction to Numerical Analysis / Elective course	Elective course
	10 ECTS credits	10 ECTS credits	10 ECTS credits

Alt: 10 credits
elective course

Alt: 10 credits
master thesis

Mandatory, 2 out of these 3 courses:

- IN4200 – High-Performance Computing and Numerical Projects
- FYS-STK4155 – Applied Data Analysis and Machine Learning
- MAT4110 – Introduction to Numerical Analysis

Suggested elective courses

CS: Bioinformatics

- IN4030 – Introduction to bioinformatics (spring)
- IN4050 – Introduction to artificial intelligence and machine learning (spring)
- BIOS-IN5410 – Bioinformatics for Molecular Biology (autumn)

Other relevant courses

- IN4020 – Database systems (spring)
- IN4120 – Search technology (autumn)
- IN4310 – Deep Learning for Image Analysis (spring)
- IN5050 – Programming Heterogeneous Multi-Core Architectures (spring)
- IN5320 – Development in Platform Ecosystems (autumn)
- IN5490 – Advanced Topics in Artificial Intelligence for Intelligent Systems (autumn)
- IN-STK5000 – Responsible Data Science (autumn)
- BIOS-IN5010 – Computational physiology (summer)
- STK4051 Computational Statistics (spring)
- STK-IN4300 – Statistical Learning Methods in Data Science (autumn)
- TEK5020 – Pattern recognition (autumn)
- MEDFL5155 – Introduction to statistics and bioinformatics for the analysis of large-scale biological data (autumn, 5 credits)

Please see studies page of the Bioinformatics in Life Science (BILS):

<https://www.mn.uio.no/bils/english/studies/>

Master thesis at SCML

- Look for available master projects on website:
<https://www.mn.uio.no/ifi/english/research/groups/scml/master-projects/>
- Presentation of new projects in September/October
- Contact potential supervisors
- Usually 60 credits (0+10+20+30 credits in each semester)
- Make contact early!

Deadlines:

- 1 December 2025: Study plan, select project and supervisor
- 1 June 2026: "Essay"
- 15 May 2027: Master thesis

Some former master theses

- A Modular Pipeline for RNA Sequencing Data Analysis
- Adding 3D structure support to immuneML
- Addressing documentation and testing for Undocumented workflows
- Aiming at heterogeneous parallel computing for bioinformatics: To overlap or not to overlap?
- Assessment of Non-negative Matrix Factorization Methods for Mutational Signature Analysis
- Automated assessment of similarity between synthetic and real Adaptive Immune Receptor Repertoire data
- Automated code generation for simulating cardiac cells with lookup tables
- Computational Discovery of Micropeptides in Brain Cells
- ChIPPipeline: Reproducible and Extensible Snakemake ChIP Sequencing Data Analysis Pipeline
- Development of Quantitative Adverse Outcome Pathways Tools for Future Risk Assessment
- Effects of splicing mutations in prostate cancer
- Evaluating the utility of Micro-C and computational simulations for enhancer-promoter interaction analysis
- Generation probability of immune receptors and full sequence implanting in adaptive immune receptor repertoires
- Implementation and evaluation of GPU-accelerated rapid brain tumor classification
- Incorporation of Dimensionality Reduction Methods into immuneML
- Integrating Clustering for Adaptive Immune Receptors in immuneML
- Introducing a Generative Model Pipeline in immuneML
- Investigating the Relationship between DNA Methylation in Whole Blood and Chemical Exposure for Predictive Modeling
- KIVS - Graph K-mer Indexer & Variant Signature Finder: Improving the performance of index creation for alignment-free genotyping
- Quantitative Adverse Outcome Pathway Modeling for Chemical Hazard Assessment
- Speeding up Genotyping through GPU Acceleration
- Streamlining the use of parallelization in a software platform for domain-tailored machine learning analyses
- VirMake: a Snakemake pipeline for viral metagenomic data analysis