

Analysis of data from high-throughput molecular biology experiments

Objective

The aim of this course is to train students to become advanced and independent users of state-of-the-art bioinformatics tools applicable on large data sets generated with state-of-the-art experimental setups.

Contents

The course contains the fundamental theory of bioinformatics analysis of large data sets from high-throughput genomics and proteomics experiments – in particular, massively parallel DNA sequencing and protein mass spectrometry: how this theory is manifest in state-of-the-art tools for handling, analyzing, and visualizing the data; how these tools are applied on real high-throughput molecular biology data; and how the outcome may be interpreted in a biologically relevant context.

The course consists of lectures, student-prepared presentations, computer-based laboratory exercises, and a project.

Specific goals

After passing the course you should be able to:

1. Describe widely used high-throughput experimental techniques employed to investigate the DNA, RNA, and protein contents of a cell, tissue, or organism.
2. Explain the theory of state-of-the-art tools/algorithms for processing data from high-throughput molecular biology experiments.
3. Choose appropriate methods and tools for processing data from high-throughput molecular biology experiments.
4. Use tools for processing data from high-throughput molecular biology experiments.
5. Interpret the results of these analyses in a biologically or medically relevant context.
6. Reflect over the choice of methods and tools and how it influences the outcome of the analyses.

Prerequisites

The following courses, or equivalent, are recommended: Bioinformatics and basic probability theory corresponding to BB2440 Bioinformatics and biostatistics (or bioinformatics corresponding to DD2396 Bioinformatics, and probability theory corresponding to SF1901 Probability theory and statistics). Computer acquaintance and programming experience equivalent to DD2397/DD2404 Applied bioinformatics.

Course information and communication

We will mainly use the **Course web** (previously known as "KTH Social", to provide updated information and course material: <https://www.kth.se/social/course/BB2490/>. Your access will be activated after course registration. Log in using your kth.se-account. Lecture handouts, reading assignments, and computer exercise instructions will be available there. Please check regularly for news and instructions.

Structure of the course

The *first half* of the course (20 Jan – 17 Feb) contains lectures, student-prepared presentations, and computer-based laboratory exercises.

The *second half* of the course (17 Feb – 18 Mar) is dedicated to project work. The second half of the course ends with a presentation of your project work (date: Friday 18 March).

Lectures

There are 11 lectures. Each lecture will require student interaction and active participation.

Seminars

There are 3 seminars during the student project (see below). Each seminar will require student interaction and active participation.

Required participation

Analysis of data from high-throughput molecular biology experiments

You are required to be present at at least 10 out of the 13 events (L2-L11 and S1-S3).

Literature

Scientific articles and web resources as assigned during the course.

This list might be subject to change. List of papers will appear on the **Course web**.

There will also be handouts from the lectures. Handouts will appear on the **Course web**.

Reading assignments

To each lecture there may be a reading or other preparatory assignment. If so, it is mandatory to prepare according to these instructions, e.g. reading the texts in the reading assignment, before the lecture. This will help you to prepare for the lecture, and will help you when you participate in the interactive parts of the lectures. Reading assignments will be communicated through the **Course web**.

Student-prepared presentations (L4 and L5-L10)

[a] At some of the lectures (L5-L10), 45 minutes will be devoted to article presentations prepared by you, the students at this course. You have to do your assigned share of this to pass the course. Please note, these presentations are *not* graded; the presentations are explicitly only for learning purposes and not for grading purposes. These presentations will be done in pairs. Each student will present at one or two occasions. Exact schedule will be communicated towards the end of the first course week. It will be available on the **Course web**.

[b] The student presentations at lecture 4 (L4) are different. Here, all of you are expected to prepare *individually* a short presentation (with *no* powerpoint slides) for each of a number of topics (typically 7-8 topics, provided on the **Course web**). A presenter for each topic will then be picked at random. Thus, everybody needs to prepare each topic, but not everyone will be chosen for presentation.

Computer exercises

There are 4 regular computer exercises where you follow the given instructions. You are supposed to work in pairs. These exercises will be graded pass/fail (see Requirements for examination). The instructions are available on the **Course web**.

Computing resources

You will be using the Uppmax computing cluster for computer exercises and during the project (<http://www.uppmax.uu.se>, specifically the Milou Cluster, <https://www.uppmax.uu.se/the-milou-cluster>). You need to get an account at Uppmax, and once you have an account you should apply for membership in the specific "course project". The instructions are available on the **Course web**.

Quiz

A quiz (kontrollskrivning) is scheduled for Tuesday 16 February at 13:00-15:00 in FB53. Please note that it starts 13:00 sharp. The quiz will contain a number of multiple-choice questions. You have to pass the quiz in order to be allowed start working on the project. In order to pass the quiz, you need to answer a certain number of questions correctly.

Project

You will receive a real data set and a brief project plan. You are also welcome – even encouraged – to bring your own data set, if you have the possibility for that. You should then analyze this data set with a relevant set of the tools, probably some of which you've encountered earlier in the course. The analysis should aim at extracting and interpreting biologically relevant information. You are supposed to work in groups of 3 students. You will be assigned a group. During the project period, there are **three project seminars (S1-S3)**. Presence is mandatory. At the project seminars you will present your project plan and your individual diaries (S1) but also discuss any problems you might run into (S2).

Each group should **present a poster** at the poster session on Friday 18 March (14:00-17:00, at Science for Life Laboratory). The poster should contain the results and conclusions from your data analysis, along with relevant information concerning how you analyzed your data and how

Analysis of data from high-throughput molecular biology experiments

you support your conclusions. One of the project seminars is devoted to poster session preparation (S3).

In addition to the poster, you should also **maintain an individual diary** of what you do in the project, and where you also can keep figures etc. that you create during the work. You should use an Internet-based documentation system for the diary (e.g., a wiki or a blogg). The diary is part of the examination, *i.e.*, it will be used together with the poster to determine your grade. This means that the teachers will need access to your diary during the project period and until one month after the course has finished (for grading purposes). The first entry of your individual diary should be presented at seminar S1.

The poster and each individual diary will be graded A/B/C/D/E/Fx/F (see Requirements for examination). More information about the project will be provided later.

Requirements for examination

The examination comprises four parts:

1. **Presence and active participation at at least 10 out of the 13 lectures and seminars**, starting counting from the second lecture (thus it applies to lectures L2 to L11, S1-S3) and including performing the student-prepared presentation(s) assigned to you. Grading: Pass/Fail
2. **Passed laboratory exercises**. Based on passing the regular computer exercises (Lab1, Lab2, Lab3, Lab4). Grading: Pass/Fail
3. **Written exam (quiz)**. A quiz is organized before the start of the projects. You need to pass this exam in order to be allowed to start working on the project. One re-exam opportunity will be arranged before project start. Grading: Pass/Fail
4. **Poster presentation and individual diary for project**. The grade will be based on the poster (one per group) and the diary (one per individual). Grading: A, B, C, D, E, Fx, F.

Grading criteria

Grading criteria will be communicated separately and made available on the **Course web**.

Plagiarism

Plagiarism is a serious offense. Each written document you hand in and which is part of any requirement for examination will be checked for plagiarism.

Contact

Lecturers, course coordinators, and examiners:

Lars Arvestad (LA), lars.arvestad@scilifelab.se, office phone 08-52481428

Olof Emanuelsson (OE), olofem@kth.se, office phone 08-52481458.

Lukas Käll (LK), lukas.kall@scilifelab.se, office phone 08-52481196.

Please make sure you've consulted the **Course web** site for BB2490 before emailing us!

Teaching assistants (TAs) (lab-handledare):

Johannes Alneberg (JA), johannes.alneberg@scilifelab.se

Matthew The (MT), matthew.the@scilifelab.se

Location

Lectures, seminars, poster session:

FA31, FA32, FD41, FB52-55 – AlbaNova University Center, main building, Roslagstullsbacken 21
SciLife – Karolinska Institute campus, Science for Life Laboratory, Tomtebodavägen 23: gamma building, floor 6, Pascal meeting room.

Computer exercises:

5o2spo – KTH main campus, Lindstedtsvägen 5, floor 5.

4v2röd – KTH main campus, Lindstedtsvägen 3, floor 4.

Schedule (also available on the Course web)

Analysis of data from high-throughput molecular biology experiments

Event	Date	Time	Room	Teach.	Contents (brief)
L1	2016-01-20	10-12	FB55	all	Introduction
L2	2016-01-22	10-12	FB53	LA	Genomics: Assembly
L3	2016-01-25	10-12	FA31	LK	Statistics of high-throughput experiments
L4	2016-01-27	10-12	FA31	all	Student presentations HT techniques
Lab1	2016-01-27	13-17	5o2spo	JA/MT	Statistics
L5	2016-01-29	10-12	FA31	LA	Genomics: Mapping (*L2)
L3 (!)	2016-01-29	13-15	FB51	LK	Statistics of high-throughput experiments
L6	2016-02-01	10-12	FB55	OE	Transcriptomics: RNA-seq (*L5)
Lab2	2016-02-01	13-17	5o2spo	JA/MT	Genomics: Mapping/assembly
L7	2016-02-03	10-12	FD41	OE	Transcriptomics: ChIP-seq (*L6)
Lab3	2016-02-05	08-12	4v2röd	JA/MT	Transcriptomics: RNA-Seq, ChIP-Seq
L8	2016-02-08	10-12	FB55	LK	Proteomics: Peptide identification (*L7)
L9	2016-02-10	10-12	FB55	LK	Proteomics: Protein quantification (*L8)
Lab4	2016-02-10	13-17	5o2spo	JA/MT	Proteomics
L10	2016-02-15	10-12	FB55	all	Stud.pres. Project information (*L9)
QUIZ	2016-02-16	13-15	FB53	-	Quiz (kontrollskrivning)
L11	2016-02-17	10-12	FB52	all	Walk-through of Quiz and Project kickoff
S1	2016-02-19	10-12	FB54	all	Project plan and individual diary
S2	2016-02-26	10-12	FB53	all	Exchange of experiences
S3	2016-03-09	10-12	FA31	all	How to prepare poster and poster presentation.
Poster	2016-03-18	14-17	SciLife	all	Poster session

Notes: L = lecture Lab = computer exercise S = seminar
 (*Ln) = student-prepared presentations with the subject introduced in lecture Ln.

Reading assignments and lecture contents might be subject to change.