Analysis of data from high-throughput molecular biology experiments
BB2490 and DD2399 (*KTH*) 7.5 hp – spring 2015

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 tools for processing data from high-throughput molecular biology experiments.
4. Apply 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.

Literature
Scientific articles and web resources as assigned during the course.
This list might be subject to change. List of papers will appear on KTH Social.
There will also be handouts from the lectures. Handouts will appear on KTH Social.

Structure of the course
The *first half* of the course (20 Jan – 18 Feb) contains lectures, student-prepared presentations, and computer-based laboratory exercises.
The *second half* of the course (13 Feb – 20 Mar) is dedicated to project work. The second half of the course ends with a presentation of your project work (date: Friday 20 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.

Reading assignments

Analysis of data from high-throughput molecular biology experiments
BB2490 and DD2399 (*KTH*) 7.5 hp – spring 2015

To each lecture there will be a reading or other preparatory assignment. It is mandatory to prepare according to these instructions, e.g. reading the texts in the reading assignment, before each lecture. This will help you to prepare for the lecture, and will help you when you participate in the interactive parts of the lectures.

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.
[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). 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).

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 subset of the tools you've learnt to handle. The analysis should aim at extracting and interpreting biologically relevant information. You are supposed to work in groups of (2 or) 3 students. You will be assigned a group. During the project start, there are **project seminars**. Presence is mandatory. At the project seminars you will present your project plan and your individual diaries (see below), but also discuss any problems you might run into.

Each group should **present a poster** at the poster session on Friday 20 March. The poster should contain the results and conclusions from your data analysis, along with relevant information concerning how you analyzed your data and how you support your conclusions.

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). Your individual diary should be presented at seminar S1.

The poster will be graded A/B/C/D/E/F and each individual diary will be graded A/B/C/D/E/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**. 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, F.

Reading and other preparatory assignments
There are reading or other preparatory assignments for **each lecture L2-L11**. See separate paper. These assignments will also be posted at **KTH Social**.

Course information and communication
We will mainly use KTH Social, https://www.kth.se/social/course/BB2490/, to provide updated information and course material. 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.

Analysis of data from high-throughput molecular biology experiments
BB2490 and DD2399 (*KTH*) 7.5 hp – spring 2015

Contact

Lecturers, course coordinators, and examinators:

Lars Arvestad (LA), lars.arvestad@scilifelab.se, office phone 08-52481428
Olof Emanuelsson (OE), olof.emanuelsson@scilifelab.se, office phone 08-52481458.
Lukas Käll (LK), lukas.kall@scilifelab.se, office phone 08-52481196.

Please use email to get in touch, if possible! And make sure you've consulted the **KTH Social** site for BB2490 first, maybe there is already an answer to your question.

Teaching assistants (TAs) (lab-handledare):

Benjamin Sigurgeirsson (BS), benjamin.sigurgeirsson@scilifelab.se
Johannes Alneberg (JA), johannes.alneberg@scilifelab.se

Location

Lectures, seminars, poster session:

FD41, FB52, FB55 – AlbaNova University Center, main building, Roslagstullsbacken 21
SciLife – Karolinska Institute campus, Science for Life Laboratory, Tomtebodavägen 23: gamma building, floor 6, room "G6571".

Computer exercises:

RB33 – AlbaNova University Center, Roslagstullsbacken 33, computer hall

Schedule

Event	Date	Time	Room	Teach.	Contents (brief)
L1	2015-01-20	15-17	FD41	all	Introduction
L2	2015-01-21	10-12	FD41	LA	Genomics: Assembly
Lab1	*2015-01-26*	*13-17*	*RB33*	*BS/KS*	*Statistics*
L3	2015-01-28	08-10	FB55	LK	Statistics of high-throughput experiments
L4	2015-01-29	08-10	FA31	all	Student presentations HT techniques
L5	2015-01-30	10-12	FD41	LA	Genomics: Mapping (*L2)
Lab2	*2015-02-02*	*13-17*	*RB33*	*BS/KS*	*Genomics: Mapping/assembly*
L6	2015-02-03	13-15	FB55	OE	Transcriptomics: RNA-seq (*L5)
L7	2015-02-05	08-10	FB55	OE	Transcriptomics: ChIP-seq (*L6)
Lab3	*2015-02-06*	*08-12*	*RB33*	*BS/KS*	*Transcriptomics: RNA-Seq, ChIP-Seq*
L8	2015-02-09	15-17	FB55	LK	Proteomics: Peptide identification (*L7)
L9	2015-02-11	10-12	FD41	LK	Proteomics: Protein quantification (*L8)
Lab4	*2015-02-11*	*13-17*	*RB33*	*BS/KS*	*Proteomics*
L10	2015-02-13	10-12	FB55	all	Stud.pres. Project information (*L9)
QUIZ	**2015-02-18**	**13-15**	**FB52**	**-**	**Quiz (kontrollskrivning)**
L11	2015-02-19	08-10	FB52	all	Walk-through of Quiz and Project kickoff
S1	2015-02-20	10-12	FD41	all	Project plan and individual diary
S2	2015-03-03	10-12	FB54	all	Exchange of experiences
S3	2015-03-09	13-15	FB55	all	How to prepare poster and poster presentation.
Poster	**2015-03-20**	**10-13**	**SciLife**	**all**	**Poster session.**

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

Reading assignments and lecture contents might be subject to change.