Assignment 2: Sequencing data

In-class discussion on October 31st and November 2nd
Individual write-up due on 16th November
TAs in charge: Christoph and Gemma

Part I. Experimental design

Inspired by the last few lectures, your experience in the lab and the group discussions, write an ‘experimental design’ in which you address the following points:

- Think of a way how the experimental methods mentioned in the course could be applied to the study of autism. Which question(s) could you address with them?

- Imagine that you need to design an experiment that investigates this question. How would you go about it? For example, what kind of biological sample(s) or tissue(s) would you begin with and why? Do developmental time points matter? What are the main steps involved in your experimental design (e.g. RNA isolation, etc.)? What kind of data to you expect to generate (e.g. genomic or expression data, etc.)? How will these data be used to address the question you defined?

When describing your proposed experimental approach, pay attention to writing style and structure, and make sure the questions above are addressed. Focus on concepts, rather than details of experimental techniques; remember that you will not be actually doing this experiment. Also, the specific question you raise does not necessarily have to be related to your first assignment nor to the final project. The aim here is to frame questions that can be reasonably addressed experimentally, and to think carefully about the experimental design so that it fits the postulated question.

General assignment policy: Consider the assignment in your team and flesh out possible approaches and solutions. Make sure that all members of your team are on the same page and able to discuss the issues. While collaboration is highly encouraged, you are required to submit your own individual write-up named as Name_Group_assignmentN to ist.core.homework2017@ist.ac.at by the due date indicated above. Clearly, your group members’ answers will be similar to yours, but you need to formulate your write-up in your own words. Importantly, we apply zero tolerance to plagiarism, which implies not literally copying the information you find, and citing all the sources correctly. There is no minimal length requirement; be concise, but give all the details you consider relevant for explaining the issues to a colleague. Inadequate assignments and assignments handed in behind schedule may have to be resubmitted, which we allow at most twice during this course.
Part II. Sequence alignment

Recall the alignment exercise in class and answer the following questions:

i) Write down a basic algorithm for solving problems of the same type as in-class Exercise 1 (using the output of the sequencer to infer gene expression levels). Your algorithm should be defined precisely, as a series of non-ambiguous simple steps, which can be in the form of a list.

ii) Explain the particular problem posed by your dataset from Exercise 2. Propose and briefly justify modifications to your algorithm from part i) that allow you to work around this problem.

iii) Would your algorithm(s) be practical for larger datasets? How would the time taken increase with the size of the problem (length of the sequence, coverage, average size of the reads,...)?

The suggested length of the full assignment is 3 pages.