1. Describe the steps of a ChIP-chip experiment. For what purposes can such an experiment be used? How is ChIP-Seq different? (6p)

2. a) What is plotted in an MA plot? (2p)

b) Have a look at the MA plot in Fig. 1. Which problems for data analysis can you see in the data set plotted in this figure? How could these kinds of errors have come to existence? (3p)

c) For each problem you listed in b), how would you handle the issue before the actual data analysis? (3p)
3. The figure on the left contains a running-sum plot from GSEA. You can assume that the result from the test is significant.

a) Is there a positive or a negative correlation between gene expression in the gene set and the class labels? How can this be seen from the figure? (2p)

b) Where is the leading-edge subset located? What information can it give? (2p)

c) How can permutations be used to estimate the significance level of the enrichment? How does an average permutation look like when plotting the running-sum? (2p)

4. Explain briefly the following:

a) RPKM normalization (2p)

b) Bicluster (1p)

c) Consensus clustering (2p)

d) Position Weight Matrix (PWM) (2p)

5. BFAST is an example of an alignment algorithm suitable for aligning short reads from current high-throughput sequencing methods. How does BFAST deal with errors in the reads? (5p)