These are sample exam questions. Their purpose is to give you the possibility to check your knowledge and understanding. This is not a comprehensive list. The problems on the exam will be similar but not exactly the same. You are strongly advised to concentrate on the problem solving approaches and methods rather than the specific solutions to these problems.

• Lecture 1

1. What is DNA? What are the DNA bases? What does base complementarity refers to and why is that important? What is RNA? What is the genetic code? How is information coded into DNA and RNA?

2. What is gene expression? Describe the process specifying at each stage what key elements are involved, what such elements do and what the role of each stage is in the overall gene expression process.

3. What are microarrays? What are they used for? What are the main technologies used in microarrays? What are the advantages and disadvantages of each such technology? Describe briefly the microarray process for cDNA microarrays/ oligonucleotides microarrays? Discuss the differences between multichannel and single channel microarrays. What are the data analysis implications of the differences between various technologies?

• Lecture 2

1. What is the general flow of the microarray process? What are the essential pieces of information that need to be stored in each phase of this process? What are the main factors influencing the
microarray data (causes of variability)? Describe and discuss 3 specific problems likely to appear
on a microarray? Describe and discuss what measures can be taken to reduce or eliminate such
effects from a data analysis point of view?

• Lecture 3

1. Describe the main steps of the image processing stage of the microarray data analysis? Define
   and discuss the main issues related to: segmentation, spot localization, signal quantification (in
   various technologies such as cDNA and Affymetrix). In the context of a digital image define
   the terms: resolution, color depth, dynamic range, quantification. What are the factors in the
   image acquisition, processing or quantification stages that can influence the data collected from
   a microarray?

2. A researcher is scanning a cDNA microarray and obtains an image with the following character-
   istics: a few spots are very bright but very many spots are not visible. A colleagues suggests the
   increasing the PMT gain setting on the scanner will many of the missing spots visible. Describe
   what happens if the PMT gain is increased. Is it true that many spots currently not visible might
   become visible? Should this be done?

3. A researcher is scanning a cDNA microarray and obtains an image with the following charac-
   teristics: most of the spots are visible and many are very bright; the background appears to be
   light gray. The researcher proceeds to the image processing and quantification stages and finds
   that most spots appear to be characterized by a high average intensity. Discuss what might have
   happened? What steps would you undertake in order to test your hypothesis and correct the
   situation.

4. Describe the problems that you can observe in the following images. For each such problem
   indicate the probable cause and how this problem will affect subsequent data analysis steps.
   Possible problems are: dust contamination, other contamination, doughnuts, missing spots (pin
   printing problem), imperfect spots (single pin defect), non-specific hybridization, background
   contamination, etc. In order to prepare for such questions, download and examine images from
1. Describe the Affymetrix microarray technology. Define the terms oligonucleotides, average difference, MM, PM, A, P, M calls, etc. Discuss the main issues related to the data analysis of the Affymetrix microarrays.

2. Is it possible to have a gene with an Absent call having an average difference higher/lower than a gene with a Present call? If no, explain why. If yes, give an example and explain what happens.

3. Compare the cDNA and oligonucleotide technologies from a data analysis perspective.

4. What main types of models that can be used to describe a phenomenon? Describe each of them and discuss their advantages and disadvantages? What is a fixed/parametric/non-parametric model?

5. Consider the data coming from set of microarray experiment? You would like to use a t-test to compare the expression values in the experiment group with the expression values in the control group. What type of model would you be using and why? Is this appropriate? Can you use a different type of model for the same problem? When would you do so?