**AP0614: Applied Bioinformatics and Post Genomics**

**& AP0610: Genomics**

**Component 1: Coursework Report on Bioinformatic Analysis of a Cancer Microarray dataset**

**(60% of the total module mark)**

**Semester 2: 2019‐2020**

# Learning Outcome:

Demonstration of understanding the processes underlying the bioinformatic investigation of a cancer microarray dataset. Specifically, ability to interpret and critically appraise the results from (i) subgroup discovery in the data, (ii) characterisation of the differentially expressed genes, (iii) demonstration of robustness of subgroups by application of a classifier to an external dataset.

**Aims**: To describe the results obtained from a multi‐part practical which comprehensively characterised a cancer, transcriptomic microarray set.

# Experimental Summary:

In the workshop on “**Microarray Analysis and application to cancer**”, you discovered subgroups within the medulloblastoma transcriptomic microarray dataset. This was followed by another workshop on “**Gene set identification and annotation**” where differentially expressed genes defining each group were identified, along with the enrichment of different gene ontologies, that will provide clues as to the nature of each subgroup, and what is driving the disease in these tumours. In the third workshop on “**Designing and validating machine learning classifiers**”, you derived a classifier from the differentially expressed genes and validated it in an external dataset.

# Instructions and Information:

* Prepare a report that summarises all aspects of the three workshops – see guidance below for report structure.
* This work should be completed individually.
* Your work should be referenced in the Vancouver style (see Cite them Right if unsure).  Indicative word count: 1500 words.
* Consult the marking rubric below to help you structure your report.
* Marks and feedback will be provided within 20 days of the submission date.

1

# Academic Misconduct:

Plagiarism and collusion are not tolerated under any circumstance and are considered as **serious academic misconduct**. For further detail refer to the ARNA regulations and the student handbook:

It is very simple for staff to identify offenders and unfortunately, staff have done so on several occasions. Once staff detect plagiarism, the perpetrator will be challenged to defend their actions and penalised accordingly. There will always be a marking penalty and in most cases this will result in a fail. It is worth noting that a fail in the second and final years can significantly affect the degree classification you graduate with.

# Word limit guidance (AS rules)

For assessments where a word limit is indicated, a student’s ability to write within the word limit is part of the assessment concerned. Where a word limit is indicated students should provide a final word count by highlighting all text included in the main body of the assessment (the main body of the assessment does not include the reference list) and simply stating that word count. The main body of the assessment includes: the title (if applicable), an abstract (if applicable), the main body of text (including any sub‐titles), in text citations, direct quotations and case studies, tables and figures (including any table/figure titles), illustrations and footnotes. Indicative word count for each section should be proportional to the indicated marks for the section in the rubric on pages 3‐4.

# Write up format:

In this report you will demonstrate that you understand the process and the significance of results you will obtained during the exercise. You don’t have to worry about the R code used in the analyses but you should be able to describe the techniques used and how they are implemented.

Preferred structure for the write‐up:

1. Introduction
2. Subgroup discovery
3. Differentially expressed genes
4. Classifier and its application to external dataset
5. Conclusions
6. References

You will be assessed for a description of each technique to demonstrate understanding, and a critical appraisal of the results. Results should be presented as figures or tables where appropriate, but don’t include every single graph we produced. Try to select the figures that most directly relate to the questions we’ve asked. Don’t be afraid to make a multi‐part figure e.g. Figure 1 A,B,C,D, if all panels relate to the same question (e.g. subgroup discovery). Ensure that your answer includes appropriate references to the supporting literature.

2

Component 1: weighting 60%: Scientific Report based on the interpretation of material delivered as part of computer workshops.

The assessment will see students will be required to critically appraise a transcriptomic microarray cancer dataset to:

• Identify tumour subgroups

• Identify subgroup-specific differentially expressed genes (DEGs)

• Characterise the functional themes of these subgroup-specific DEGs

• Identify a gene signature that can identify subgroups in an external dataset

Computer workshops will provide instruction and training in these techniques.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***COMPONENT*** | |  |  |  |  | **GRADE BAND** | |  |  |  |
| **0‐29%** | **30‐39%** | **40‐49%** | **50‐59%** | **60‐69%** | **70‐79%** | **80‐89%** | **90‐100%** |
| ***1***  ***INTRODUCTION*** | **MARKS AVAILABLE** | **15** | *Trivial*  *engagement*  *with data analysis process* | *Multiple major deficiencies in description of*  *techniques and process* | *Key single major deficiency in*  *description of*  *techniques and process* | *Multiple minor deficiencies in description of*  *techniques and process* | *Single minor deficiency in*  *description of*  *techniques and*  *process or poor*  *representation of outcomes* | *Concise, accurate*  *description of techniques,*  *process and outcomes* | *Interesting presentation* | *Thoroughly engaging in all components* |
|  |  |  |  |  |  |  |  |
| ***2***  Subgroup discovery | **20** | *Trivial*  *engagement* | *Multiple omissions of*  *methodologies and limited*  *description of materials* | *Single omission of key*  *methodology or*  *poor description of materials* | *Poor delivery of discipline*  *presentation*  *conventions or frequent*  *omissions of detail* | *Inconsistent adherence to discipline*  *presentation*  *conventions or multiple*  *omissions of detail* | *Generally accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Concise and accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Virtually flawless* |
|  |  |  |  |  |  |  |  |
| ***3***  Differentially expressed genes | **20** | *Trivial*  *engagement* | *Multiple omissions of*  *methodologies and limited*  *description of materials* | *Single omission of key*  *methodology or*  *poor description of materials* | *Poor delivery of discipline*  *presentation*  *conventions or frequent*  *omissions of detail* | *Inconsistent adherence to discipline*  *presentation*  *conventions or multiple*  *omissions of detail* | *Generally accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Concise and accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Virtually flawless* |
|  |  |  |  |  |  |  |  |

3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***COMPONENT*** | |  |  |  |  | **GRADE BAND** | |  |  |  |
| **0‐29%** | **30‐39%** | **40‐49%** | **50‐59%** | **60‐69%** | **70‐79%** | **80‐89%** | **90‐100%** |
| ***4***  Classifier and its application to external dataset |  | **20** | *Trivial*  *engagement* | *Multiple omissions of*  *methodologies and limited*  *description of materials* | *Single omission of key*  *methodology or*  *poor description of materials* | *Poor delivery of discipline*  *presentation*  *conventions or frequent*  *omissions of detail* | *Inconsistent adherence to discipline*  *presentation*  *conventions or multiple*  *omissions of detail* | *Generally accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Concise and accurate*  *description of*  *materials and methods with*  *minor deviations from*  *presentation*  *conventions or*  *omission of detail* | *Virtually flawless* |
|  |  |  |  |  |  |  |  |
| ***5***  Conclusions | **MARKS AVAILABLE** | **15** | *Limited understanding; poor conclusions* | *Limited identification of*  *value of analyses*  *of the dataset.*  *No ideas for future development.* | *Limited discussion of*  *value of analyses of the dataset.*  *Limited ideas for future development.* | *Some discussion of value of*  *analyses of the dataset. Ideas*  *expressed for future development* | *Clear discussion of value of*  *analyses of the dataset. Ideas and approach*  *expressed for future development* | *Strong statement clearly*  *summarising the findings of data analyses. Ideas and approach*  *expressed for future development* | *Interesting presentation* | *Thoroughly engaging in all components* |
|  |  |  |  |  |  |  |  |
| ***6***  ***GENERAL PRESENTATION & REFERENCES*** | **10** | *Very poor manuscript*  *preparation*  *disregarding author instructions* | *Limited adherence to author instructions* | *Several discrete deviations from author instructions* | *Multiple minor deficiencies in presentation* | *Almost faultless adherence to author instructions* | *Faultless adherence to*  *instructions plus a good use of graphics/tables* | *Great presentation with*  *development of your own graphics/tables* | *An overall visual treat!* |
|  |  |  |  |  |  |  |  |

4