MSc Genomic Medicine

Module 7: Bioinformatics, Interpretation and Data Quality Assurance in Genomic Analysis

2017

Assessment 1: Full Analysis Report - 75% of total module marks

Assessment 2: Molecular Genetic Diagnostic Report – 25% of total module marks

You will be required to write and submit a full analysis report and a molecular genetic diagnostic report based on your analysis of a region of whole exome sequence data and submit using the eAssignments website **www.assignments.soton.ac.uk** by 17.00 **GMT on Thursday, 4th May 2017.**

There is a **penalty** for late submission (see BlackBoard website). Candidates should not leave submission until the last minute, and should undertake this in good time in case of problems. If any problems are experienced with uploading a submission, students should contact the Exams team (**exams.fm@southampton.ac.uk)**. It should be noted that work may only be submitted in the format specified by the eAssignment submission system for each assignment.

The Faculty will be submitting all electronically submitted assignments to analysis by anti-plagiarism software. The University of Southampton considers plagiarism an act of academic malpractice (or academic dishonesty). As such, where proven, it is subject to the implementation of academic procedures. The University of Southampton policy on Academic Integrity can be found at **www.calendar.soton.ac.uk/sectionIV/part8a.html** and should be referred to. There is additional information in the Academic Integrity section within the Study Resources section of our main Blackboard website.

Applying for Special Considerations and Deadline Extension Request

Should you experience difficult circumstances that may have a negative effect upon performance or ability to meet a deadline or to sit an examination, you can apply for special consideration.
If you wish to apply please be aware that requests must be submitted as soon as possible, preferably before the exam or deadline date that may have been affected by difficult circumstances. Requests submitted later than five working days after any assessment or deadline will only be considered if students are able to demonstrate exceptional reasons for the delay. Written evidence has to be provided to support a request. Special Considerations requests which are incomplete or do not enclose evidence in support of that request will not be considered by the Special Considerations Board. Forms together with the evidence should be sent to the Exams and Assessment team (exfm@soton.ac.uk or Mailpoint 801, Southampton General Hospital, Tremona Road, Southampton SO16 6YD or delivered to the Faculty of Medicine, University of Southampton, AB224.

You are also advised to keep copies of all documentation you submit and it's strongly advise you to see your Senior Tutor should you require any further advice. More detailed information can be found here http://www.southampton.ac.uk/assets/imported/transforms/content-block/UsefulDownloads_Download/0E1B9FC4E5544CF9BA10660FE451FFE2/Special%20Considerations%20Guidance.pdf and also within the regulations http://www.calendar.soton.ac.uk/sectionIV/special-considerations.html

Presentation guidelines for electronic submissions

Your assignment should be marked with your student identification number only as it will be marked anonymously.

There is a word **limit** for each assessment:

- **full analysis report** 1500 words not including tables, figures and references
- **molecular genetic diagnostic report** fill in the blanks and short answer questions (500 words, not including tables, figures and references)

Please state your word count on your assignment, which should include the main body of the text excluding titles, subtitles, tables, figures, captions and references. A penalty for exceeding the allowed word count may be applied. All assignments must be pdf converted before submission and the text must be submitted in the following standard format:

  - Arial 11 point font size
  - 1 inch margins left and right, top and bottom
  - 1.5 line spacing with normal spacing of words

Referencing

Candidates must reference their work correctly. The correct usage of referencing is one criterion used to assess work. All Faculty of Medicine assignments should use the approved Vancouver style to reference their work.

If you have any queries about the assessment, please contact the MSc coordinator:

- Rebecca Poole (R.L.Poole@soton.ac.uk)
- Bobbi Moore (B.J.Moore@soton.ac.uk)

Assessment

A male patient aged 16 has been referred for whole exome sequencing (WES). The patient has Low-set, posteriorly rotated ears, a high narrow palate, bilateral cryptorchidism, deep philtrum, short neck, pulmonic stenosis, micrognathia and hypertelorism. Informed consent has been given to report primary and secondary findings.
Use Galaxy, a web-based suite of bioinformatic tools, and methodology from the workshops to analyse the WES data and to detect a causal variant that relates to the patients phenotype. On completion of the analysis, you will write a full analysis report and a molecular genetic diagnostic report based on the assessment criteria (outlined below). Full marks in both assessments will be awarded for concise answers that demonstrate a thorough understanding and, where appropriate, justify the methods used and place the results into context.

Assessment 1: Full analysis report

The full analysis report (maximum 1500 words not including tables, figures and references) will count towards 75% of the final mark and should include the following sections:

1) **Alignment of Next Generation Sequencing (NGS) data to the reference human genome sequence**.
Use this section to describe and evaluate the alignment process and any quality control performed. Include a brief description of the computer programs and software settings that were used. Explain the task(s) performed by the programs and their importance and give details regarding any datasets that were used during the process.

2) **Identification of sequence variation (single nucleotide variants, and small insertions and deletions).**
Describe the process used to call variants including a brief description of the tool(s) used, any important software settings and a justification for the choice of tool(s). Report on the number of variants called and describe and record the metrics that indicate the performance of variant calling.

3) **Annotation of sequence variation with respect to clinical information major data sources of variation (eg Exome Variant Server, dbSNP, 1000 genomes) and predictors of functional significance and pathogenicity (eg: SIFT and Polyphen-2).**
Write a short summary of the annotation process including a description of the tool(s) used, their version number and any important software settings that were used. Describe the databases of sequence variation and predictors of functional significance that you consider most useful for annotation.

4) **Filtering annotated variant call files to identify a shortlist of potentially pathogenic variants with functional relevance.**
Describe the rationale used to filter variants into a shortlist of potentially pathogenic variants with functional relevance and any online resources that were used in the process. Use the phenomizer program (http://compbio.charite.de/phenomizer/) to interpret the patients phenotype and generate a list of candidate genes. Include a table to show the number of variants that remain after each step of filtering. Identify a single variant that you determine is the most likely to be causal and include an IGV screen shot of the variant region. Investigate the gene and predicted functional consequences of the variant and describe these findings and the evidence supporting variant pathogenicity. Comment on any additional work required to determine that the genetic variant is the cause of the patient's phenotype. Repeat the filtering process without restriction to a list of candidate genes and comment on any relevant secondary findings.

5) **Quality assessment of raw sequence data, aligned sequence data, and validity of filtered variants.**
Comment on the raw sequence data and its quality. Describe the data quality for the causal variant. Discuss any limitations of the analysis and any improvements that could be made.

6) **Create a bioinformatic pipeline that can be used to replicate the analysis.**
Include a diagram of the Galaxy workflow in your report with annotation of the steps and details on the input, output and parameters used. Use Galaxy to make your history and workflow(s) accessible via a web link and include these links at the end of your report.

Mark breakdown for the Assessment 1 – full analysis report

1.	Alignment of NGS data to the reference human genome sequence	14
2.	Identification of sequence variation (SNVs, and small insertions and deletions)	14
3.	Annotation of sequence variation with respect to clinical information, major data sources of variation and predictors of functional significance and pathogenicity	14
4.	Filtering annotated variant call files to identify a shortlist of potentially pathogenic variants with functional relevance	14
5.	Quality assessment of raw sequence data, aligned sequence data, and validity of filtered variants	14
6.	Creation a bioinformatic pipeline that can be used to replicate the analysis	5
Total		**75%**

Assessment 2: Molecular genetic diagnostic report

After completing Assessment 1, the next step is to create and populate a Molecular Genetic Test Report. Please use Claustres et al. (2014) "Recommendations for reporting results of diagnostic genetic testing (biochemical, cytogenetic and molecular genetic)" *European Journal of Human Genetics* (2014) 22, 160–170, as a guide (http://www.nature.com/ejhg/journal/v22/n2/full/ejhg2013125a.html). Please use HGVS nomenclature in your report (http://www.hgvs.org/mutnomen/recs.html). The pdf of this paper "Claustres2014_reporting.pdf" may also be posted on Blackboard in the Assessment section. This report carries a total of 15 marks.

To respond to the short answer questions, please read Wallis et al. (2014) "Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics" *Association for Clinical Genetic Science*. (http://www.acgs.uk.com/media/774853/evaluation_and_reporting_of_sequence_variants_bpgs_june_2013_-_finalpdf.pdf). The pdf of this paper "Wallis2014_best_practice.pdf" may also be found on Blackboard in the Assessment section. Each short answer question is worth a single mark except for the last question which is worth 2 marks. Combining with the Molecular Genetic Test Report with these short answer questions results in a total of 25 marks which will count towards 25% of the final mark.

Mark breakdown for Assessment 2 – molecular genetic diagnostic report

1	Molecular Genetic Test Report – create and populate	15
2	Best Practice Guidelines – Short answer questions	10
Total		**25%**

Molecular genetic diagnostic report form

You must create and populate this report form …

SHORT ANSWER QUESTIONS ON BEST PRACTICE GUIDELINES IN THE DIAGNOSTIC SETTING FOR REPORTING GENOMIC VARIATION (Wallis 2014_best_practice.pdf).

1) What should be the policy of laboratories working within the public health sector with respect to the variants they identify (excluding known non-pathogenic variants)?
2) Name one essential practice upon the discovery of a novel sequence variant.
3) When using matched controls to determine whether a variant exists in a population of normal chromosomes what is it essential to consider?
4) What information does segregation analysis provide?
5) What concurrent factors are strong indicators of pathogenicity?
6) When should inter-species comparisons be performed?
7) What is an unacceptable practice when using *in silico* methods to predict pathogenic effects of a previously unclassified variant?
8) Which individuals should receive reports on variants of unknown significance (VUS)?
9) As part of the molecular genetics team, what would you do if you found a second pathogenic mutation unrelated to the primary diagnosis?