Programme Structure Summary

Class Size – there were 19 students on the programme in 2016/17

Programme Structure - the MSc outcome will require 180 credits total (full-time only)

- 60 credits ‘core’ courses - Semester 1: Sept. - Dec.
- 60 credits ‘optional’ courses - Semester 2: Jan. - May.
- 60 credits ‘research project’ - Summer: May - August

The 180 credits can only be made up by doing all these components. Each component comprises several courses that students can enroll themselves onto in this University’s student management system, MyCampus, once they have been accepted onto a ‘Plan’ that allows appropriate course selection.

PGDip and PGCert outcomes

The PG Diploma will require 120 credits i.e. the taught programme

- 60 credits ‘core’ courses - Semester 1: Sept. - Dec.
- 60 credits ‘optional’ courses - Semester 2: Jan. - May.

The PG Certificate will require 60 credits-worth of courses from the taught programme

- for students with no prior programming experience, this must be:
  - 60 credits ‘core’ courses - Semester 1: Sept. - Dec. (you also have to do two exams in the April/May diet)

- for students with prior programming experience, this must be made up from:
  - 30 credits ‘core’ courses (Foundations of Bioinformatics and Omics and systems Approaches in Biology) - Semester 1: Sept. - Dec.
  - 30 credits ‘optional’ courses - Semester 2: Jan. - May.

Programme title and remit

Many of the topics we offer are those that you would find in almost any programme on Bioinformatics or on ‘omics’ analysis - you would expect nothing less, as these areas are a staple in modern biology and relevant computing skills are sought after by employers and supervisors. However, there are several key considerations that have shaped our view of what this programme should offer:

- Some students will primarily want to do some ‘omics’ analysis as a preparation for general life sciences PhD projects - our programme now caters for this more readily
- Some students, on the other hand, want a career in Bioinformatics - our programme is designed with these students in mind and several alumni now have careers in Bioinformatics
- Many researchers know how to do analysis at individual ‘omics’ levels, but fewer understand the added value of integrating over several omics levels simultaneously - this is an added dimension that we feel is important, and it explains the ‘polyomics’ (a.k.a ‘multiomics’) term in our programme title. Glasgow is known for its approach here, and we have a good amount of teaching from the staff of ‘Glasgow Polyomics’, the research organisation and facility recently set up to handle post-genome ‘omics’ experiments and analysis of many kinds at the University of Glasgow
- Although we see it as important that students can step back and take a ‘systems’ view of biological problems, we have no systems biology courses currently within the programme.
The 60 credits of ‘core’ material in Semester 1 are made up of the following compulsory courses:

**Courses delivered by the School of Computing Science:**

20 credit course - **Programming (COMPSCI 4039)**
(Shared with the M.Sc. I.T. Programme)
Course Leader: Dr Alessandro Vinciarelli
Summary of Main Content: Learning to programme in Java. (We would love to be able to teach this in Python, but at the moment there are no suitable courses available - we are working to change that). Covers programming basics, primitive data types, objects; use of algorithms involving repetition and conditional execution; development of well-structured programs and data structures, with attention to maintainable, robust, reliable, and reusable code, appropriate documentation, and thorough testing; solving problems in programming; object-oriented design; human-computer interaction.

10 credit course – **Database Theory and Application (M) (COMPSCI 5076)**
(Shared with the M.Sc. I.T. Programme)
Course Leader: Dr Rosanne English
Summary of Main Content: Provides students with opportunities to develop an understanding of modern methods of managing computerised information. This includes the principles and practice of relational database systems, distributed databases, and NoSQL databases. Mostly taught using PostgreSQL with some MySQL. Covers design of relational database using entity relationship diagrams; use of database management systems (DBMS) to construct and manage relational databases; use of Structured Query Language (SQL) to construct SQL statements for communicating with relational databases; application of normalisation techniques to relational databases to reduce information repetition; handling of database transaction management, concurrency control, and security; analysis of the execution efficiency of SQL to optimise queries across centralised and distributed databases; the different types of NoSQL databases; evaluation of the most appropriate data solution for a given context; examination of issues related to distributed and multi databases.

**Courses delivered by the School of Life Sciences:**

15 credit course - **Foundations of Bioinformatics (BIOL 5170)**
Course leader: Dr Mark Bailey
Summary of Main Content: Molecular biology of the genome, RNA and protein expression products of the genes; what molecular data look like and how to manipulate them, analysis of DNA and protein sequences, phylogenetic trees, the basics of statistical concepts, data distributions and tests as used in bioinformatics, the Unix operating system and use of command line interfaces, doing statistics using the programme ‘R’, programming computers to carry out repetitive tasks by using scripting languages (e.g. Perl), setting up and interacting with databases in MySQL

15 credit course - **Omics and Systems Approaches in Biology (BIOL 5174)**
Course leader: Dr Pawel Herzyk
Summary of Main Content: This course provides a detailed introduction to the experimental design and practice employed in modern ‘omics’ approaches to biology and to ways in which data generated in such experiments are analysed. This covers genomics, transcriptomics, proteomics, metabolomics and systems biology approaches.
For each omics level, we cover:
- the conceptual basis of the discipline
- the instrumentation/machines used
- key applications and research questions that each type of technology can be used to answer
- how to visualise the resulting data using commonly used software packages

A number of areas are also covered that relate to several omics levels simultaneously:
- how different large-scale data sets can be integrated/co-analysed in order to use integration across molecular levels as a basis for biological inference
- the nature and scale of the challenges facing bioinformaticians working with omics data.

In terms of topics, coverage in this course includes: basic concepts in Next Generation Sequencing (NGS), intro to genomics, intro to RNA expression analysis using microarrays, intro to RNA expression analysis using NGS (‘RNA-seq’), protein structure analysis and structure databases, intro to proteomics, intro to metabolomics, intro to how ‘omics’ data are analysed using bioinformatic analysis pipelines, more on stats and omics analysis, intro to mathematical modelling for systems biology.

The 60 credits of ‘optional’ material in Semester 2 comprise the following choices:

The were 13 courses available to choose from in 2016/17 – there is a ‘cluster’ system in operation, so there is a fixed timetable in Semester 2, with 10 credit courses occupying two consecutive 3-week blocks, followed by 2 further blocks of 5 weeks each in which 20 credit courses run (see programme structure summary diagram below). The 60 credits total must therefore comprise two 10 credit courses followed by a 20 credit course, followed by a 20 credit course, from the range available in each slot. The courses are listed by Block and slot below. Several of these optional courses are based around a scientific ‘scenario’ heavily influenced by research carried out within Univ. of Glasgow research groups and involving a large component of handling datasets derived from or inspired by real recent research data analysed by those groups. Students will be exposed to the biological scientific background to the data and the research questions being tackled, to the general experimental approaches and analytical strategy, to the methodology, programs and tools used in the analysis, and to the ways in which biological inferences that answer the initial research questions are made from the results of the analysis; they will also get substantial hands-on experience of multiple stages of the analytical pathway. In these cases, the teaching will mainly be carried out by group leaders and research staff involved in the research that forms the basis for each scenario.

Please note that the mix of courses available in 2017/18 may not be identical to this list, as some courses may not run, if staff have left the University, and other courses may be approved between now and the start of the session.

- **Identification of Disease-Causing Genetic Variants (BIOL 5300)**
  
  Course leader: Dr Mark Bailey
  
  Summary of Main Content: - A hands-on tour of the computational challenges involved in mapping genetic variants that predispose to traits of interests in organisms of importance to health, agriculture and farming. It will cover genetic data handling and manipulation, the principles behind linkage disequilibrium and its role in genetic association analysis, the key stats underlying experimental evidence for genetic association, all steps in the
pipeline from raw genotype data to analysis output, interpretation of genetic association analysis output data to infer whether evidence for association has been found and the strength of the evidence, the principles involved in doing genetic association analysis at the genome-wide level in a GWAS design, and the steps involved in going from an association ‘hit’ to evidence for function of a variant and its role in disease predisposition.

- **Drug Discovery (BIOL 5222)**
  
  Course leader: Dr Alan Bilsland

  **Summary of Main Content:**
  
  The course aims to provide students with a critical understanding of the various stages involved in the pre-clinical drug discovery process. Students will learn to critically evaluate published data and appraise the current methods and strategies used for drug discovery.

  The course aims to provide students with a critical understanding of the various stages involved in the pre-clinical drug discovery process with regard to the requirements, methods, challenges and limitations for identifying and validating ‘druggable’ targets, for identifying, validating, and optimising new compound leads, and the importance of pharmacokinetic/pharmacodynamic (PKPD) profiling and toxicology testing, as well as wider health economic aspects. The knowledge gained will enable students to critically evaluate literature on current methods, techniques, and strategies used for drug discovery, and to appraise their advantages and disadvantages for targeting a specific disease. The course will cover the main phases and decision points of the pre-clinical drug discovery pipeline; how current drug families target different types of biological pathways and activities, and the concept of a “druggable” target; the methods and strategies used for target validation, lead identification and optimisation, and in vivo efficacy testing; using data generated as part of the processes of target validation, lead optimisation and efficacy testing; design of additional experiments or tests to complement existing data for a particular drug; current issues and problems related to the drug discovery and development pipeline from an industrial and health economic perspective.

- **Animal Models of Disease (BIOL 5238)**
  
  Course leader: Dr Alan Mowatt

  **Summary of Main Content:** The course aims to provide students with a critical understanding of the technologies and techniques used to develop animal models of human inflammatory and infectious disease. Students will understand the place of animal models in exploring disease pathogenesis and therapy, and will discuss the ethical issues relating to the use of animals in medical research.

  The course covers comparative anatomy; the anatomy of the immune response; technologies used in the development of new animal models for human inflammatory and infectious diseases - gene targeting, transgenesis, genome editing etc.; technical approaches used in evaluating animal models (e.g. imaging, omics approaches, flow cytometry); comparison and evaluation of widely used animal models of human inflammatory (e.g. rheumatoid arthritis, Crohn’s disease, multiple sclerosis, asthma) and infectious disease (e.g. malaria) and how these models have provided crucial mechanistic insights; the ethical basis for the use of animals in medical
research and role of animal models in drug discovery and development

- Using Chemical Structure Databases in Drug Discovery for Protein Targets (CHEM 5042)

  Course leader: Dr Adrian Lapthorn

  Summary of Main Content: The drug discovery pipeline is now heavily dependent on the *in silico* screening of small molecule chemical library databases at a number of points along the route. This course will describe the biophysical and bioinformatics procedures used to identify lead and lead-like chemical compounds based upon their properties and interactions with protein target binding sites. The course shows how protein sequence data and 3D protein structure and biophysical data can be integrated to inform drug design and the drug discovery workflow. The course will cover modelling interactions between proteins and small molecules of relevance to drug development, data coding formats and organisational structure of chemical databases, interactions with such databases, database mining, sequence and structural data manipulation, formulation of search criteria based on chemical and protein features, atomic structure coordinates, files and databases, recognition and prediction of active sites and binding sites and generation of images of protein with ligand bound in the active site, modelling of interactions *in silico* and the limitations of this approach, the principles underlying manipulation of protein structures using docking algorithms to show molecular interactions design and implementation of *in silico* screening protocols to select candidate ligands from libraries online, constraint rules for the selection of potential small molecule ligands, construction of potential interaction maps, construction of virtual compound libraries, and identification of potential lead molecules for drug development. Students will have the chance to put many of these analysis concepts into practice during extensive computer lab practicals and will develop analytical skills, practical database construction and other computing skills and the ability to assess critically, and in the appropriate biological context, procedures for integrating structural data concerning proteins and small molecule ligands.

- RNA-seq and Next Generation Transcriptomics (BIOL 5177)

  Course leader: Dr Pawel Herzyk

  Summary of Main Content: This course will examine transcriptomic NGS data analysis. Students will get hands-on experience of how such RNA-seq data are generated, manipulated, and analysed both bioinformatically and statistically, using a scenario based around real datasets emerging from studies conducted within Glasgow Polyomics. It will cover RNA-seq in more detail than in the core course above, and then go on to cover statistical approaches to RNA-seq data analysis, a comparison of different analysis methods and analysis of their performance, application of a range of software tools, use of genome browsers to visualize RNA-seq results, interpretation of statistical analysis results by mining annotation databases, and use of Perl and R tools to construct transcriptomics data analysis pipelines.
- Omic Analyses for the Biomedical Sciences – from Genomics to Metabolomics (BIOL 5197)

Course leader: Dr Richard Burchmore

Summary of Main Content: - This course provides an introduction to workflows for the resolution and characterisation of complex mixtures of biomolecules, from DNA to small molecule metabolites. The course will emphasize the potential and challenges of omic approaches and will include data handling tasks and demonstration.

The course aims to provide students with a critical understanding of a range of modern “omics” technologies and applications. The course will introduce students to genomic, transcriptomic, proteomic and metabolomic techniques, and the analytical approaches that can be employed to examine the data output from these approaches. The relative benefits and challenges of each -omic approach will be presented, with exemplar data sets. Examples of the application of omic approaches in a variety of relevant biological systems will be presented, to give students an appreciation of the type of output generated and of typical strategies for data analysis and interpretation.

Learning objectives relate to information flow in biology, the relative pros, cons and challenges associated with the different levels of -omic data collection, strategies for characterisation of a genome/transcriptome/proteome/metabolome, workflows to identify, quantify and characterise molecules in complex mixtures, the importance of and strategies for validating proteomic data, identification of appropriate applications for different omic approaches, and designing an experimental strategy to exploit an omic analysis and the importance of controls and validation.

- Bio-Imaging for Research Scientists (BIOL 5261)

Course leader: Dr Francis Burton

Summary of Main Content: - This course provides an introduction to the science of imaging and image analysis in life sciences. The course includes lectures and practical classes covering range of imaging modalities.

The course aims to provide students with both knowledge and practical skills in biological image analysis (BioImaging). Students will be introduced to a wide range of techniques and technologies for collecting images of cells and biological tissues. Tuition will be given in the appropriate design of protocols and appropriate methods of quantitative image analysis.

Learning objectives relate to: use of ImageJ for 2D and 3D image processing and analysis, design of effective and appropriate imaging experiments and analysis protocols, structure, capture and storage of digital images, the use of fluorescent probes and issues surrounding their selection, and the advantages and disadvantages of different imaging modalities.

- Clinical Genomics (MED 5425)

Course leader: Dr Maria Jackson

Summary of Main Content: - This course provides an overview of the clinical applications of genomic approaches to human disorders, particularly in relation to clinical genetics, discussing the methods and capabilities of the new technologies. Tuition and hands-on experience in data analysis will be provided, including the interpretation of next generation sequencing reports.
The course covers the structure and function of the human genome (e.g. enhancers, epigenetics, chromatin remodelling, non-coding regions and the ENCODE project); diagnostic analysis of the human genome using standard and high-throughput technologies (current standard techniques - MLPA and aCGH; and exome and genome sequencing using NGS approaches) - their capabilities and limitations; clinical bioinformatics - making sense of the data (quality checks, trios, gene panel filtering, PHRED scores, PED and BED files etc.); data analysis and evaluation of the pathogenicity of single nucleotide variants (SNVs) using online prediction algorithms based on sequence conservation and on biophysical properties of amino acids, Grantham matrix etc.; familiarity with additional terminology (e.g. BAM files, CRAM files, BCFs, VCFs, SNVs, CNVs, and many more); ethical aspects of whole genome analysis (including issues around the reporting of ‘incidental findings’); stratified medicine, pharmacogenomics and consequences of genomic variants for therapeutic drug responses and adverse effects.

- **Pathogen Polyomics (BIOL 5299)**

  Course leader: Prof. Mike Barrett

  Summary of Main Content: This course will provide hands-on experience of next generation sequencing, proteomics and metabolomics analysis in the context of investigations into the mechanisms by which a human parasite, *Leishmania*, develops resistance to drugs used to treat it. The course will specifically show how integrative analysis across these omics levels can be utilised to reveal more about the biology of this system than each level can alone. In this course, genomes of parasites resistant to drugs are compared with those of their wild type progenitors. We then use software that mines metabolomics data to identify metabolic changes between the parasites. Correlation analysis applied to the genetic and metabolomics data then allows identification of those changes associated with the resistance phenotype. The material covered will cover hands-on next generation sequence analysis, the analysis of whole genome sequence data, sequence read assembly and data mining to learn about the functional attributes of genomic sequences, proteomics data analysis and software, informatics approaches to the problem of large-scale protein identification, analysis to identify differences in protein abundance between samples, informatics approaches to metabolite identification from metabolomics data, ways to infer the components and connectivity of metabolic networks within cells, ‘polyomics’-style integration of omics approaches to enrich biological inference, modelling of candidate protein structures and bound ligands.

The **60 credit project (MSc Bioinformatics, Polyomics and Systems Biology Project; BIOL 5173)** is carried out in the summer months after the April/May exam diet and students are embedded within their project supervisor’s research group

Aims: The project aims to equip students with a range of advanced skills in relation to the formulation, planning and practicalities of short research projects. It will give students an insight into how research is carried out in one or more of the fields of bioinformatics, omics analysis and systems biology, and practice in selected research techniques and data analysis approaches. It will give students the opportunity to develop and enhance their analytical and organizational skills through hands-on research and the opportunity to develop their ability to work independently. It will provide practice in the writing of substantial project reports.
Programme structure diagram and timetable blocks  
(New timetable, 26/04/17)

**Semester 1**  
60 credits

- **Block 1**  
  **3 wks**  
  Jan. 2018  
  - BIOL 5170** Foundation of Bioinformatics  
    - 15 CR  
    - Mark Bailey  
    - SLS  
    - **and**  
    - COMPSCI 4039 Programming  
    - 10 CR  
    - Alejandro Vinzantelli  
    - SCS

- **Block 2**  
  **3 wks**  
  Feb. 2018  
  - BIO 5530 Omics and Systems Approaches in Biology  
    - 15 CR  
    - Pawel Herzyk  
    - SLS  
    - **and**  
    - COMPSCI 5076 Database Theory and Application  
    - 10 CR  
    - Roseanne English  
    - SCS  
  - Students must do all 4 of these courses in Semester 1

**Semester 2**  
60 credits; all students can choose from any of the courses in each of these blocks

- **Block 1**  
  **3 wks**  
  Jan. 2018  
  - BIO 5310 Identification of Disease-Causing Genetic Variants  
    - 10 CR Blended/Shareable  
    - Block 1  
    - Mark Bailey  
    - SLS

- **Block 2**  
  **3 wks**  
  Feb. 2018  
  - CHEM 5042* Using Chemical Structure Databases in Drug Discovery for Protein Targets  
    - 10 CR Blended  
    - Block 2  
    - Adrian Lapthorn  
    - SChem

- **Block 3**  
  **5 wks**  
  Feb./Mar. 2018  
  - BIOL 5177* RNA-seq and Next-Generation Transcriptomics  
    - 20 CR  
    - Block 3  
    - Pawel Herzyk  
    - SLS

- **Block 4**  
  **5 wks**  
  Mar./Apr. 2018  
  - BIOL 5299* Pathogen Polyomics  
    - 20 CR Blended  
    - Block 4  
    - Mike Barrett  
    - III

**Summer**  
60 credits  
**14 wks**  
May-Aug. 2018  
- BIO 5173* MSc Bioinformatics, Polyomics and Systems Biology Project  
  - 60 CR  
  - Mark Bailey  
  - SLS

* Indicates courses that only MSc Bioinformatics students take

**Students must pick ONE of these courses in Block 1**

**Students must pick this course in Block 2**

**Students must pick this course in Block 3**

**Students must pick this course in Block 4**
Learning and Teaching Approaches Used in This Programme

- lectures (both didactic and interactive)
- computer practical classes, software demonstrations and exercises
- guided independent study
- individual and group presentations
- field trips
- small project work
- a substantial project (MSc only)

Assessment Methods Used in This Programme

- coursework essays and written exercises
- written examinations - essay questions, short answer questions and small problem analysis questions
- computer laboratory reports
- written reports on small projects
- group and individual presentations
- computer lab performance during the main project (MSc only)
- written report on the main project (MSc only)
- viva examination on the main project (MSc only)

Study protocols, types of teaching and timetables:

- For a 10 credit course, 100 hours of work by the student are assumed, including time spent in private study and time spent preparing for assessments and in assessment.

- For a 20 credit course, 200 hours of work by the student are assumed, including time spent in private study and time spent preparing for assessments and in assessment.

- For the 10 credit courses, there will be generally be about 15-25 hours of contact time with teachers (for the 20 credit courses, it will be about 25-40 contact hours). Types of teaching will vary between courses, but most will have at least 5-10 hours of lectures and tutorials/student presentations and at least 5-10 hours of supervised computer practical. Students will be expected to do many more hours on the computer themselves working from worksheets etc., with staff and demonstrators on hand to supervise progress at intervals, and course leaders helping to round things off at the end of each course.

- In Semester 1, we have to fit in with the School of Computing Science timetable and their two courses are long and thin, running for a few hours each week over most of the semester. The two School of Life Sciences courses therefore operate similarly.

- In Semester 2, the 10 and 20 credit optional courses are ‘short and fat’ - the teaching takes place over 3 weeks (10 credit courses) or over 5 weeks (20 credit courses) during the teaching term (January to early May), and you are only studying one course at any one time (you may be working towards assessment hand-in deadlines for previous courses also, though). Where there are optional courses up against each other in the same timetable slot, any course with fewer than 5 students enrolled will probably not run.

- The project in the summer lasts for 14 weeks. Students are embedded in the research group of their supervisor and the project occupies them full-time. Although meetings with the supervisor and team will be regular, a high level of independence in the design and execution of the work is expected (increasing as the project goes on) from the student. Projects may be carried out within Glasgow, elsewhere in the UK, or in any approved lab worldwide.
Rationale for the Programme - what needs are we hoping to meet?

The rationale behind the programme included the following thoughts about what students on the programme and its graduates will need (in addition to other factors thought important by the academic staff contributing to the programme):

What do the students want their own attributes to be after the degree Programme?
- generic skills pertinent to employability in any employment sector
- practical computing skills, including programming
- a solid conceptual foundation for data analysis in post-genome biology
- knowledge of programming languages they could use again in a range of jobs, or in further study
- competency (via hands-on experience) in a range of bioinformatics and analysis tools commonly used in post-genome biology

What are the unique selling points (USPs) and key sellable features of our Programme?
- the scenario-based optional course system, which deeply embeds real, local research stories in the teaching
- the high quality projects where students are embedded in what may be world-leading groups in their fields
- the fact that over the last few years, graduates of this programme have had a high rate of success in getting PhD Studentships or employment as core Bioinformaticians