

Bioinformatics facility

Summary

The Bioinformatics facility comprising two Bioinformatics experimental officers and an academic lead aims to support research across the Faculty of Health and Medical Sciences by providing advice in experimental design and costing for research bids, and skills in data processing, handling and analysis. This document outlines: 1) the terms of reference for the facility, 2) the skills and resources available, and 3) access to the facility and the prioritisation of tasks.

1. Terms of reference

Staff members of the Bioinformatics facility will:

- Provide experimental design, data processing and handling, and analysis support to researchers accessing the Bioinformatics facility as outlined in Section 3.
- Establish and maintain analysis pipelines, including associated documentation, working towards standard operating procedures (SOPs) for routine (frequently requested) analyses.
- Provide training to researchers performing routine tasks/analyses and/or where there is a demand for a given topic, provide half-day workshops.
- Assess developments within the field of high throughput data analysis (e.g. processing standards of RNA-seq) via regular literature searching and attendance of conferences. Significant changes, changes likely to have an impact on any existing or future pipelines/SOPs, will be reported back to researchers accessing the facility via written reports.
- Liaise with faculty IT to maintain the required computing and storage hardware and software of the Bioinformatics server(s).

Where a contribution to the Bioinformatics server(s) has been made (e.g. through direct allocation of funds to increase server capacity via a research grant) the facility will also ensure that data is stored and backed up in keeping with the University's data management policies, as supported by faculty IT.

Where no contribution has been made to the Bioinformatics server(s) researchers must seek their own long term solution for the storage and management of their data. In this instance, data will be moved to the Bioinformatics server(s) only to complete the required tasks and will be deleted upon project (task) completion.

The head of the Bioinformatics facility will:

- Act as line manager for the staff members of the Bioinformatics facility.
- Regularly meet with the staff members of the Bioinformatics facility to ensure the needs of the team and the faculty are met.
- Provide guidance on the development of pipelines and SOPs that are of interest and/or strategic priority to the faculty.
- Consult and liaise with researchers and members of the Bioinformatics facility to address identified issues in the provision of data analysis e.g. issues in the prioritising of work and/or misunderstandings.
- Provide costings or facilitate costings for grant bids which require support from the Bioinformatics facility.
- Oversee and manage the budgets associated with the Bioinformatics facility.
- Oversee and manage the software licences associated with the Bioinformatics facility.

- Oversee the 'billing' for work carried out to external (to the faculty) partners/contractors.
- Prepare reports on the activity of the Bioinformatics facility, as requested by HoS.
- Provide justification for staff additions and/or further training of staff members of the facility.
- Liaise with faculty IT to ensure the maintenance of the required computing and storage hardware and software of the Bioinformatics server(s).

2. Bioinformatic resources and skills available

Resources available to the Bioinformatics facility include:

- Two FT staff members.
- 8 Linux servers comprising 2 - 64 CPUs, 1-128GB RAM and 16 -15000GB of storage, with this capacity increasing over time, based on the contributions from research grants.
- Licenced software; Metacore (GeneGO) for the analysis of biological systems.

Experienced in processing of data from:

- DNA- and RNA-seq, including Genome assembly, GWAS, SNP analysis, differential expression, peak identification, metagenomics
- Microarrays, including differential expression, ChIP-on-chip, genotyping/SNP
- Metabolomics data from mass spectrometry instruments
- Phenomic data from Biolog
- Clinical measurement data

Analytical skills:

- Exploratory data analysis (incl. PCA, descriptive statistics, histograms, boxplots)
- General Data mining (incl. Clustering, multidimensional scaling, artificial neural networks, fuzzy logic)
- Data visualization (incl. custom made and standard visualisation plots)
- Statistical analyses (incl. mixture models, ANOVA, Kaplan-meier estimate, power calculations)
- General sequence analysis and manipulation (alignment, motif search)
- Machine learning and classification (incl. short time-series, linear models)
- 'Systems analysis' (flux balance analysis of genome scale metabolic networks (GSMNs), and general modelling and simulation using tools such as SurreyFBA, QSSPN)

Note: The skills highlighted above are not exhaustive and the team welcomes the chance to expand and enhance their skills to support the faculty. If you would like to discuss your own data/required skills please do get in touch at bioinformatics@surrey.ac.uk.

3. Access

Requests for bioinformatics support can be sent to bioinformatics@surrey.ac.uk, where the team endeavour to respond to emails within 24 hours during the working week. Note that requests to support a new project should come from the PI, where information regarding the source of funding and allocation to the facility (if applicable) is required. This information is used to prioritise requests (see below), and facilitates the monitoring of the facility's activities and budget.

All publications or presentations that include work that has been supported by the bioinformatics facility **must** acknowledge the facility according to the following format: "I/We thank the Bioinformatics facility at University of Surrey; [name of staff member] for [tasks/analyses performed]".

Prioritising Requests

The Bioinformatics facility prioritises and categorises requests based on the following definitions:

I. Request related to a funded project work, highest priority

Where funds have been directly allocated to the Bioinformatics facility, either as a percentage of the team's time included in a grant or directly costed and paid for via work to external (non-RCUK) funders. This work may comprise both routine work and novel, ad-hoc analyses. Funded projects receive top priority and the bioinformatics team will block-book time to meet the requirements of these projects/requests.

To ensure fair management, allowing for all groups to receive efficient and effective service, the team require that PIs communicate with the team so that:

- 1) The team are aware that funds have been awarded and thus a percentage of the team's time needs to be allocated to the project.
- 2) In advance, the team are aware of the likely date that the data will become available, project starts and/or set deadlines for given objectives to be completed.

Work falling into this category is considered 'part of the service' of the bioinformatics facility and therefore authorship on any publications arising from the team's work is down to the PIs discretion (acknowledgement is required, as detailed above). As a suggestion, where the work performed by the team forms a critical aspect of a publication, on which conclusions of the manuscript are predominantly based, the offer of authorship is recommended.

For those planning a future project, please see Category III below.

II. General technical advice, high priority

Requests related to general technical advice, such as how to run specific software and/or which statistical analysis is the most appropriate, will receive a timely response (within 24 hours during the working week).

III. Request related to a future project : research proposal support, medium priority

Power calculations, Preliminary data analysis and Experimental design, where the outputs of such work will directly support the submission of a bid. Timing of such requests should be no less than **4 weeks** in advance of bid submission.

Those planning on submitting a bid for a project that will rely on bioinformatics analysis/expertise in the future should discuss their requirements **at least four weeks** ahead of grant submission. This step will ensure the data can be handled effectively and that the data generated (and bid) is of highest quality (i.e. experimental design/power calculations are appropriate). **Failure to engage with the bioinformatics team, to discuss future projects and/or experiment, may result in the team's inability to analyse the data.**

IV. Request related to an unfunded project involving routine analyses, low priority

Requests of work on projects having no access to funds, such as an undergraduate or postgraduate project, or analysis of legacy data. Unfunded projects receive the lowest priority and the bioinformatics team will only work on such projects *following* their work on higher priority projects. Tight deadlines for turnaround cannot be imposed for unfunded projects, the meeting of specific deadlines is the responsibility of the PI and should be discussed with the team in advance.

Unfunded routine work is classed as work requiring the use of existing pipelines and/or scripts taking no more than 3 working days to complete. Such routine work is seen as 'part of the service' and will not require authorship (it is at the PIs discretion) but does require acknowledgement (see above).

Where a PI requires repeated routine work for a single project it is expected that the bioinformatics team will train the researcher(s) (student(s)) on that project to enable them to carryout repeated analyses independently (with the bioinformatics team providing technical support only). Half-day workshops are possible where there is demand for a given topic from a number of researchers.

V. Request related to an unfunded project involving novel or ad hoc analyses, lowest priority

Requests of work on projects having no access to funds, such as an undergraduate or postgraduate project, or analysis of legacy data. Unfunded projects receive the lowest priority and the bioinformatics team will only work on such projects following their work on higher priority projects. Tight deadlines for turnaround cannot be imposed for unfunded projects, the meeting of specific deadlines is the responsibility of the PI and should be discussed with the team in advance.

Unfunded novel or ad hoc analyses that require more than two working weeks to complete are not seen as part of the service and will be considered as a collaborative project. The collaboration needs to be discussed with the team, including viable turnaround/timings, given the team's higher priority workload, and will be *mutually* agreed. Collaborative work requires authorship on any resulting publication, where authorship position is dependent upon the percentage of the manuscript relying on the bioinformatics work, and will be agreed between the PI, the staff member(s) of the Bioinformatics facility and head of Bioinformatics facility.