Institute for Apprenticeships & Technical Education

BIOINFORMATICS SCIENTIST

Reference Number: ST0649

Details of standard

Bioinformaticians are scientists - specialists who use computational, data analytical and data mining techniques which are applied to a range of problems in the life sciences, for example, in pharmaceutical companies in the process of drug discovery and development. Roles require scientists who understand life sciences, and who can work computationally with diverse and large volumes of data derived from different life science activities - and role names and descriptions often reflect this by using slightly different names for what is broadly the same computational skill-set. For example, bioinformatics, computational biology, computational toxicology, Health informatics, Medical informatics, Agri-informatics. This range of titles reflect the importance of life-science-specific knowledge coupled with the underlying (and sometimes specifically-adapted) data science, statistics and computational skills.

Broadly, bioinformatics is: Research, development, or application of computational tools and approaches for expanding the use of life science, (inc. biological, chemical or health) data, including those to acquire, store, organise, archive, analyse, or visualise such data; in such a way that aids development and application of data-analytical and theoretical methods, mathematical modelling and computational simulation techniques to the study of such biological systems.

A bioinformatician is often part of a collaborative group or team of scientists, drawing together life scientists, statisticians and computational infrastructure specialists. Consequently, the bioinformatician must be able to work across these disciplinary boundaries. The main duties of such a role would include the ability to:

- Work as part of an interdisciplinary team to support life science experiments from the design stage through to data analysis and biological interpretation.
- Develop suitable plans for the storage and management of biological data, including annotation and metadata, and implement these through upload to public repositories and/or implementation of local databases.
- Work with private and/or public bioinformatics resources, taking into consideration the legal, ethical or confidentiality aspects of their use.
- Critically evaluate raw biological data and prepare it for suitable analysis.
- Determine the most suitable method for computational analysis, considering biological and experimental relevance, and where necessary build and test analytical pipelines or write and test new algorithms.
- Analyse biological data and to interpret the result in the context of the experimental design and, where necessary, in a broader biological context through integration with complimentary (often public) data.
- Carry out all data handling, processing and analysis with consideration of commercial practices and guidance including documentation, reproducibility regulated processes and intellectual

property (IP).

- Communicate the results and their context to interdisciplinary scientific peers and participate in dissemination through writing papers and giving talks. Further aid the understanding of others through supervision and training.

Requirements: Knowledge, Skills and Behaviours

The Bioinformatics Scientist will have knowledge and understanding of:

1. A topic aligned with the life science field, and the core experimental platform/data generating technologies in the chosen field.

2. How research is conducted in bioinformatics and within the broader context of interdisciplinary life sciences.

3. The technical limitations and the underlying biological and experimental assumptions that impact on data quality.

4. Details of omic-scale/big-data-driven life science making use of core platform technologies.

5. The responsibilities of working in production/industry environments managing scientific data – including regulated environments (good practice, and IP/confidentiality requirements).

6. Current approaches for modelling and warehousing of life science data.

7. Requirements for responsible, legal or ethical access and use of biological data, including general data protection (GDPR) considerations, identifiable personal genomic & healthcare data, and geographic biodiversity-related data concerns.

8. Ontologies and their use.

9. Retrieval and manipulation of biological data, including data mining, from public repositories.

10. Techniques to integrate, interpret, analyse and visualise biological data sets.

11. Bioinformatics analysis methodologies and expertise in common bioinformatics software packages, tools and algorithms – including workflow management tools.

12. Common bioinformatics programming languages; algorithm design, analysis and testing.

13. The use of suitable version control tools, software sustainability practices and open source software repositories.

14. Licensing limitations on the use of bioinformatics software and data such as open source, commercial and academic usage restrictions.

15. Database design and management, including information security considerations and big-data technologies.

16. Relevant big-data and high performance computing platforms including Linux/Unix, local and remote High Performance Computing (HPC), and cloud computing.

17. Application of statistics in the contexts of bioinformatics and life science data analysis.

18. Statistical and mathematical modelling methods, and key scientific and statistical analysis software packages.

19. General data science approaches to life science problems, such as machine learning and artificial intelligence (AI).

20. The importance of data governance, curation, information architecture and ensuring interoperability.

21. Differences in the knowledge-base of diverse audiences, and the most appropriate means of effectively communicating scientific and technical information.

22. Communication models and techniques which can be employed in a collaborative research environment to effect change at individual, team and organisational level eg. active listening skills, teamworking, influencing and negotiation skills.

A Bioinformatics Scientist can:

23. Work with multi-disciplinary colleagues to design life-science experiments that will generate data suitable for subsequent bioinformatics analysis.

24. Provide guidance to experimental scientists on data generation methodology and handling to ensure the quality of data produced.

25. Recognise and critically review the format, scope and limitations of different biological data.

26. Define the required metadata to be collected for specific datatypes and analytical approaches.

27. Design and implement appropriate data storage formats and associated database structure.

28. Choose appropriate computational infrastructure and database solutions - including internal or external/cloud resources.

29. Store and analyse data in accordance with ethical, legal and commercial standards, including checking who has access.

30. Curate biological data using suitable metadata, ontologies and/or controlled vocabularies.

31. Make use of suitable programming languages and/or workflow tools to automate data handling and curation tasks.

32. Maintain a working knowledge of a range of public data repositories for biological data.

33. Prepare data for submission to appropriate public bioinformatics data repositories as required, being aware of IP and/or ethical and legal issues.

34. Carry out data pre-processing and quality control (QC) to prepare datasets for bioinformatics analysis.

35. Determine the best method for bioinformatics analysis, including the selection of statistical tests, considering the research question and limitations of the experimental design.

36. Identify and define appropriate computing infrastructure requirements for the analysis of such biological data.

37. Apply a range of current techniques, skills and tools (including programming languages) necessary for computational biology practice – and;

38. Contribute to (where appropriate, lead) research to develop novel methodology.

39. Build and test analytical pipelines, or write and test new algorithms as necessary for the analysis of biological data.

40. Document all data processing, analysis and implementation of new methods in accordance with good scientific practices and industry requirements for regulatory process and IP.

41. Interpret the results of bioinformatics analysis in the context of the experimental design and, where necessary, in a broader biological context through integration with complementary (often public) data.

42. Obtain data sets from private and/or public resources – considering any legal, privacy or ethical aspects of data use.

43. Carry out the analysis of biological data using appropriate programmatic methods, statistical and other quantitative and data integration approaches – and visualise results.

44. Communicate and disseminate bioinformatics analysis and results to a range of audiences, including multi-disciplinary scientific colleagues, non-scientific members of management, external collaborators and stakeholders, grant/funding bodies and the public as required.

45. Supervise and mentor colleagues and peers to develop bioinformatics knowledge relevant to their specific life science subject experience.

46. Communicate orally and in writing, and collaborate effectively with interdisciplinary scientific colleagues, and management functions to monitor and manage people, processes or teams.

47. Manage their own time through preparation and prioritisation, time management and responsiveness to change.

A Bioinformatics Scientist will be mindful of:

48. Professional standards in the workplace in relation to: ethics and scientific integrity, legal compliance and intellectual property, respect and confidentiality, and health and safety.

49. The need to continuously develop their knowledge and skills in relation to scientific developments that influence their work, ensuring they continue to provide relevant analyses, including emerging techniques where appropriate.

50. The ongoing need for awareness of technical advances in the broader scientific field that may present opportunities for personal and / or organisational development.

51. The wider context (policy, economic, societal, technological, legal, cultural and environmental) in which scientific research operates, recognising the implications for professional practice.

52. The need to be enthusiastic, self-confident, self-aware, empathic, reliable and consistent to operate effectively in the role.

53. The requirement to persevere, have integrity, be prepared to take responsibility, to challenge areas of concern, to lead, mentor and supervise.

Duration

It is anticipated that the duration of the apprenticeship will be typically 30 months.

Entry Requirements

Individual employers will set the selection criteria, however in most cases applicants will have a background in a life sciences subject or informatics/computer science. Apprentices without Level 2 English and Maths will need to achieve this level prior to taking the endpoint assessment. For those with an education, health and care plan or a legacy statement, the apprenticeship English and Maths minimum requirement is entry Level 3 and British Sign Language qualifications are an alternative to an English qualification for whom this is their primary language.

Qualifications

On completion, apprentices will hold a Level 7 Master's Degree qualification (a MSc., MSt., MRes., or a MSc by Res. depending on the provider institution) in Bioinformatics aligned to a science or technology discipline relevant to their job role.

Level

This is a Level 7 degree apprenticeship

Review Date

Three years after approval of the standard.

Find an apprenticeship

Postcode (optional)

Version log

VERSION	DATE UPDATED	CHANGE	PREVIOUS VERSION
1	29/01/2019	Funding band first published - standard now approved for delivery	Not available
1	18/12/2018	Assessment plan first published	Not available
1	17/05/2018	Standard published	Not available