Bioinformatician / Data Scientist

Location: Hinxton, Cambridgeshire.

Job Type: Full-time

Work Model: Hybrid working. Some time required on-site

About Broken String Biosciences Ltd

Broken String Biosciences is a Cambridge-based genomics company with the goal of developing safer cell and gene therapies by assessing the stability of the genome. We are leveraging state of the art genomic sequencing technologies to better understand the factors that determine on and off-target activity during CRISPR genome editing, with the vision to build a technology platform that will drive the development of CRISPR gene therapies that are safer by design. Starting with our first patented technology, INDUCE-seq, which for the first time enables precise digital mapping of DNA double strand breaks in the genome, Broken String Biosciences is building a suite of genomics technologies that will assess all elements of genetic safety during the genome editing process. This will provide the information required to design CRISPR gene therapies that are fundamentally safer, ensuring that this potentially life-saving technology can be delivered to patients.

With offices based at the BioData Innovation Centre (BIC) on the world-renowned Wellcome Genome Campus, this is an exciting opportunity to join a well-funded and fast-paced start-up in the CRISPR and genomics space at a very early stage.

Position overview

We are seeking a **motivated and talented Bioinformatician / Data Scientist with expertise in statistical modelling** to process and interpret high-throughput sequencing data generated by the Broken String laboratory team. This data will be used to characterise the induction of DNA double strand breaks (DSB) by gene editing for a wide range of cell- and gene therapies. The patented INDUCE-seq technology produces genomic DNA break data that is highly specific and quantitative, providing a unique understanding of the consequences of gene editing against the background of endogenous break events. The successful candidate will have the opportunity to apply their knowledge of bioinformatics and computational biology to gain insight into this data that allows our partners to better understand the efficacy and safety of these therapies currently in development. Joining an integrated team of laboratory, bioinformatics and machine learning scientists, you will work closely with colleagues and industry-leading collaborators to build reproducible pipelines to analyse the data and apply classical statistical methods, in addition to machine learning approaches, to assess the significance of break induction. This role will leverage access to rich data sets from previous and on-going collaborations, applying biological knowledge and statistical rigour to gain a better understanding of the mechanisms and outcomes of a range gene editing technologies.

You will be joining an energetic and forward-looking team applying cutting-edge laboratory, bioinformatics, and data science techniques to help improve the next generation of cell- and gene therapies.

As a **Bioinformatician / Data Scientist** you will initially focus on three main aspects of the company's activities.
Firstly, you will work with other bioinformaticians to develop, run and interpret the outcomes of reproducible pipelines that translate raw sequence data into human-readable outputs tailored to customer and collaborator requirements.

Secondly, you will apply your expertise in statistics to the outputs in order to better evaluate the significance of events identified within the data.

Thirdly, the role will require you to work closely with others in the data science team to generate data sets that will be utilised for the development of deep learning models. Additionally, you will assess the performance of these models.

You are the ideal candidate if you are looking to deploy your technical and scientific skills within an ambitious start-up and have impact as a part of a growing team.

Role & Key Responsibilities

- Build bioinformatics pipelines using best practice software engineering to process NGS data in the cloud.
- Predict genome editing off-targets introduced by CRISPR-Cas9 and other related technologies using an existing pipeline
- Develop novel algorithms to be able to distinguish different classes of DSBs.
- Prepare data for visualisation and secondary analysis, as well as evaluating the statistical significance of these results.
- Assist with building and analysing the outputs of machine/deep-learning models

Essential Competencies

- An MSc or equivalent industry experience in computer science, bioinformatics or computational biology.
- Experience in applying bioinformatics to omics datasets.
- A working knowledge of workflow managers for processing data using approaches that are both parallelisable and high throughput.
- Advanced Python programming skills.
- Familiarity with data formatting, reshaping and visualisation methodologies using Python (pandas, seaborn) and/or R libraries (tidyverse, ggplot2).
- An in-depth knowledge of statistical approaches as applied to the evaluation of biological datasets. In particular a thorough knowledge of how to apply statistical tests of significance and the appropriate distributions to use to achieve this.
- An excellent communicator, both verbal and written, with the ability to collaborate with a multidisciplinary team across different fields and convey concepts to external collaborators.

Desirable Competencies

- A PhD or equivalent industry experience in computer science, bioinformatics or computational biology.
- Experience with CRISPR genome editing methods and data.
- Familiarity with cloud computing, preferably AWS.
- A working knowledge of the software development life cycle including version control, testing and CI.

Salary: Competitive with benefits and participation in company share scheme
Reporting to: Head of Bioinformatics