VACANCY – DATA SCIENTIST / SENIOR DATA SCIENTIST

About Enhanc3D Genomics

Enhanc3D Genomics is a functional genomics spinout company from the Babraham Institute (Cambridge, UK) leveraging a disruptive technology to profile three-dimensional (3D) genome folding at high resolution. Understanding DNA organisation and long-distance interactions allows to link gene enhancers and non-coding genetic variants to their target genes and holds promise to unlock disease-related genetics for therapeutic discovery.

http://enhanc3dgenomics.com/

Team
Investors: Start Codon, Bioqube Ventures, Private Investors
Founders: Prof. Peter Fraser, Dr. Stefan Schoenfelder, Dr. Mikhail Spivakov
CEO: Dr. Debora Lucarelli
Chairman: Dr. Frank Armstrong

Role Description

Enhanc3D Genomics is an innovative and dynamic company with diverse and highly engaged staff. We believe in fostering great teamwork to maximise our collective skills and experience. We are passionate about realising the power of 3D genomics by developing new cutting-edge technologies for therapeutic discoveries.

We are looking for a forward thinking, enthusiastic, highly motivated **data scientist / senior data scientist** to join the technology team tackling complex data analysis problems across the whole spectrum of Bio R&D. Enthusiasm and passion to work as part of multi-disciplinary teams to solve research problems combining data science and machine learning expertise are keys to this role.

The ideal candidate will have demonstrated experience in developing and deploying new data processing pipelines and machine learning models, as well as analysing, visualising and interpreting complex biological data. A good understanding of data mining and machine learning are key to this role. You will be involved in collaborative projects with academic and commercial research groups to deliver high-quality data reports and develop ideas and technologies, contributing to our IP portfolio. You will be flexible and keen to support the technology development while championing quality in all aspects of your work.

An organised professional with exceptional interpersonal and communication skills, enabling you to successfully operate in a business environment where confidentiality and discretion are paramount.

Key accountabilities
- Perform data analyses using the Enhanc3D Genomics proprietary pipeline
- Produce reports for internal research and external collaborations

- Process, analyse and interpret NGS data
- Work with wet-lab scientists during experimental design and analysis phases, including understanding their goals, providing statistical support, and assisting with querying public resources (databases, literature) to contextualise and improve findings
- Algorithm design and development, data processing, statistical analysis and visualisation of complex and large datasets
- Programming and statistical data analysis and visualisation experience in R or Python with one or more common ML frameworks (Tensorflow, PyTorch, etc…)
- Familiarity with common bioinformatic databases and tools
- Communicate clearly and openly with stakeholders to ensure we are a data-driven organisation
- Enhance data collection procedures to include information that is relevant for building analytic systems
- Write protocols, risk assessments, documentation and keep track of development software (e.g. Git) and analyses
- Work across teams and proactively engage in knowledge sharing and peer support, including training
- Attend meetings and provide written and verbal reports for team and projects developments
- Flexibility to perform other duties which may be required

Qualification and Skills

Essential	Desirable
For the senior post a PhD + postdoc, or equivalent experience with formal training and proven expertise, in a data science or biotech engineering	Knowledge of continuous integration workflows and platforms within Data Science. Understand the importance of versioning
Excellent knowledge of Python and R	Knowledge of version control and collaborating with developers using GitHub
Experience with Next Generation Sequencing platforms and / or OMICS datasets	Interest in and knowledge of Cybersecurity
Experience with and understanding of Hi-C, capture Hi-C, HiChIP	A social personality that contributes to an open, positive, collaborative working climate, and a strong desire to drive personal development
Proficient with common data science libraries/toolkits such as Pandas, SciKit-Learn, Numpy, Keras etc. Experience with Data Visualisation, Analysis, Warehousing tools (Looker/ Stitch / DBT / Snowflake or similar)	Proven experience of working in a fast-paced research-driven commercial environment or in collaboration with industry

Good Understanding of Data Mining and Feature Engineering techniques	Ability to deliver quick wins alongside implementing more advanced modelling techniques
Good Understanding of statistics i.e. statistical distributions, hypothesis testing	A strategic, inquisitive and innovative mindset with the ability to sense commercial opportunities in our data and exploit accordingly
Highly motivated and innovative scientist with effective influencing and problem-solving skills	Delivery skills with the ability to work well under pressure in an agile environment
Excellent oral and written communication skills	Technical expertise in either bioinformatics, computational biology, biostatistics, mathematical modelling or data science
Excellent organizational and record keeping skills	Experience building data science models for B2B SaaS platform
Good Understanding of Machine Learning Pipelines/Algorithms, particularly unsupervised/semi-unsupervised learning techniques	Experience with pipeline development and software virtualisation (e.g. docker, singularity)

For more information please contact Debora Lucarelli dlucarelli@enhanc3dgenomics.com.