Genomics Data Curator – Maternity Cover

Congenica is a pioneering digital health company enabling genomic medicine, revolutionizing the way genetic diseases are characterized and diagnosed, and providing life-changing answers. Our end to end solutions, from sample to report, enable you to deliver world-class genomic medicine services and make important clinical decisions that transform the lives of patients and their families.

The Congenica product presents all the information that a Clinical Scientist needs to investigate the genomic evidence of rare diseases. A key aspect of the system is the UI that selects from the extremely large volumes of data associated with a genome sample, combines it with complex reference data and displays it to the user in an easy to access fashion.

Job Purpose

Congenica is applying machine learning to genomic data to advance the capability of its clinical decision support software. The machine learning team is delivering AI automation that will have a huge impact on patients' lives.

Genomics Data Curator is responsible for selecting, acquiring and preparing data that inputs into the machine learning model. Working in an immediate team of 3 (including you), you will be accountable for the delivering of clean, well-prepared data sets to our ML engineer and ML data scientist to enable them to be immediately highly productive. You will be to primary person for all data preparation and selection requirements for the machine learning colleagues, setting the highest possible standards of data curation.

Main Responsibilities

- Prepare data from monthly SQL database refreshes and external data sets to enable model retraining, applying and improving QA processes
- Debug data quality issues to enable improvements to the model's performance

- Manipulate data from a variety of sources, working within our own cluster and in external environments where necessary. (edited)
- Analyse data to inform the team of the viability and suitability of data for certain challenges
- Research and recommend new data sets that can help us solve specific challenges
- Apply for data access, managing the written application process and access requirements
- Document your work comprehensively in the company's wiki
- Build trusted relations with colleagues, partners and stakeholders

Employee Profile

Essential attributes candidate must have on entering the role

Knowledge, Skills & Abilities

- Skilled in preparing, integrating and manipulating high-quality, very large (TBs) genomic data sets for further analysis.
- Deep understanding of the field of genomics and the data gathered.
- Skilled in researching and reviewing suitable data sources to solve a challenge.
- Experience with PostgreSQL/ relational databases.
- Exceptional attention to detail.

Behavioural Qualities

- Energised by being part of an innovation team with a specific commercial remit to achieve.
- Able to work under time pressure, clearly prioritising work to achieve goals.
- Self-motivated and results-driven, problem-solver.
- Friendly, approachable and builds positive personal and organisational relationships
- Excellent written and verbal communication.
- Enthusiastic, hardworking, well organised and able to prioritise.
- Continuous learning and application and sharing of new skills and knowledge.
- Professional appearance and able to produce high standards of work

Desirable attributes candidates may have on entering the role

Qualifications

- Masters degree in genomics or similar scientific field

Knowledge, Skills & Abilities

- Knowledge of machine learning and statistical tools that will be using the data once prepared (e.g. Tensorflow, R).
- Working with AWS.
- Programming in Python.
- Knowledge of data privacy and security requirements.
- Knowledge of statistics.

Related Experience

- Start-up or scale-up experience.
- Predictive model building experience.