



<b>Number of Positions:</b>	1-3
<b>Job Description:</b>	Data Scientist, AI Specialist, Bioinformatician
<b>Location:</b>	-Virginia (Blacksburg or in close proximity is preferred) -Telework option will be considered for the right candidate(s)

---

#### About Us:

GATACA, LLC is a new-frontier biotech company--located in SW Virginia near Virginia Tech-- that develops advanced bioinformatics tools for virologists to identify mutating variants of disease-causing viruses from sequence data. We are a nimble, dynamic team designing and developing algorithms and a pipeline to support the computational side of virology research and clinical investigations of pathogenic viruses. Working alongside a prestigious group of world-renowned scientists, we have tailored our pioneering bioinformatics pipeline—GAT (GATACA Assembly Tool) to be virus-specific. In future iterations, following successful deployment of its virus implementations, GAT will be adapted to other disease indications (e.g., cancer), applications (e.g., clinical trial screening), multi-omics data, functional genomics and structural bioinformatics to drive new target evaluation, and clinical research.

We are looking for an experienced data scientist and bioinformatician to join our group. This opportunity describes either a single position (for the multi-talented candidate) or multiple positions. The position(s) will provide the candidate(s) with an opportunity to apply their computational and mathematical skills to develop innovative methods for analyzing complex virology data to identify novel mutant strains, targets and biomarkers associated with diseases. Successful candidate(s) will possess significant experience in data science research including mathematical modelling, machine learning (ML) and statistical data mining techniques, and will be seasoned in bioinformatics analytics such as assembly, alignment and phylogenies.

***We prefer one multi-talented person to fill this project-focused job. However, we realize this is a very niche skill set, so please do not shy away from applying if you are uniquely qualified for one or the other of these specialties! We will consider all qualified candidates for full- and/or part-time work.***

#### Project Overview:

The successful candidate(s) will continue on-going development, testing and optimizing of bioinformatics tools in the GAT pipeline, and will develop machine learning (ML) modules. The candidate(s) will identify and curate relevant datasets to build a database tailored to specific ML training tasks, and will develop at least three ML modules previously designed by our team to work with the assembly algorithms in GAT in an integrated fashion.

The development environment is in Linux, and geared toward large-capacity Linux servers.



### **Principal Responsibilities:**

The principal overriding outlook is to maintain a vision to expand all pilot steps toward long-term software goals. To this end, the candidate(s) will be in regular correspondence with the team through all stages of conception and development, and will be responsible for:

- Implementing, modifying and optimizing existing tools and developing novel analytical approaches to analyze Next Generation Sequence (NGS) samples, both simulated and patient-derived.
- Creating and managing simulated NGS datasets from existing reference sequences, including mixed ratios of different strains. Maintain and expand knowledge of high-throughput genomic data sources.
- Integration and meta-analysis of public and internal data sets to generate testable hypotheses.
- Being skilled in or open to learning to identify diverse multi-layered omics data types (e.g. genomics, transcriptomics, proteomics) that would augment or complement NGS in identifying rare, hidden or novel mutations, novel drug targets, biomarkers and therapeutic mechanisms.
- Lead the development of predictive models, using mathematical, ML/AI and statistical approaches, for biomarker identification and development.
- Support and/or assist the bioinformatics team in implementing innovative statistical methods and data integration approaches. Support and/or assist the Architect to integrate all developed tools.
- Staying up-to-date with novel analytical methodologies, tools and applications.
- Contributing to the general development of GATACA's bioinformatics platform and capabilities.

### **Essential Skills:**

- Degree or significant experience in numerical discipline (applied mathematics, computational statistics, physics, data science, engineering, computer science or similar).
- Knowledge of and experience with genomics, DNA files and NGS (essential for Bioinformatics).
- Knowledge of and experience with ML methods and Natural Language Processing (essential for Data Science).
- Knowledge of and experience in Deep Learning methods (essential for Data Science).
- Knowledge of and experience with searching and retrieving genomic data (e.g., PubMed).
- Programming proficiency with Python.
- Experience in developing algorithms and workflows.
- Critical and curiosity-driven thinking to solve problems as they arise.
- Excellent written and oral communication skills, and ability to communicate with computational and experimental scientists.

### **Desired Skills:**

- Knowledge and experience with NGS files are recommended for ML tasks, though not essential if ML skill set and experience are high.
- Experience with NGS in mixed population samples a plus (e.g., multi-strain pathogens; metagenomics, etc).
- Analytical approaches for analyzing beyond NGS (e.g., multi-omics and multimodal data sets).
- Demonstrated working knowledge of statistical methods.



*We offer learning and career development opportunities as well as a highly competitive salary.*

To apply, please either apply directly on our website ([www.gatacallc.com](http://www.gatacallc.com)) --OR--send your CV and a cover letter clearly addressing your skillset to [jcraig@gatacallc.com](mailto:jcraig@gatacallc.com)

*We are recruiting into 1-year fixed term contract position(s), to expand to permanent position(s) depending on successful completion of projects and on qualifications*

For information on GATACA, LLC please visit: [www.gatacallc.com](http://www.gatacallc.com).