**Project #1: Bioinformatics Aspect of Assay Development**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about to improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. Lots of learning while having a good time is guaranteed!

Project Description: Come learn the bioinformatics side of developing sequencing assays used for drug development. You’ll work with assay development lab members and bioinformaticists to evaluate sequencing data and determine if the assay is suitable for intended use. The range of projects are suitable for students with basic programming skills to advanced computing skills.

Applicant Skills:

* Interest in learning bioinformatics and statistical methods.
* Familiarity with next generation sequencing assays and outputs.
* Familiarity with unix or a programming language will be helpful but not required.
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.

**Project #2: Optimization of Generative AI tools for Method Validation Reports**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: Q2 Solutions develops genomics assays for use in clinical trials. As part of this, we create official documentation about the performance of our assays and their acceptability for the intended use. These Method Validation Reports contain a large amount of information that would be useful for our product development teams and sales to query. In addition, writing these reports is a significant time burden that could be eased by an AI tool to generate draft reports based on examples of relevant sections from previous reports and incorporating data from standard outputs generated by in house software for analysis of assay performance. The applicant would work with the mentor to expand, assess, and improve the performance of generative AI tools for querying and/or writing Method Validation Reports.

Applicant Skills:

* Understanding/course work on one or more: computer programming, statistical modeling, or machine learning.
* Familiarity working in Linux/UNIX environment, including shell scripting and bash (preferred).
* Proficiency in Python
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.
* Exposure to next generation sequencing, multi-analytic instrumentation or other \*omic data preferred.
* Experience with natural language processing, large language models, retrieval augmented generation, or similar is preferred, but not required.

**Project #3: Docker creation for software**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about to improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: Come learn about dockers with us and help us develop a Bioinformatics docker toolkit. Docker packages an application with all its dependencies and makes it easy to run a software on a server without compatibility issues. You’ll learn about advantages of dockerization, best practices for dockerization, and help us dockerize a variety of bioinformatics programs that we use regularly. You can use the knowledge to upgrade your pipelines with docker container calls.

Applicant Skills:

* Understanding/course work on one or more: computer programming, statistical modeling, or machine learning.
* Familiarity working in Linux/UNIX environment, including shell scripting and bash (preferred).
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.

**Project #4: Reference Gene detection from RNA-seq Datasets**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about to improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: The applicant will gain R skills while helping us re-write an existing script to make the script dynamic. The applicant will learn about RNA-sequencing datasets, outlier detection, normalization of expression data, and basic statistical analyses. The result of the project will be identification of a set of reference genes with low standard deviation. Script may need to be dockerized.

Applicant Skills:

* Understanding/course work on one or more: computer programming, statistical modeling, or genomics
* Familiarity with R programming language
* Familiarity working in Linux/UNIX environment, including shell scripting and bash (preferred) – if interested in dockerization
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.

**Project #5: High Confidence Regions of the Human Genome under Different Sequencing Conditions**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about to improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: The applicant will examine existing benchmark genomic regions from NIST, T2T and HG38 references, other information sources (such as RepeatMasker), and other information such as mapping scores and determine optimized high confidence regions for different modes of sequencing using Illumina technology. For example, with FFPE samples and MRD samples, genomic inserts are usually 150b +/- or even less in length. For PCR-based WGS, inserts can be 200-240b +/- in length. For PCR-free WGS, inserts can be 300b or longer +/- in length. This output will be used to filter VCF files or filter reads prior to variant calling as false positives greatly increase in lower confidence regions when using short-read sequencing as even short reads can vary substantially in length due to the nature of the tissue, its preservation, and protocol. In this case, we want to be able to measure insert length characteristics and then filter variant candidates appropriately given this measurement, or filter reads prior to variant calling.

Applicant Skills:

* Understanding/course work on one or more: computer programming, statistical modeling, or genomics
* Proficiency in a programming language: Python or R
* Familiarity working in Linux/UNIX environment, including shell scripting and bash (preferred).
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.

**Project #6: Ensemble Calling of Structural Variants**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: Structural variants, including large (>50 bp) insertions, deletions, duplications, inversions, and translocations are challenging to detect accurately with short read sequencing. We have assessed several individual structural variant calling tools, but each has weaknesses. One strategy to improve sensitivity of structural variant calling is to combine the results from multiple tools; however, simply taking all results can lead to a large increase in false positive calls. The applicant would help determine a strategy for combining outputs to create a final dataset that has improved sensitivity while minimizing the impact to false positive rate. This may include simple voting or filtering strategies, or more complex statistical methods, based on the applicant’s experience. The goal is to make an initial attempt at ensemble structural variant detection and benchmark the results on well characterized samples, comparing to our previous results for individual tools.

Applicant Skills:

* Understanding/course work on one or more: genomics, genetics, next generation sequencing. Statistical modeling or machine learning a plus.
* Exposure to next generation sequencing, multi-analytic instrumentation or other \*omic data
* Familiarity with a programming language: Python, R, or bash
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.

**Project #7: Comparison of Structural Variant Benchmarking Depending on Reference Genome**

About Q2 Solutions: Q2 Solutions is a leading clinical trial laboratory services organization passionate about improving world health. The translational genomics bioinformatics department provides analysis solutions for insights from genomic data such as biomarker discovery, DNA or RNA signature identification and interpretation, immune landscape, pan-viral pipelines, pathway analyses, and more.

Purpose: The chosen applicant will gain practical work experience within an industry setting comprising individuals of diverse talents across computer science, molecular biology, statistics, and genomics. He/she will be part of a dynamic bioinformatics team, advancing the state of art in genomic data analysis with multiple projects available to apply their research and analytical skill sets to help us advance the field. The applicant will also learn principles of software development in an industry setting and the art of transitioning from research to production environments.

Project Description: Structural variants, including large (>50 bp) insertions, deletions, duplications, inversions, and translocations are challenging to detect accurately with short read sequencing. In developing assays for use in clinical trials, high sensitivity and low false positive rates are essential for our clients. We have assessed our WGS assay for structural variant calling performance using well characterized samples, with our standard reference genome, hg38. However, most of the samples were characterized by third parties using the hg19 reference genome. We have seen significant differences in perceived performance of small variant calling depending on the reference genome, due to differences such as failed or inaccurate lift-over, or variant alleles from one reference genome becoming the reference allele in another version of the reference genome. The applicant would repeat the benchmarking analysis for structural variants using hg19 and identify and help explain the differences between reference genomes that may cause issues when benchmarking structural variants. This will improve our understanding of the complexity of structural variant benchmarking and help us to improve our methods for assessing performance of our assays.

Applicant Skills:

* Understanding/course work on one or more: genomics, genetics, next generation sequencing.
* Exposure to next generation sequencing, multi-analytic instrumentation or other \*omic data
* Familiarity with a programming language: Python, R, or bash
* Experience giving presentations of analysis results to non-technical audiences.
* Appropriate verbal and written communication skills to function within a professional work environment.