Public Health Laboratory Recruitment Guide:

Bioinformatics Scientist Positions

www.aphl.org

Summary

While education level can contribute to how good a candidate might be, years of experience in data mining, pipeline construction and next generation sequencing (NGS) analysis are really what differentiates a beginner versus an expert in bioinformatics. At a minimum, a candidate should have prior experience performing NGS data analysis. Unless you are planning to only use commercial software packages such as Geneious or CLC Genomics Workbench, knowledge of Unix operating systems and coding experience is essential. As such, we recommend including a few questions regarding past coding projects, use of bioinformatics tools and pipeline construction during recruitment. You also need to be aware that prokaryotic and eukaryotic genomics are very different in terms of NGS analysis; candidates that have only conducted analysis involving humans (such as cancer) may have difficulty transitioning to bacterial genomics (and vice versa). When advertising positions, tailor them to your needs, aims and budget.

You will likely receive applications from different academic backgrounds. An applicant who majored in computer science would be strong in software development and database management and should be able to easily deal with system administration and information technology (IT) issues, but may lack biological knowledge and have to rely heavily on laboratory staff for pipeline validation or determining what is biologically relevant in the data. At the other end of the spectrum are candidates coming from biological sciences who have picked up bioinformatics along the way during their career; these candidates may understand the biological significance of the data (e.g., phylogenetic analysis, antimicrobial resistance, etc.) and be able to communicate well with clinical laboratories, but may not be as efficient in terms of programming skills and dealing with IT issues. Candidates with a degree in Bioinformatics were educated with a mix of computer science and genetics course work, giving them strong technical knowledge of bioinformatics in general; however, these programs do not always provide practical experience that allows students to apply their skillset to different biological questions and public health problems.

Enclosed are bioinformatics position descriptions, based on the various competencies around sequencing data quality control, data analytics and pipeline design, with which bioinformaticians should be expected to be proficient. In addition to bioinformatics skillsets, experience in training and communication across different areas of expertise are also important for a bioinformatician in a public health laboratory setting.

All materials in this guide are meant to be a template for a job posting, but can and should be modified to fit your laboratory's specific needs and resources. Download a Word version of this guide for easy editing.

Position Description

Program/Department: **Public Health Laboratory**
Position Title: **Bioinformatics Scientist**
Reports to: **Laboratory Director / Operations Director / Division Director**
Previous Incumbent: **None (New position)**

Job Position Summary

The bioinformatics scientist within the public health laboratory will be in charge of analysis of NGS data produced by the Center. The incumbent is expected to apply state-of-the-art computational methods and assist in the development of innovative approaches to software challenges that touch all aspects of NGS data analysis. This includes, but is not limited to: pathogen identification, outbreak tracking, mapping drug-resistance determinants, development of novel genetic tests, variant analysis, genome annotation and phylogenetics. The analysis may involve microbial, viral, fungal, parasites and human NGS data. In addition, he/she may be expected to serve as the statistics resource for the Center's research and regulatory programs. Coding expertise, efficient communication skills, basic knowledge of microbiology, training and outreach are necessary requirements for this position.

Laboratory Infrastructure Requirements

- Access to a high-performing computing (Linux-based) is essential for performing high-throughput bioinformatics analysis. This can be achieved through the purchase of a local computer cluster, collaboration with a local university or agency that houses high-performing computing, or through cloud access such as Amazon AWS, Microsoft Azure or Google Cloud Platform.
- High volume storage capabilities since many bioinformatics analyses tends to generate significant amounts of disk data. This data should also preferably be mirrored or backed up to prevent any accidental losses of information.
- High bandwidth and sharing capabilities to easily transfer data and analytic results between the interested parties.
- The bioinformatics scientist should ideally have administrative privilege on the high-performing computing to facilitate the installation and update of the multiple software used for analyzing NGS data.

Essential Job Duties

- Maintain quality assurance/control of NGS data generated by the Center to ensure reproducibility, sensibility and accuracy of the downstream bioinformatics analyses. This includes the verification of the overall quality score of the sequencing reads, presence of adaptors, carry-over and contamination.
- Conduct bioinformatics analysis of NGS data using community-accepted or independently verified bioinformatics tools and development of pipelines for high throughput analysis.
- Use best practices for code development and documentation.
- Educate, train and provide guidance to in-house staff on the interpretation of results.
- Guide the development of policies and procedures for clinical test implementation that require bioinformatics analysis.
- Summarize and communicate results in a comprehensive way for clinical staff, scientific meetings and journal publication.

Job Position Competencies

Data Quality Assessment (10%)

1. *Assess quality of raw sequencing reads.*
   Identify and implement tools to evaluate overall quality score of the sequencing reads, presence of adaptors, carry-over and contamination.

2. *Maintain and ensure appropriate sequence reads quality prior to downstream analyses.*
   - Identify and implement tools to execute quality control, including but not limited to trimming of the sequencing reads, adaptor removal and contamination sequencing reads removal.
   - Setup and deploy guidance for re-sequencing when the raw sequencing data quality is not satisfied.

3. *Evaluate sensitivity and specificity of bioinformatics analysis results.*
   - Develop evaluation plans for bioinformatics analysis sensitivity and specificity.
   - Select appropriate data, including simulated data, to conduct sensitivity and specificity test, integrate the results into the evaluation reposts.

4. *Implement controls in the bioinformatics analysis and pipeline construction to fulfill the Clinical Laboratory Improvement Amendments (CLIA) requirement.*
   - Implement positive analysis controls in the bioinformatics analysis and pipeline construction
   - Implement negative analysis controls in the bioinformatics analysis and pipeline construction.
   - Set up, evaluate, maintain and update positive and negative controls.

Data Analysis and Pipeline Construction (50%)

1. *Develop and implement existing and new computational methods and tools for NGS data.*
   Demonstrate knowledge of and utilize current best-practice tools for genome assembly, quality control, containerization and workflow management.

2. *Write scripts using common coding languages such as Python and/or R.*
   - Develop in a Linux environment using shell scripting and bash syntax.
   - Use cloud computing platforms (AWS, GCP, Azure) and run pipelines in a cloud environment.

3. *Deploy existing pipelines and perform required analysis.*

4. *Utilize data visualization tools and develop reports for superiors and partners.*

5. *Interact directly with other scientific staff to ensure appropriate interpretation of results in a scientific context.*

Workforce Training (10%)

1. *Training design*
   - Develop needs assessment questions and documents.
   - Set up reasonable overall training goals and timeline based on the evaluation of a needs assessment.
   - Integrate principles of adult learning for use in designing training.
   - Collaborate with subject matter experts to gather content.

2. *Training content development*
   - Identify necessary training topics and tools/software based on the learning objectives.
   - Assemble instructional materials including knowledge principles, data, online resource, demo, etc.
   - Set up needed resources (i.e., facilities, computers, software, software licenses, etc.) prior to the training.
   - Integrate multiple types of training materials into training design.
   - Tailor training topics and materials to the knowledge differences of various learners.

4. *Training implementation*
   - Apply principles of learning to training implementation and delivery.
   - Use the most effective presentation tools and techniques.
   - Answer training-related questions raised by learners in a timely manner.
   - Use examples, exercise, questions, etc., to increase the knowledge absorbance of learners.

5. *Training evaluation*
   - Develop tools and questions to evaluate the overall training, as well as learner knowledge and skill development. Integrate training goal assessment into the training evaluation.
   - Track and compile evaluation data into summative training reports.
   - Evaluate and analyze training reports. Summarize the participants' achievement of training objectives and the training's overall strengths and weaknesses.
   - Develop an improvement plan based on results of the training report evaluation.

Communication (20%)

1. *Communication techniques*
   - Deploy formal written and oral communication strategies.
   - Apply logical structure to written communications.
   - Tailor and translate bioinformatics discipline-specific terms, concepts, results, etc., into accessible information for non-bioinformatics audiences in both written and oral communications.
   - Consume questions, requests and other communication from non-bioinformatics audiences and transfer the information into bioinformatically-actionable items.

2. *Communication technology*
   - Utilize technology to communicate information to internal and external partners.
   - Use designated technology for sharing information.
   - Revise and improve communication technology when needed for maximum communication efficiency.

3. *Communication professionalism*
   - Display professional demeanor in all situations with professional peers.
   - Determine and select essential information to share based on communication recipients.
   - Prepare professional written reports (e.g., memos, abstracts, posters, scientific manuscripts, etc.) and oral presentations (e.g., meetings, conferences, training presentations, etc.) by preparing drafts of all public communication materials.

Computer System Maintenance (10%)

Education and Experience Requirements

Entry-level Positions

Supervision
Will need close supervision.

Required Skills/Expected Tasks
Able to perform basic QA/QC analysis and utilize commercial software packages to perform SNP calling, genome assembly, limited phylogenetic analysis and whole genome analysis. May be able to work at the command line level to perform some bioinformatics analysis or build pipeline with limited capabilities.

Qualifications
Bachelor's degree with no or limited experience.

Intermediate-level Positions

Supervision
Will need some supervision and guidance during the design and development phase of projects.

Required Skills/Expected Tasks
Able to develop, test and implement custom bioinformatics pipelines. Capable of analyzing a wide range of data from different sources. Excellent knowledge of at least one scripting language and Unix, and may be able to work with statistical packages such as R. Can provide some guidance to the clinical labs on requirements needed (such as type of sequencing or number of reads needed) to perform the bioinformatics analysis.

Qualifications
Bachelor's degree with 2 or more years of experience OR Master's degree.

Advanced-level Positions

Supervision
Will need little direct supervision.

Required Skills/Expected Tasks
Can develop complex bioinformatics solutions with no or limited supervisions. Can handle multiple projects at any given time. Will act as a resource for guidance during the design, development and testing phase of projects. Can work independently or be an integral part of a research team. Will participate in the production of high quality scientific communication and grant proposals. Can train and lead lower level bioinformatics staffs.

Qualifications
Master's degree with 2 or more years of experience OR PhD.