ASSOCIATION FOR MOLECULAR PATHOLOGY
Providing global expertise in molecular testing that drives patient care
6120 Executive Blvd., Suite 700, Rockville, MD 20852
Tel: 301-634-7939 | Fax: 301-634-7995 | amp@amp.org | www.amp.org

August 24, 2020

Heather Stang, MS, MT, Division of Laboratory Systems
Centers for Disease Control and Prevention
1600 Clifton Road NE
Mailstop V24–3
Atlanta, GA 30329

Attn: Docket No. CDC–2020–0051.

Comments submitted electronically at www.regulations.gov

Dear Ms. Stang:

Thank you for the opportunity to submit these comments in response to the Request for Information (RFI) Concerning Personnel and the Retention of Next Generation Sequencing Data in Clinical and Public Health Laboratories. AMP is an international medical and professional association representing approximately 2,500 physicians, doctoral scientists, and medical technologists who perform or are involved with laboratory testing based on knowledge derived from molecular biology, genetics and genomics. Membership includes professionals from the government, academic medicine, private and hospital-based clinical laboratories, and the in vitro diagnostics industry.

AMP actively advocates for the modernization of the Clinical Laboratory Improvement Amendments (CLIA) regulations to ensure that oversight of laboratory medicine aligns with the current state of the field. For that reason, we commend the Agency for initiating this discussion within the Clinical Laboratory Improvement Advisory Committee (CLIAC) and for providing this comment docket.

Utilization of large data sets, including those derived from next generation sequencing (NGS), requires an investment in bioinformatics expertise, and infrastructure to manage, store, and access the data appropriately. AMP is pleased that the RFI recognizes the importance of NGS-based testing in the diagnosis, monitoring, and treatment of patients with multiple conditions and diseases and that the Agency understands the role of NGS in many areas of medicine, including pediatrics, oncology, and microbiology as well as the essential role of informatics in analyzing data and reporting NGS results. We hope the guidance and information provided in this letter will help inform your efforts to define the personnel requirements to meet a laboratory's informatics needs when performing NGS testing.

(1) What are the roles and responsibilities for all personnel performing bioinformatics or pathology/ laboratory informatics activities? What training is considered essential for each of the roles? What competencies are considered essential for each of the roles? What minimum educational requirements (degrees or courses) are required for each of the roles?

While there is some minimal overlap, it is important to first distinguish between the laboratory's clinical informatics, bioinformatics, and other software related activities as well as the skills needed to carry out each role. In a recent *Journal of Molecular Diagnostics* editorial, AMP distinguished between bioinformatics and clinical informatics.[1] As CLIAC considers recommendations related to informatics, it should be cognizant of these distinctions as well:

> *Bioinformatics* is the science of collecting, analyzing, and managing complex biological data using informatics techniques and the specialty at the center of bioinformatics pipeline development; translational bioinformatics is a subset that develops interpretative methods to optimize the transfusion of data into clinical data.
>
> *Clinical informatics* is the application of informatics and information technology to deliver health care services; it applies to the storage, management, communication, and display of complex molecular and genomic data in clinical information systems.

In general, there are distinct roles within a laboratory's workflow related to bioinformatics/informatics which include:

- *Data acquisition*
  - *Skills: Knowledge of basic hardware and software engineering, security and system design*
  - *Activities:*
    - *Establish secure, automated, and fault-tolerant data hand-shake between laboratory instrumentation and place where the data is analyzed, i.e., computer servers.*
- *Data analysis and presentation*
  - *Skills: Significant knowledge of bioinformatics, software engineering and genomics concepts/domain expertise (e.g., cancer, hereditary diseases, virology) and the role of genetics therein; expertise in hardware scaling, hardware and software redundancy, network security.*
  - *Activities:*
    - *Create, select, evaluate, and validate various computer algorithms for quality control and data analysis*
      - *Define the usage of existing computational algorithms for CLIA environment*
      - *Write programming code blocks for data integration suitable for CLIA environment*

[1] https://jmd.amjpathol.org/article/S1525-1578(19)30194-1/fulltext

        - *Develop new computer algorithms for data analysis that are suitable for a CLIA environment (i.e. are versioned and well-documented)*
    - *Present analyzed data to clinical personnel and professionals with limited computer/technical skills to enable clinical decision-making*
    - *Secure data integration into LIS for clinical report sign-outs*
- *Data Storage and retention:*
    - *Skills:*
        - *Significant knowledge of hardware, software and bioinformatics concepts related to genomic data representation and file formats*
        - *Significant understanding of laboratory accreditation requirements and best practices for data storage, retention, and verification of integrity*
        - *Significant knowledge of applicable regulatory requirements specifically pertaining to PHI, HIPAA, HITECH, data privacy, etc.*
    - *Activities:*
        - *Train on all necessary specifications for genomic data representation (e.g. FASTQ, SAM/BAM, CRAM, and VCF)*
        - *Determine what data (file) type to retain and where to retain*
        - *Evaluate type of storage (fast or slow storage) for balance between reliability and cost*
        - *Determine how long to retain identified data in the context of federal, state and local regulatory requirements as well as institutional policies.*
        - *Determine level of redundancy of data backup (primary, secondary, on-site and off-site data storage in case of an emergency)*
        - *Establish a disaster recovery (DR) plan for fail-safe and secure data storage and retrieval*
        - *Determine appropriate data transfer (network) infrastructure needed for storage and retrieval.*

In order to fill all the necessary roles, a laboratory may have individual staff serve in one or more roles, depending on the laboratory and its unique needs. Additionally, a laboratory may hire, contract for, or otherwise outsource individuals with expertise in any of the following areas:

Clinical genomic bioinformaticists: develop/select/validate/document/evaluate software for bioinformatics pipelines that generate data for interpretation of individual patient specimens

Clinical genomic data scientists: extract, transform, analyze and manipulate bioinformatics-generated data across populations

Clinical informaticists: focus on the transmission, organization, display and use of genomic data in Laboratory Information Systems and Electronic Health Records

Software engineers and other personnel: develop and maintain applications, web servers, databases, hardware and storage infrastructure, and security

Clinical genomic bioinformaticists and clinical genomic data scientists must have expertise in genetics/genomics to develop accurate and robust pipelines and algorithms for genomic data. Clinical informaticists focus on usability and software design, including that of usability of genomic data. A pathology clinical informaticist may focus on the design of the application being used to sign out genomic tests as well as the usability of that data once it crosses over to the EHR. Other personnel focus on infrastructure needs including database administration, website design, software design, and interfaces with electronic health records. For one staff member to tackle all of these needs and roles, it would require an expertise in software engineering, biological sciences (genetics), and statistics; however, it is extremely rare to find one person with expertise in all of these areas, and the work effort for NGS testing is too much for any one person to realistically sustain. Ideally, laboratories would employ multiple people to meet all of their clinical informatics and bioinformatics needs. Unfortunately, budget and market constraints often restrict hiring and retention. In the end, many laboratories either hire a bioinformaticist or outsource bioinformatics to a third party company to focus on analyzing the data. On occasion, an internally hired bioinformaticist is tasked with clinical informatics responsibilities as well.

Bioinformaticists and data scientists working on clinical sequencing assays ideally have a Master's degree or higher in a related informatics, computer science, genetics, or pathology specialty. They should be overseen by qualified physicians or doctoral scientists with significant training and experience in clinical genomics bioinformatics, in addition to clinical expertise. Given the lack of clinical education and subspecialty board certification specific to these informatics fields (further discussed below), the laboratory director should have discretion regarding the educational requirements and professional experience required for a given position, and CLIAC recommendations should not be overly prescriptive.

Additionally, it is important to note that while laboratory directors may not have direct expertise in bioinformatics/informatics, they have the responsibility to ensure that bioinformatics/informatics software solutions are fit for purpose as supervisors of all laboratory activities. Responsibilities include ensuring appropriate division of responsibilities in bioinformatics/informatic activities and verifying robust assay performance as intended. Additionally, laboratory directors should consider adhering to industry standards such as the Software Development Life Cycle (SDLC).

(2) What are the challenges for recruitment and retention of bioinformatics or pathology/laboratory informatics personnel?

The initial hurdle to hiring informatics personnel is convincing human resources and hospital administration of the need to allocate funds for a purpose that does not directly lead to increased revenue. Recruitment of talent poses a significant challenge due to competition from potentially lucrative opportunities at software technology companies such as social media companies, online retailers, etc. This is particularly pronounced for departments in smaller cities who, overall, have much greater difficulty accessing individuals with this sort of talent/skillset. Retention of talent is also difficult given the lack of opportunities for promotion within the laboratory for this type of position. Data professionals often transition to the technology industry early in their career. Additionally, those data

professionals who do choose a clinical academic setting often devote some proportion of their time and attention to academic pursuits, as it is one of the few (perhaps the only) paths to career advancement available to a data scientist in a hospital setting. Not all clinical laboratories provide the resources or support such pursuits, further impeding recruitment and retention. Similarly, assignment of responsibilities outside of primary training area (e.g., bioinformaticist is asked to develop and manage web application for supporting laboratory operations) also negatively impact employee retention. The lack of available subspecialty certifications for these personnel is also a barrier to both advancement and salaries. Finally, the highly regulated and controlled clinical environment is at odds with the culture of adaptation and rapid innovation associated with data and programming professionals.

To assist with recruitment and retention of informatics employees, it would be helpful for there to be additional certification options or another accreditation program. This would provide opportunities for professional development and promotion, and would create confidence and consistency in the expertise that these professionals offer to clinical laboratories. Additionally, this would further be supported by the creation of graduate degree programs designed specifically for clinical use of these technologies in genomics. The need for both bioinformatics and data science professionals will continue to grow as uptake of genomic testing increases, and investing in educational programs today will assure a steady stream of talented professionals to meet the expanding need over time. While this is not within the purview of CLIAC, a recommendation to universities to expand training into these fields would be a helpful step for laboratories advocating both within human resources to create employment opportunities and also within graduate programs to design curricula to train future professionals.

(3) What are examples of how NGS data files are used in addition to generating a clinical test result?

Within the clinical realm, these data are also helpful with maintaining laboratory accreditation requirements, routine quality control checks, specimen selection for assay validation, validating pipeline upgrades, interacting with third party vendors via application program interfaces (APIs), and infectious disease surveillance. Regarding the last point, appropriate re-analysis of NGS data sets upon request can be useful for longitudinal tracking of chronic patient cases in infectious diseases (e.g., tracking emerging resistance in viral sub-populations in HIV patients, or serial microbiome analysis to track dysbiosis). Sequence data can be very valuable to fuel molecular epidemiological analysis for infectious disease surveillance and outbreak analysis, for example identification of nosocomial infections and targeted interventions for infection control in hospitals. Depending on consent and required institutional review board approval, NGS data files may also be used for biomedical research conducted by the laboratory, the laboratory's partners, or by clinical laboratory teams and researchers worldwide via public databases.

(4) What NGS data files should be retained for quality assurance, repeat analyses, or subsequent analyses? How long should these NGS data files be retained?

A major limiting factor to storing large data files over time is their size and the corresponding cost to ensure adequate secure storage space. Acknowledging this concern and the need to provide laboratories with sufficient resources to meet these storage requirements, AMP believes that unless federal or state law or regulations stipulate otherwise, the raw variant call format (VCF) files and the assay specific browser extensible data (BED) files should be retained for a *minimum* of ten years. The binary sequence alignment map (BAM) files and FASTQ files should be stored for a minimum of two years. The choice of data types to store should be coupled with the appropriate use of data compression techniques as well as type of storage (fast vs slow storage) to optimize cost.

(5) What are the challenges and approaches for laboratories to maintain and utilize previous versions of sequence analysis software?

There are several challenges to maintaining and utilizing older versions of sequence analysis software that AMP members encounter in their laboratory practice. First, there are challenges associated with a laboratory obtaining sufficient positive controls that span the breadth of the NGS test, as well as challenges in developing pipelines that detect all of the variant types in all regions of the assay. Secondly, operating systems and software are continuously updated, and each update has to be validated by the laboratory prior to being put into use. After validation has been approved and the test put into production, it can be difficult or impossible to access or run a previous version of software on existing or updated hardware. Thirdly, some laboratories outsource bioinformatics to commercial software. The laboratories are unable to control how to maintain and utilize previous versions, it all depends on the design of the commercial software.

Additionally, CLIA requires that the versions of all software used when validating a test must be the software used after the test goes into production. Most if not all NGS tests require multiple software programs in the bioinformatics pipeline, which are run either in series or in parallel. Maintaining numerous versions of software requires time, space, and financial resources from an already resource-limited laboratory. Laboratories must maintain detailed documentation of each version of the software used, including its parameters and any differences between the versions and changes in software behavior. Some software used in clinical practice may be used beyond the support lifetime of the original developer, preventing use with a pipeline that has updated other software components.

Lastly, storage of the dates of continuously updated external content, especially reference data that is integrated into a pipeline from an outside source (e.g., ClinVar), is critical, and being able to reevaluate data previously retrieved from a data source that has since been updated can be difficult to impossible. As taxonomy, identifiers, and evaluations of variant pathogenicity change over time, it is critical that previous versions of reference databases are not only maintained but also accessible to laboratories, particularly for databases funded by the United States government (e.g., dbSNP, ClinVar). Currently, there are no standards for how many prior versions of such external databases should be maintained and no accreditation or regulatory requirements for this. Therefore, a standard for the number and/or length of time that prior versions of databases should be maintained would be extremely helpful.

Thank you very much for the opportunity to submit these comments in response to your recent request for information on the bioinformatics personnel needs for laboratories providing NGS-based testing. If AMP may be of further assistance, please don't hesitate to contact Tara Burke at tburke@amp.org.

Sincerely,
Karen E. Weck, MD, FCAP
President, Association for Molecular Pathology