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

November 1, 2021

Francis S. Collins, MD, PhD   Janet Woodcock, MD
Director   Commissioner
National Institutes of Health   Food and Drug Administration
9000 Rockville Pike   10903 New Hampshire Avenue
Bethesda, Maryland 20892   Silver Spring, Maryland 20993

Attn: Notice Number: NOT-OD-21-162

Delivered electronically to: https://osp.od.nih.gov/rfi-comment-resource-gaps-for-radiomics/

Dear Drs. Francis S. Collins and Janet Woodcock,

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

On behalf of the Association for Molecular Pathology (AMP), thank you for the opportunity to submit comments in response to the National Institute of Health (NIH) and the Food and Drug Administration (FDA)'s Request for Information (RFI) on "Critical resource gaps and opportunities to support Next Generation Sequencing (NGS) test development, validation, and data interpretation, including the use of technologies such as artificial intelligence (AI)/machine learning (ML)." 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 members are at the intersection of molecular diagnostics and frequently conduct, design, and interpret NGS-based tests, so we are pleased to engage in this important conversation to address critical resource gaps and explore opportunities that support the use of NGS to help meaningfully move the field of molecular diagnostics forward. We first want to acknowledge that many NGS-based tests are currently an essential component of medical practice, and there are already high-quality standards for performing and interpreting NGS-based tests. We hope any outcomes from this RFI process build from existing efforts and resources to support advances in NGS technology. Moreover, use of NGS technology spans many areas of medicine, including oncology, inherited diseases and disorders, and infectious diseases and while gaps exist across the board, it is important to note that resource needs will likely differ across disciplines.

Additionally, most NGS-based tests are laboratory-developed testing procedures (LDPs) and therefore, meet the regulatory requirements promulgated by the enactment of the Clinical Laboratory Improvement Amendments (CLIA). CLIA regulations set strict standards to ensure the high quality of

LDPs conducted in clinical laboratories across the United States while enabling patients to receive the necessary and timely information impacting their health care. These tests are currently recognized as the standard of care in diagnostics for all types of conditions including cancer, inherited disorders and diseases, and infectious diseases. These tests meet or exceed CLIA standards, and/or other federal, state, and professional practice standards, as well as provide clinically significant information to patients. Many have been demonstrated to be of highest quality by peer review through the College of American Pathologists (CAP) laboratory inspection processes.

Developing high-quality reference samples, tools, and infrastructure that can enhance clinical laboratories' ability to conduct test validation, ensure quality control, and engage in proficiency testing of NGS-based tests is critical to AMP members. We outline our comments on critical resource gaps and opportunities to support NGS tests below.

Topic 1: Development of reference samples, tools, and infrastructure for clinical and translational research using NGS

NGS technology is an important tool used in clinical and research settings, which inform patient care. Advancements in NGS technology continue and there are unique challenges associated with analyzing and applying NGS data that require more sophisticated analyses and standards due to the sheer size of the human genome and diversity of microbial genomes. We strongly believe that establishing better quality reference samples, tools, and infrastructure can address these challenges. We applaud NIH and FDA for identifying this gap and for working with stakeholders to make meaningful contributions in this area. To assist with your efforts, AMP provides the following recommendations:

Reference Materials: AMP has long maintained that reference materials are a key component to the design and validation of NGS tests and has advocated for more funding and resources be devoted to their development. AMP's advocacy on this area began over twenty years ago with a request for funding for the National Institute of Standards and Technology (NIST) to develop reference materials and through AMP representation on the NIST Genome in a Bottle (GIAB) Steering Committee. AMP also hosts a Reference Materials Forum annually[1] and participates in the Joint Initiative for Metrology in Biology Coronavirus Standards Working Group. Additionally, a number of AMP members independently participate as part of the NIST GIAB and/or Genetic Testing Reference Materials Coordination Program (GeT-RM) projects. Finally, AMP highlights the need for reference materials into all our published guidelines, when appropriate, and states support for reference materials in a number of published papers[2,3,4,5,6]. AMP is a leader in the field of, and advocate for, development of NGS reference materials and looks forward to engaging with NIH and FDA on this topic.

---

[1] https://amp19.amp.org/program/special-events/amp-reference-materials-forum/
[2] https://www.jmdjournal.org/article/S1525-1578(13)00131-1/fulltext#secsectitle0045
[3] https://www.jmdjournal.org/article/S1525-1578(12)00175-4/fulltext
[4] https://www.jmdjournal.org/article/S1525-1578(15)00172-5/fulltext#secsectitle0060
[5] https://www.jmdjournal.org/article/S1525-1578(17)30025-9/fulltext#secsectitle0055
[6] https://www.jmdjournal.org/article/S1525-1578(17)30373-2/fulltext

A significant gap exists for NGS reference materials for infectious diseases. As the use of NGS technologies for infectious disease diagnosis continues to increase, laboratories are recognizing the paucity of reference materials as a barrier to providing these clinically useful tests[7][8][9][10]. Additionally, there is also a gap in the number of well-curated and complete genomes that are used in reference databases to identify pathogens. This situation is especially dire for fungal pathogens. Thus, AMP supports more attention and resources be devoted to bolstering reference materials and databases.

Currently, many molecular professionals create their own controls to meet a range of needs, to account for specimen type and preservation method. We also recognize that the cost of high-quality reference samples, and the appropriate length of time for their use, impacts laboratory test validation practices. As such, AMP strongly supports the further development of high-quality and affordable physical samples and *in silico* approaches to be used during the workflow process. This would allow for the use of a variety of reference samples at multiple points of the NGS workflow to identify systematic sequencing and analysis errors, such as sequencing artifacts and short-read misalignment, which often cannot be overcome by high sequencing coverage.[11] AMP also recommends investments in developing reference samples that have been validated and well-characterized for the detection of challenging variant types, such as large insertions and deletions[12]. Additionally, limits of detection and accuracy can vary widely for different NGS technologies, types of variants, microorganisms (e.g., virus vs. bacteria), and nucleic acids (e.g., RNA vs. DNA) ,[13] thus, in order to effectively assist clinical laboratory practice, NIH and FDA should develop reference samples that can adequately evaluate limits of detection. Furthermore, reference samples should be comparable to patient samples to reduce bias in calibrated measurements and prevent inaccurate diagnoses.[14]

Data Analytical and Interpretation Software: There is a shortage of professionals with bioinformatic expertise to meet the current demand, which means that clinical laboratories often experience challenges in accessing in-house expertise for data interpretation.[15] Some NGS companies have sought to fill this need by creating proprietary data analytic software. However, these types of software are expensive and many laboratories may not be able to afford them. Therefore, AMP believes that there is an opportunity for NIH and FDA to create open-sourced and broadly accessible clinical bioinformatic tools to improve clinical laboratories' bioinformatics capabilities. Additionally, while the needs for these open-sourced/accessible bioinformatic tools exist across all subspecialties of molecular pathology, the specifics might differ based on the species assessed (e.g., human vs. microbe). There is also a need for expanded training programs for bioinformatics professionals, which could be established by NIH, either as intramural programs or through partnerships with academic centers.

Additionally, molecular professionals lack access to clinically-validated virtual NGS data interpretation tools for assessing variant data. While many exist for research purposes, very few have been validated for clinical use. Thus, AMP urges NIH and FDA to invest in efforts to create virtual data interpretation tools that have been clinically validated for use in patient care.

[7] https://meridian.allenpress.com/aplm/article/141/6/776/129024/Validation-of-Metagenomic-Next-Generation
[8] https://academic.oup.com/cid/article/66/5/778/4460112
[9] https://academic.oup.com/jid/article/221/Supplement_3/S331/5571876
[10] https://www.jmdjournal.org/article/S1525-1578(15)00172-5/fulltext
[11] https://www.nature.com/articles/nrg.2017.44
[12] https://www.jmdjournal.org/article/S1525-1578(17)30373-2/fulltext
[13] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6941185/
[14] https://www.nature.com/articles/nrg.2017.44
[15] https://www.nature.com/articles/s41591-020-0935-z

Finally, when considering the data analytical and interpretation software for NGS data in the clinical setting, it is critical to focus on how these data will be incorporated into the Electronic Health Record (EHR) to ensure that data are accessible and interpretable by the health care teams. AMP has identified multiple challenges with how NGS genomic data, particularly discrete variant data, are displayed within EHRs and made interoperable[16]. AMP strongly encourages NIH and FDA to address solutions for genomic data availability and interoperability in the EHR as it relates to NGS within the scope of this project.

Data Storage: The sheer amount and sensitivity of data produced by NGS-based tests calls for advances in data storage infrastructure, security, and standardization. Across organizations, there are different NGS data storage practices. This includes increased use of cloud-based storage systems, which have significant advantages for storing NGS data due to their low per-GB prices and minimal fixed costs. These systems may be a good choice for many laboratories to store NGS data;[17] however, ensuring compliance with Health Insurance Portability and Accountability Act (HIPAA) and related regulations is complex.[18] AMP recommends that NIH and FDA provide guidance and education on the use of data storage systems and how best to comply with any relevant federal requirements.

Topic 2: Application of AI/ML to the interpretation of NGS data and multi-domain data

AMP acknowledges the potential value of AI/ML for the interpretation of NGS data. NGS technology has advanced significantly since its first inception and now is a highly valuable tool for patient care. However, as NGS technology evolves and science advances, there is a vital need for sophisticated data analysis methods that can analyze and interpret the large volumes of data resulting from NGS-based tests. AMP is pleased to see that NIH and FDA are interested in information on the role of AI/ML in analyzing NGS data and ways to potentially advance the field, as we believe there are several opportunities that NIH and FDA can address.

AI/ML capabilities can only be as good as the data used to support their development. To enable the use of AI/ML for the interpretation of NGS data, variant data sharing efforts need to be improved and expanded. Despite the widely recognized understanding that data sharing is critically important for advancing the use of genetic and genomic information in patient care, there are considerable barriers that prevent many institutions from contributing to public databases. In January 2020, ClinGen reported that only fifteen clinical laboratories met their minimum data sharing requirements.[19] In recognition of this low level of participation in data sharing, AMP recently published a series of recommendations[20] for policymakers, clinical laboratories, institutions, and other stakeholders to resolve numerous data sharing barriers identified by our members for the purposes of advancing the conversation about how to make data more broadly accessible.

Specific to your efforts, AMP would also like to emphasize the need for simplifying and streamlining patient consent standards for laboratories in instances when consent is required, as well as data preparation for the purposes of safeguarding and sharing of NGS data. Data sharing is a labor-intensive and costly activity that is not reimbursed. The processes needed to prepare information for sharing – for

[16] https://www.jmdjournal.org/article/S1525-1578(21)00327-5/pdf
[17] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7276491/
[18] https://www.nature.com/articles/gim201392
[19] https://www.clinicalgenome.org/tools/clinical-lab-data-sharing-list/
[20] https://www.amp.org/AMP/assets/File/advocacy/AMP_Position_Variant_Data_Sharing_7_29_2021.pdf

example de-identifying a patient's health information before submission to a variant database; ensuring that the data meets at least the minimum requirements expected from ClinVar or other data sharing entities; and data reformatting required for submission to the online database – are highly variable and require considerable resources. First, AMP calls for clear guidance on best practices for establishing consent for the sharing of de-identified protected health information to minimize hesitancy. Secondly, AMP urges NIH, FDA, and other relevant federal agencies to work with clinical, academic, and public interest stakeholders to convene a standards-setting body to accelerate harmonization of variant database elements. We believe these activities will aid in resolving some of the disincentives to data sharing.

Further, genetic databases need to be well-curated. Current curation practices for the associated molecular pathology subspecialties vary among the professional communities. For example, some variant databases currently lack uniform standards and expert curation, which can decrease confidence or utility in these databases. Lack of confidence in the accuracy, quality, or completeness of a database, especially as new information is compiled and reclassified, can lead to hesitancy among clinicians to both contribute to and utilize databases. AMP calls for standardized guidelines and appropriate financial support to ensure that data are continuously evaluated and updated. Further, operators should work to apply curation processes to remove obvious data outliers, request evidence when it was not included in the original submission, and follow up with individuals who have submitted older data to verify relevance (i.e., that variant data has been appropriately updated).

Including diverse data from a wide range of participants in genetic databases will also be crucial in developing AI/ML algorithms that are appropriate for patients from all ethnicities and races. We are concerned that the lack of diversity in many genetic databases is leading to algorithms that are only clinically applicable to specific groups whose data are represented, furthering health disparities for those underrepresented groups. There are already known issues with ancestry bias in clinical databases that have resulted in less informative and more costly genomic testing for patients of non-European ancestry.[21] Substantial efforts must be made to increase the access of traditionally underserved groups to molecular pathology services. For instance, AMP has called upon stakeholders to establish data sharing policies that emphasize the importance of contributing data which represents the diversity of their patient populations. We believe that federal agencies, such as NIH and FDA, can play a large role in advancing these efforts and acknowledge that both NIH and FDA are already aware of these issues. As example, NIH has made considerable efforts to ensure that they are taking an inclusion approach for the All of Us program. As you work towards resolving gaps to support the application of AI/ML to the interpretation of NGS data, AMP recommends that data diversity become an area of focus.

AMP also recognizes that given the nature of NGS data and the necessary protections to safeguard privacy, there will never be "complete" sharing of NGS data. This could cause issues with how widespread an AI/ML algorithm can be applied and, in some situations, an algorithm may only be appropriate for the dataset for which it was developed. To work towards the use of AI/ML that can have more widespread applications, the development of AI/ML algorithms should not be considered a one-and-done activity; instead, datasets and AI/ML algorithms should be continuously updated, evaluated, and curated to ensure ongoing relevance and quality. Professional guidelines and resource support should be provided to clinical laboratories to help with these efforts as smaller clinical laboratories alone may not have the capabilities to update and curate data for AI/ML algorithms.

---

[21] https://pubmed.ncbi.nlm.nih.gov/30209973/

Topic 3: Existing resources that could be leveraged to fill resource gaps

There are a broad range of public and private resources that could potentially be leveraged, such as but not limited to, ClinVar[22], ClinGen[23], Genome in a Bottle (NIST)[24], The Cancer Genome Atlas Program (TCGA)[25], Catalogue of Somatic Mutations in Cancer (COSMIC)[26], OncoKB[27], and cBioPortal[28]. These resources are focused either broadly on genomics or specific to cancer genetics, and the resources have been used to aid in the validation and characterization of genetic variants. To better leverage existing resources to fill resource gaps, AMP believes that the NIH and FDA should solicit feedback from publicly funded NGS-related resources to learn about any resource gaps they have identified and also the possibility that these programs could potentially fill existing resource gaps.

Thank you again for the opportunity to provide these comments in response to the RFI. We are hopeful that the NIH and FDA take action to develop NGS reference samples, tools, and the needed infrastructure to ensure that clinicians have vital resources to aid in their abilities to use NGS technology for clinical applications. We hope that you will consider AMP and our members as a resource as you take the next steps in this effort. If you have any questions, please do not hesitate to contact Tara Burke, PhD at tburke@amp.org.

Sincerely,

Antonia R. Sepulveda, MD, PHD
President, Association for Molecular Pathology

---

[22] https://www.ncbi.nlm.nih.gov/clinvar/
[23] https://clinicalgenome.org/
[24] https://www.nist.gov/programs-projects/genome-bottle
[25] https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
[26] https://cancer.sanger.ac.uk/cosmic
[27] https://www.oncokb.org/
[28] https://www.cbioportal.org/