Letters to the Editor

Standardizing Next-Generation Sequencing Experiments and Analysis Methods

To the Editor:

The past few years have seen the emergence of new strategies for high-throughput DNA sequencing that have invigorated life science research. Because of these technological advances, the available sequencing data continue to expand. Additionally, cost reductions have made access to sequences of entire genomes easier for established laboratories and even easier for small research groups. A lack of consensus exists, however, on how best to design and analyze next-generation sequencing (NGS) studies and how to compare these data to previous and future work. Comparing different sequencing platforms and analysis methods, and thus their interpretations, is becoming more complicated. The main purpose of such studies is somehow getting lost in the process of experimental design as biologists confront new territory with new strategies and tools to be compared more easily and allow investigators to better understand the interpretations and nuances that derive from each method.

Three reports on autism spectrum disorder recently published in Nature have described exome-sequencing approaches (2–4). The 3 reports concluded that genomic associations were weak and that individual candidate varying loci were incongruent. The 3 studies were different with respect to their approach to exome-enrichment methods, the sequencing platform used, the algorithm and parameters used in calling single-nucleotide polymorphisms, and pathway analysis. These studies confound the question of whether the differences are due to different sample populations or different methods of analysis. At present there is no right or wrong approach, but imagine having a positive standard. By using such a standard sample, we would be better able to compare these studies and determine whether the lack of agreement was simply due to differences in sample populations or to analysis techniques. We would then have the option to compare the 3 studies (and any future studies that use the standard) and then determine their consequences and interpretations to extract the most valuable information. If such standards existed, we could better compare 2 approaches or data from different groups and make appropriate decisions about sequences, genetic variation, or methodologic errors. With the introduction of less expensive DNA-sequencing methods, cost is no longer a justifiable reason as long as the costs support the integrity of the study. The availability of a large quantity of the biological sample would serve as the standard is critical, and it will probably require commercial production, which might produce batch-to-batch variation in either cloned DNA or a fully characterized and stable cell line. NIST expects to publicly issue the samples as fully characterized standards by December 2014 (5); therefore, it will be possible to mandate such characterized standards for all NGS studies. Manufacturers that supply a standard sequence for each batch of cloned standard DNA will make it an attractive choice over the standard cell line, which may be susceptible to genetic drift and not have a batch-to-batch standard sequence supplied by the US Food and Drug Administration or NIST.

Furthermore, most sequencers reserve a lane for a technical control sample. For example, PhiX is currently used in the control lane of the Illumina Genome Analyzer II sequencer. If PhiX were replaced with a biological standard sample (cloned characterized standard DNA or a characterized standard cell line) and those sequences were mapped and assembled with the method used for the rest of the samples, then new approaches could be better vetted, and a lack of agreement could be better understood. Simply replacing a sample of a technical standard sample with a more meaningful biological standard sample would allow us to better resolve many questions without appreciably increasing the cost.

Thus, we propose checkpoints and guidelines for standardizing NGS studies (Fig. 1) that we believe will help emphasize the importance of the positive standard in
NGS studies and will help in establishing a consensus.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References


Enusha Karunasena2
Harold R. Garner2*
Virginia Bioinformatics Institute
Virginia Polytechnic Institute and State University
Blacksburg, VA
* Address correspondence to this author at:
Virginia Bioinformatics Institute
Virginia Polytechnic Institute and State University
Washington St., MC0477
Blacksburg, VA 24061-0477
Fax 540-231-2606
E-mail garner@vbi.vt.edu

Previously published online at DOI: 10.1373/clinchem.2012.189241

Jasmin H. Bavarva2
Wyatt McMahon2
Megha J. Bavarva2

Fig. 1. NGS checkpoints for experimental design and analytical methods.