The international MAQC Society launches to enhance reproducibility of high-throughput technologies

To the Editor: Reproducibility is a fundamental hallmark of good science. The US Food and Drug Administration (FDA)-led Microarray and Sequencing Quality Control (MAQC/SEQC) consortia conducted three projects[1–3] to assess the reliability and reproducibility of genomics technologies, including microarrays, genome-wide association studies, and next-generation sequencing. Here, we announce that this decade-long effort has led to the formation of a new international society, the Massive Analysis and Quality Control (MAQC) Society (http://www.maqcsociety.org), which is dedicated to quality control and analysis of massive data generated from high-throughput technologies for enhanced reproducibility.

The goals of the MAQC Society are twofold: first, to advocate and facilitate enhanced reproducibility across multiple experiments, laboratories, and data analysis methods via the development and application of quality control practices and standard analysis protocols for biomedical data; and second, to advance understanding and best practices in the analysis of massive data from emerging big data technologies applied in drug development, clinical applications, and safety/risk assessment. The MAQC Society provides a platform for discussing issues related to these goals, organizing collaborative activities around them, and informing the general public on the results and implications of these activities. It values scientific dialog and cooperation with other national and international communities (e.g., societies and organizations) of similar focus to promote scientific research, education, and communication of reproducible science. In addition, we will devote substantial effort to support, encourage, and mentor the career development of young professionals deeply engaged in vocations aligned with the MAQC Society's goals.

The Society was initially announced in the MAQC/SEQC meeting, but the larger goal is to obtain the participation of scientists across all biomedical fields who have an interest in enhancing reproducibility in the context of high-throughput approaches. The first Society meeting, which was held at SAS Institute (Cary, NC, USA) on April 12, 2017, and had a focus on 'Reproducible Genomics', gathered over 100 scientists (mostly MAQC/SEQC consortium members). The Society's membership currently comprises 35% academia, 31% government and regulatory agencies, 23% technology companies, 6% pharmaceutical companies, 4% clinicians, and 1% 'other' categories. Inspired by the productive and open discussion sessions at the first meeting, the Society seeks to foster a culture by which the members can cooperate to advance reproducibility by implementing standard protocols, pipelines, and best practices. We are promoting needed behaviors, such as the proper giving and receipt of feedback, to promote better practices in the life sciences.

To extend its mission, the MAQC Society plans to establish a series of case studies by which its members will share examples of published manuscripts with data and source code freely accessible to the community. Case studies will span various data types and tools to best represent the diversity of reproducibility frameworks; the larger goal is to facilitate the development and adoption of best practices within the larger research community. There is little doubt that genomics has changed our way of studying disease and health; however, without the implementation of robust data pipelines and analytical frameworks, we are concerned that the ultimate clinical utility of genomic technology may be compromised by a lack of rigorous data analysis.

Across the landscape of clinical medicine, drug development, and genomic technology development, reproducibility is the foundation for its translation to clinical utility and regulatory application. In the era of precision and predictive medicine, the research community needs more rigorous science for discovering effective therapies, identifying responder or adverse event-sensitive patient populations; indeed, clinical progress essentially depends on reliable and reproducible results. Error, whether it is human error, computational error, or technical error, and imprecise protocols could lead to irreproducible or inconsistent results that may contribute to patient risk or death. These concerns pervade high-throughput omics technologies[4–16], such as microarrays[1,2], next-generation sequencing[3], metabolomics, and proteomics for both preclinical and clinical studies.

Among various issues encountered, computational reproducibility becomes increasingly challenging in this field. This is simply due to the fact that the size of data is so massive that the manual inspection of data quality and analysis results is often impossible; thus, reproducibility remains largely at the mercy of whichever algorithm is used, which often lacks necessary benchmarking and metrics to assess reproducibility. Furthermore, a plethora of statistical methods have been published in the omics era and are typically promoted in terms of balancing sensitivity and specificity. However, reproducibility is seldom emphasized. The urgent unmet clinical need for better medicines, improved clinical tests, and accurate precision medicine, compounded by the alarming number of irreproducible studies, precipitates the need for a framework of reproducible science. We must explicitly consider reproducibility, a fundamental hallmark of good science, as a third dimension in addition to sensitivity and specificity.

Community-wide standardization and quality control efforts have recently been initiated in response to concerns on the lack of reproducibility in the generation, analysis, and interpretation of 'big data'[5,8,12,13,17] (https://elifesciences.org/collections/reproducibility-project-cancer-biology;

http://www.nature.com/news/reproducibility-1.17552; https://f1000research.com/gateways/PRR). The newly launched International MAQC Society will strive to work with various scientific communities to develop consensus on best practices for enhanced reproducibility in generation, analysis, and interpretation of massive data from increasingly innovative biomedical fields. More information about the MAQC Society can be found at http://www.maqcsociety.org.

ACKNOWLEDGMENTS

The first MAQC Society meeting was supported by the SAS Institute, Burroughs Wellcome Fund, and the National 863 Program of China (2015AA020104, to L.S.).

DISCLAIMER

The views presented in this article do not necessarily reflect current or future opinion or policy of the US Food and Drug Administration. Any mention of commercial products is for clarification and not intended as endorsement.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Leming Shi[1,2], Rebecca Kusko[3], Russell D Wolfinger[4], Benjamin Haibe-Kains[5–7], Matthias Fischer[8,9], Susanna-Assunta Sansone[10], Christopher E Mason[11], Cesare Furlanello[12], Wendell D Jones[13], Baitang Ning[14] & Weida Tong[14]

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai, China. [2]Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China. [3]Immuneering Corporation, Cambridge, Massachusetts, USA. [4]SAS Institute Inc., Cary, North Carolina, USA. [5]Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [6]Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, Ontario, Canada. [7]Ontario Cancer Research Institute, Toronto, Ontario, Canada. [8]Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne, Medical Faculty, Cologne, Germany. [9]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. [10]Oxford e-Research Centre, Engineering Science Department, University of Oxford, Oxford, UK. [11]Weill Cornell Medicine, New York, New York, USA. [12]Fondazione Bruno Kessler, Trento, Italy. [13]Q[2] Solutions, Morrisville, North Carolina, USA. [14]National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas, USA.
e-mail: weida.tong@fda.hhs.gov.

1. Shi, L. *et al. Nat. Biotechnol.* **24**, 1151–1161 (2006).
2. Shi, L. *et al. Nat. Biotechnol.* **28**, 827–838 (2010).
3. Su, Z. *et al. Nat. Biotechnol.* **32**, 903–914 (2014).
4. Begley, C.G. & Ellis, L.M. *Nature* **483**, 531–533 (2012).
5. Anonymous. *Nat. Biotechnol.* **30**, 806 (2012).
6. Collins, F.S. & Tabak, L.A. *Nature* **505**, 612–613 (2014).
7. McNutt, M. *Science* **343**, 229 (2014).
8. Anonymous. *Nat. Biotechnol.* **33**, 319 (2015).
9. Alberts, B. *et al. Science* **348**, 1420–1422 (2015).
10. Nosek, B.A. *et al. Science* **348**, 1422–1425 (2015).
11. Baker, M. *Nature* **533**, 452–454 (2016).
12. Beaulieu-Jones, B.K. & Greene, C.S. *Nat. Biotechnol.* **35**, 342–346 (2017).
13. Di Tommaso, P. *et al. Nat. Biotechnol.* **35**, 316–319 (2017).
14. Ioannidis, J.P.A. *et al. PLoS Med.* **2**, 0696–0701 (2005).
15. Micheel, C.M. *et al. Evolution of Translational Omics; Lessons Learned and the Path Forward.* (National Academies Press, Washington, DC, 2012).
16. Ioannidis, J.P.A. *et al. Nat. Genet.* **41**, 149–155 (2009).
17. Wilkinson, M.D. *et al. Sci. Data.* **3**, 160018 (2016).

Challenges and recommendations for epigenomics in precision health

To the Editor: In March 2017, US life insurance company, GWG Life (Minneapolis) started to require policy owners to submit saliva samples. The company was not interested in the genes that their customers inherited, but in the epigenetic state of genes in the form of DNA methylation, for which it had licensed an epigenomic technology to predict an individual's health and life span[1,2]. This raises the issue of individual epigenomic profiles being used to charge more or less for insurance coverage—or to deny life insurance altogether. The silver lining is that the epigenetic profile is not fixed in stone; you may improve your epigenome by changes in diet, exercise, or other modifications.

Seven years after epigenetics was featured on the cover of *Time* magazine, the development of novel epigenomic techniques has led to a better understanding of how the epigenome changes across individuals and health states[3]. For example, researchers have shown that the T cells of the immune system change at the epigenomic level during aging or functional exhaustion[4,5]. The response of certain cancer patients to a drug treatment can be predicted from DNA accessibility of target loci during treatment[6]. Whether a person smokes cigarettes, and thus may be at risk for myriad cancers, can be inferred from DNA methylation patterns[7]. The origin of cell-free DNA, released into the blood by damaged tissues can be gleaned from nucleosome positioning in those fragments[8]. Thus, in the coming years, we can expect epigenomic findings to be used increasingly in determining diagnosis, treatment course, and even the cost of insurance. Here, we describe five recommendations for continued development in this field, formulated with an interdisciplinary group of experts toward realizing the full potential of epigenomic medicine.

The authors' perspectives for this piece came together via the Centers of Excellence in Genomic Science (CEGS), sponsored by the US National Human Genome Research Institute (NHGRI). CEGS aim to develop and disseminate novel genomic and epigenomic technologies. As technology developers, we recognize that the field of epigenomics is rapidly maturing in both the technological and biological sciences and many further applications of epigenomic technology have been proposed for both clinical and commercial purposes. For the technology to be optimally designed for discovery or research purposes as well as robust application, priorities for use across these fields should be considered together at the outset. Thus we sought to bring together an interdisciplinary group consisting of both developers and end users of epigenomic technology to propose priorities for the field. The Center for Personal Dynamic Regulomes at Stanford, California, joined with the CEGS investigators from Harvard Medical School (Boston), Massachusetts Institute of Technology (Cambridge, MA, USA), Dana-Farber Cancer Institute (Boston), Massachusetts General Hospital (Boston), the Salk Institute (La Jolla, CA, USA), and the University of Chicago, as well as thought leaders in academic medicine, executives from companies specializing in diagnosis, lifestyle, and data analysis, and biosecurity and bioethics experts. Our recommendations reflect discussions that began with a consideration of clinical and consumer needs and moved to technical feasibility, commercial opportunities, and regulatory and ethical considerations.

The promise and challenge of epigenomics

Precision medicine promises to greatly improve individualized medical care, and this promise hinges not only on genetic tests and therapies, but also epigenetic insights. Although the massive power of DNA sequencing has largely been applied to genome and exome sequencing as a means to trace sequence variants in myriad diseases, such applications do not capture