

## **The International MAQC Society is launched to enhance reproducibility of high-throughput technologies**

Leming Shi<sup>1,2</sup>, Rebecca Kusko<sup>3</sup>, Russell D. Wolfinger<sup>4</sup>, Benjamin Haibe-Kains<sup>5,6,7</sup>, Matthias Fischer<sup>8,9</sup>, Susanna-Assunta Sansone<sup>10</sup>, Christopher E. Mason<sup>11</sup>, Cesare Furlanello<sup>12</sup>, Wendell D. Jones<sup>13</sup>, Baitang Ning<sup>14</sup>, Weida Tong<sup>14,\*</sup>

<sup>1</sup>School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai, China

<sup>2</sup>Collaborative Innovation Center for Genetics and Development and State Key Laboratory of Genetic Engineering, Fudan University, Shanghai, China

<sup>3</sup>Immuneering Corporation, Cambridge, Massachusetts, USA

<sup>4</sup>SAS Institute Inc., Cary, North Carolina, USA

<sup>5</sup>Princess Margaret Cancer Centre, Toronto, Ontario, Canada

<sup>6</sup>Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, Ontario, Canada

<sup>7</sup>Ontario Cancer Research Institute, Toronto, Ontario, Canada

<sup>8</sup>Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne, Medical Faculty, Cologne, Germany

<sup>9</sup>Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany

<sup>10</sup>Oxford e-Research Centre, University of Oxford, Oxford, UK

<sup>11</sup>Weill Cornell Medicine, New York, New York, USA

<sup>12</sup>Fondazione Bruno Kessler, Trento, Italy

<sup>13</sup>Q2 Solutions, Morrisville, North Carolina, USA

<sup>14</sup>National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas, USA

\* Correspondence should be addressed to [weida.tong@fda.hhs.gov](mailto:weida.tong@fda.hhs.gov).

**Abstract:**

Reproducibility is a fundamental hallmark of good science. The FDA-led MicroArray and Sequencing Quality Control (MAQC/SEQC) consortia conducted three projects to assess the reliability and reproducibility of genomics technologies including microarrays, genome-wide association studies and next-generation sequencing. Its decade effort has led to the formation of a new international society, Massive Analysis and Quality Control (MAQC) Society ([www.maqcsociety.org](http://www.maqcsociety.org)), which is dedicated to quality control and analysis of massive data generated from high-throughput technologies for enhanced reproducibility. The Society strives to work with various scientific communities to develop consensus on best practices for generating reliable and reproducible data and its analysis suitable for translation from increasingly innovative biomedical fields into biological application, clinical practice, and regulatory decision-making.

A decade of collaborative effort by the FDA-led MicroArray and Sequencing Quality Control (MAQC/SEQC) consortia<sup>1-3</sup> on improving reproducibility of genomic analysis has yielded a new professional organization. The international Massive Analysis and Quality Control (MAQC) Society ([www.maqcsociety.org](http://www.maqcsociety.org)) was formally launched on April 12, 2017 at SAS Institute, Cary, North Carolina, USA.

Commented [BH1]: Sentence is incomplete. Do you want to keep this?

The MAQC Society is dedicated to quality control and analysis of massive data generated from high-throughput technologies, with a special emphasis on generating reliable and reproducible data and its analysis suitable for translation into biological application, clinical practice, and regulatory decision-making. The goals of the Society are twofold: (1) 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 (2) 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 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. The Society 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, the Society will devote substantial effort to support, encourage, and mentor the career development of young professionals deeply engaged in vocations associated with the Society to develop values compatible with the Society's goals.

The establishment of this new society was initially announced in the MAQC/SEQC consortium and its invitation for participation will be extended to all the scientists from the biomedical fields with a special passion for reproducible science. The first Society meeting with a theme of "Reproducible Genomics" gathered over 100 scientists (most are the MAQC/SEQC consortium members), which was composed of 35% academia, 31% government/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 is fostering 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 receiving of feedback to enable better science through constant forward evolution. To extend its mission, the 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, which will facilitate the community as a whole to develop best practices and generate reproducible results. There is little doubt that genomics has changed our way of studying disease and health and, in many ways, such an expectation is not misplaced. However, the ultimate clinical utility of genomic technology should not be restricted by potential blunders in 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; clinical progress essentially depends on reliable and reproducible results. Errors, 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<sup>4-15</sup> such as microarrays<sup>1-2</sup>, next-generation sequencing,<sup>3</sup> 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 are often impossible and thus the reproducibility is largely in the mercy of algorithms used where the metrics to assess reproducibility have not been established. 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"<sup>5,8,12-15</sup>. We applaud such efforts of the large scientific community including *Nature Biotechnology* in addressing such challenges in a proactive manner. Indeed, *Nature Biotechnology* and Nature Publishing Group have dedicated three special issues and three Web Focuses ([nature.com/nbt/focus/maqc/](http://nature.com/nbt/focus/maqc/), [nature.com/focus/maqc2/](http://nature.com/focus/maqc2/), and [nature.com/nbt/collections/seqc/](http://nature.com/nbt/collections/seqc/)) to the MAQC project, arguably one of the most ambitious public-private-academic collaboration efforts aiming to evaluate and improve the reproducibility and reliability of microarray and next-generation sequencing technologies and the related bioinformatics approaches.

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 [maqcsociety.org](http://maqcsociety.org).

#### **Acknowledgement:**

The first MAQC Society meeting was supported by SAS Institute and Burroughs Wellcome Fund.

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

#### **References:**

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. Editorial. *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. Editorial. *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.                      Reproducibility                      Project:                      Cancer                      Biology  
(<https://elifesciences.org/collections/reproducibility-project-cancer-biology>).
15. Challenges in irreproducible research (<http://www.nature.com/news/reproducibility-1.17552>)