

# Standardization in next-generation sequencing - Issues and approaches of establishing standards in a highly dynamic environment

Christoph Endrullat

Technical University of Applied Sciences, Molecular Biotechnology and Functional Genomics, Wildau, Germany (Current affiliation: University of Glasgow, College of Medical, Veterinary and Life Sciences, Glasgow, United Kingdom)

## Introduction

2<sup>nd</sup> generation sequencing or better known as next-generation sequencing (NGS) represents a cutting-edge technology in life sciences and current foundation for unravelling nucleotide sequences. Since advent of first platforms in 2005 the number of different types of NGS platforms increased in the last 10 years in the same manner as the variety of possible applications. Higher throughput, lower cost and better quality of data were the incentive for a range of enterprises developing new NGS devices, whereas economic issues and competitive pressure, based on expensive workflows of obsolete systems and decreasing cost of market leader platforms, resulted simultaneously in accelerated vanishing of several companies. Due to the fast development, NGS is currently characterized by a lack of standard operating procedures, quality management/quality assurance specifications, proficiency testing systems and even less approved standards along with high cost and uncertainty of data quality. These aspects represent major obstacles for essential implementation of NGS in important areas such as clinical diagnostics, where reliable results, traceable and reproducible data as well as fast processing are crucial. On the one hand, appropriate standardization approaches were already performed by different initiatives and projects in the format of accreditation checklists, technical notes and guidelines for validation of NGS workflows. On the other hand, these approaches are exclusively located in the US due to origins of NGS overseas, therefore there exists an obvious lack of European-based standardization initiatives. Furthermore, market leaders like Illumina, Inc. aim to establish their own standards and are likely not willing to be forced by a general standard in NGS. An additional problem represents the validity of promising standards across different NGS applications. Due to highest demands and regulations in specific areas like clinical diagnostics, the same standards, which will be established there, will not be applicable or reasonable in other applications. These points emphasize the importance of standardization in NGS mainly addressing the laboratory workflows, which are the prerequisite and foundation for sufficient quality of downstream results, i.e. sequence data. Therefore, this presentation exhibits specifically standardization opportunities in the upstream pipeline and provides only a brief remark in the context of standardization in bioinformatics. To summarize, the presentation will give an overview about current standardization efforts, enlighten the problems associated with developing standards and show examples about possible standards and *de facto* standards primarily addressing the NGS upstream pipeline.

## Methods

This work was based on a platform-dependent and -independent systematic literature review as well as personal communications with i.a. Illumina, Inc., ISO/TC 276 as well as DIN NA 057-06-02 AA Biotechnology regarding current standardization efforts in NGS.

## Results

Prior formulation of specific standard proposals, the problems of standardization in NGS itself were identified and interpreted. These problems comprised, amongst others, the fast development of new NGS technologies, the willingness for adopting standards in an industrial environment and the validity of general standards across different NGS applications. Therefore, a variety of different standardization approaches and projects from organizations, societies and companies were identified, which aimed to tackle the aforementioned issues:

- New York State Department of Health
  - Developed and published the NGS guidelines for somatic genetic variant detection
- Genome in a Bottle Consortium
  - Described reference data, methods and standards for NGS
  - Defined possible reference materials for quality control
- Association of Biomolecular Resource Facilities – NGS Work Group
  - Identified and evaluated optimal strategies for NGS projects
- College of American Pathologist's
  - Developed and published the molecular pathology checklist
- External RNA Control Consortium
  - Developed RNA spike-in samples for quality control
- Genomics Standards Consortium
  - Developed and published the minimum information about a genome sequence specification
- Sequencing Quality Control Project
  - Developed benchmarks with reference samples for platform comparisons
- Human Genome Project
  - Developed the Bermuda Standards
- NGS – Standardization of Clinical Testing Work Group
  - Developed guidelines and recommendations for implementation of NGS into diagnostics laboratories

The standardization approaches of the aforementioned groups were based on suggestions for standards, reference materials and guidelines:

- NGS guidelines for somatic genetic variant detection
  - Include validation parameter like i.a. accuracy, repeatability and reproducibility
  - Define reference materials for quality control
- Clinical Laboratory Improvement Amendments guidelines
  - Exhibit validation parameters for NGS
- Molecular pathology checklist
  - Describes standards for i.a. documentation, validation and confirmatory testing
- Minimum information about a genome sequence specification
  - Shows information requirements for data submission
  - Involves i.a. depth of coverage, overall quality and taxonomy
- Bermuda Standards
  - Include standards for sequence fidelity

Finally, respective proposals for standards, *de facto* standards and quality documentation issues were collected and developed, serving as a tangible contribution mainly addressing the upstream pipeline in order to ensure sufficient final data quality:

- DV<sub>200</sub> and RNA Integrity Number for quality control of formalin-fixed, paraffin-embedded RNA samples
- φX174 DNA as a spike-in sample for quality control of sequencing experiments
- OD<sub>260/280</sub> and OD<sub>260/230</sub> for sample purity determination and quality control
- FASTQ, SAM/BAM and VCF file formats as *de facto* standards in the bioinformatic
- Q30/Q40 as *de facto* standards/benchmarks for sequence quality
- Technical note for quality documentation of sequencing experiments
- Laboratory accreditation checklists for validation of sequencing workflows

## Conclusions

There is already a distinct number of NGS standardization efforts present; however, the majority of approaches target the standardization of the bioinformatics processing pipeline in the context of “Big Data”. Therefore, an essential prerequisite is the simplification and standardization of wet laboratory workflows, because respective steps are directly affecting the final data quality and thus there exists the demand to formulate experimental procedures to ensure a sufficient final data output quality. Additionally, comprehensive development of standards within the upstream part will likely lead to an accelerated standardization process in the data processing pipeline. Finally, NGS reflects as a prime example the difficulties of standardization in a highly dynamic environment, thus successful establishment of NGS standards will have the potential to be adopted into other areas of life sciences.

## References

1. Endrullat, C., Glökler, J., Franke, P., Frohme, M. Standardization and quality management in next-generation sequencing. *Appl. Transl. Genomics*, 2016, 10, 2–9. doi:10.1016/j.atg.2016.06.001
2. Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., Lu, F., Lyon, E., Voelkerding, K.V., Zehnbauser, B.A., Agarwala, R., Bennett, S.F., Chen, B., Chin, E.L.H., Compton, J.G., Das, S., Farkas, D.H., Ferber, M.J., Funke, B.H., Furtado, M.R., Ganova-Raeva, L.M., Geigenmüller, U., Gungelmann, S.J., Hegde, M.R., Johnson, P.L.F., Kasarskis, A., Kulkarni, S., Lenk, T., Liu, C.S.J., Manion, M., Manolio, T.A., Mardis, E.R., Merker, J.D., Rajeevan, M.S., Reese, M.G., Rehm, H.L., Simen, B.B., Yeakley, J.M., Zook, J.M., Lubin, I.M. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.*, 2012, 30, 1033–1036. doi:10.1038/nbt.2403
3. Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E., Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med. Off. J. Am. Coll. Med. Genet.*, 2013, 15, 733–747. doi:10.1038/gim.2013.92
4. Aziz, N., Zhao, Q., Bry, L., Driscoll, D.K., Funke, B., Gibson, J.S., Grody, W.W., Hegde, M.R., Hoeltge, G.A., Leonard, D.G.B., Merker, J.D., Nagarajan, R., Palicki, L.A., Robetorye, R.S., Schrijver, I., Weck, K.E., Voelkerding, K.V. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch. Pathol. Lab. Med.*, 2015, 139, 481–493. doi:10.5858/arpa.2014-0250-CP