Understanding Genomic Sequencing Workflows Through Data-Driven Analysis and Pilot Design

Author: Claudia A. Rodríguez Rosario
Advisor: Miriam Pabón, Ph.D.
Master in Manufacturing Competitiveness
Graduate Project EXPO, May 2026

Abstract

This study developed and designed a data-driven process improvement pilot for COVID-19 genomic sequencing at the Research Institute of the Puerto Rico Science, Technology & Research Trust, based on workflow mapping and analysis. The workflow includes key stages such as RNA extraction, library preparation, sequencing, and bioinformatics analysis, where variability can propagate across interconnected processes and impact overall performance. Based on this analysis, the study develops a structured framework to identify sources of variability and proposes a data-driven pilot design to evaluate potential process improvements in a controlled environment.

Introduction

Next-Generation Sequencing (NGS) has become a pivotal tool for SARS-CoV-2 genomic surveillance and public health response. As sequencing demand increased in Puerto Rico following the COVID-19 pandemic, laboratory workflows required greater coordination across clinical laboratories in the island. Variability throughout the sequencing workflow can impact turnaround time, workflow consistency, and resource utilization. This study evaluates operational variability within the COVID-19 sequencing workflow at the Research Institute of the Puerto Rico Science, Technology & Research Trust through workflow mapping and data-driven process analysis.

Background

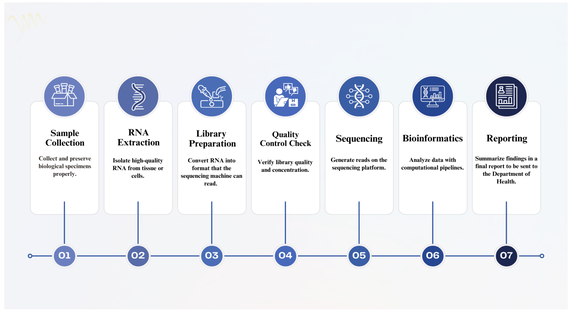
The SARS-CoV-2 sequencing workflow consists of multiple interconnected processes including RNA extraction, library preparation, quality control (QC), sequencing, bioinformatics analysis, and reporting. This study was conducted in collaboration with the Research Institute of the Puerto Rico Science, Technology & Research Trust, a state-of-the-art research facility focused on advancing scientific research, genomic surveillance, and public health response efforts in Puerto Rico. During library preparation, samples are converted into a sequencing-ready format through a series of preparation and validation procedures designed to support sequencing consistency and data reliability. QC procedures evaluate sequencing readiness before progression into sequencing analysis [1], [2]. Samples that fail QC checkpoints may require reprocessing, which can introduce workflow interruptions, increase turnaround time, and delay final reporting for up to 2 days.

Problem

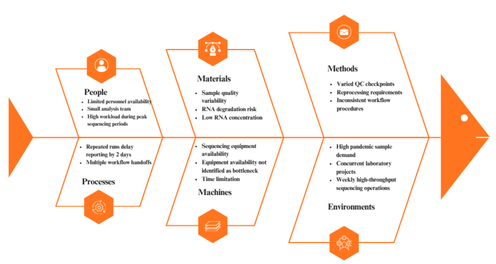
Increasing sequencing demand has amplified workflow variability, rework, and operational delays throughout the SARS-CoV-2 sequencing process, highlighting the need for a structured workflow improvement framework.

Methodology

This study utilized a data-driven case study approach to evaluate variability within the SARS-CoV-2 genomic sequencing workflow at the Research Institute of the Puerto Rico Science, Technology & Research Trust. Workflow mapping, operational metrics, and process performance analysis were used to identify inefficiencies across RNA extraction, library preparation, sequencing, bioinformatics analysis, and reporting [3]. A root cause analysis was also conducted to pinpoint operational factors contributing to workflow variability, rework, and reporting delays throughout the sequencing process [4].



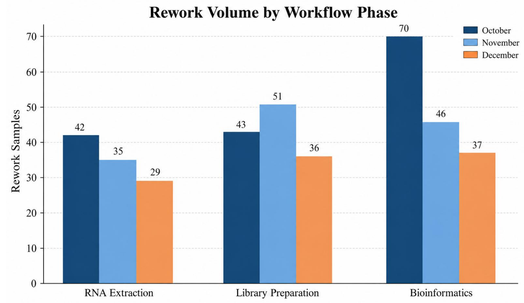
Workflow variability across sequencing stages can impact reporting time and process consistency.



Root cause analysis identified key operational factors contributing to workflow variability and reprocessing.

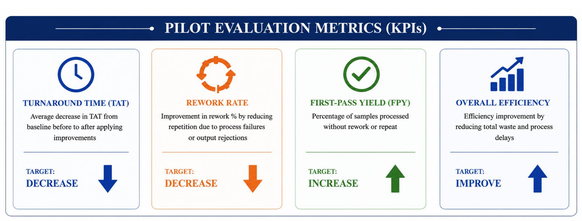
Results and Discussion

Process performance analysis was used to evaluate rework and operational variability across sequencing workflow phases.



Based on the identified sources of variability, the study developed a structured pilot framework focused on process standardization, quality control optimization, and workflow monitoring to support improved workflow consistency and operational efficiency.

Key performance indicators (KPIs) were identified to evaluate the effectiveness of the proposed pilot implementation.



Conclusions

This study identified workflow variability and rework as major contributors to operational inefficiencies within the SARS-CoV-2 genomic sequencing workflow at the Research Institute of the Puerto Rico Science, Technology & Research Trust. Results demonstrated that bioinformatics-related rework, workflow coordination challenges, and personnel limitations significantly impacted turnaround time, workflow consistency, and sequencing throughput. Through workflow mapping and operational analysis, the study validated staffing constraints as a primary operational bottleneck and established a structured foundation for future process optimization and workflow improvement efforts.

Future Work

Future work should focus on implementing the proposed pilot plan and monitoring key performance indicators (KPIs) to evaluate improvements in workflow efficiency, turnaround time, and process consistency. Continued operational analysis may also help support future staffing expansion efforts to improve sequencing throughput and reduce workflow delays.

Acknowledgements

The author would like to acknowledge Dr. Miriam Pabón for her guidance and support throughout this project. Special appreciation is extended to Dr. Lorraine Rodríguez and the laboratory personnel of the Research Institute of the Puerto Rico Science, Technology & Research Trust for their collaboration and operational insight. As discussed throughout this study, “the findings validated what the laboratory team had already identified as the primary operational bottleneck within the sequencing workflow.”

References

[1] Illumina, Inc. (2022). Illumina COVIDSeq RUO kits reference guide (Document No. 1000000126053 v08). Illumina.
[2] Next-Generation Sequencing (NGS) | Explore the technology. (n.d.). [Online] Available: https://www.illumina.com/science/technology/next-generation-sequencing.html
[3] Harte, Andrew. “Improving Lab Performance with Six Sigma | BSM.” Bsmlean.com, 18 Sept. 2018, [Online] Available: https://bsmlean.com/blog/andrew-harte/improving-lab-performance-six-sigma
[4] Holmes, C., McDonald, F., Jones, M., Ozdemir, V., & Graham, J. E. (2010). Standardization and Omics science: Technical and social dimensions are inseparable and demand symmetrical study. OMICS a Journal of Integrative Biology, 14(3), 327–332. [Online] Available: https://doi.org/10.1089/omi.2010.0022