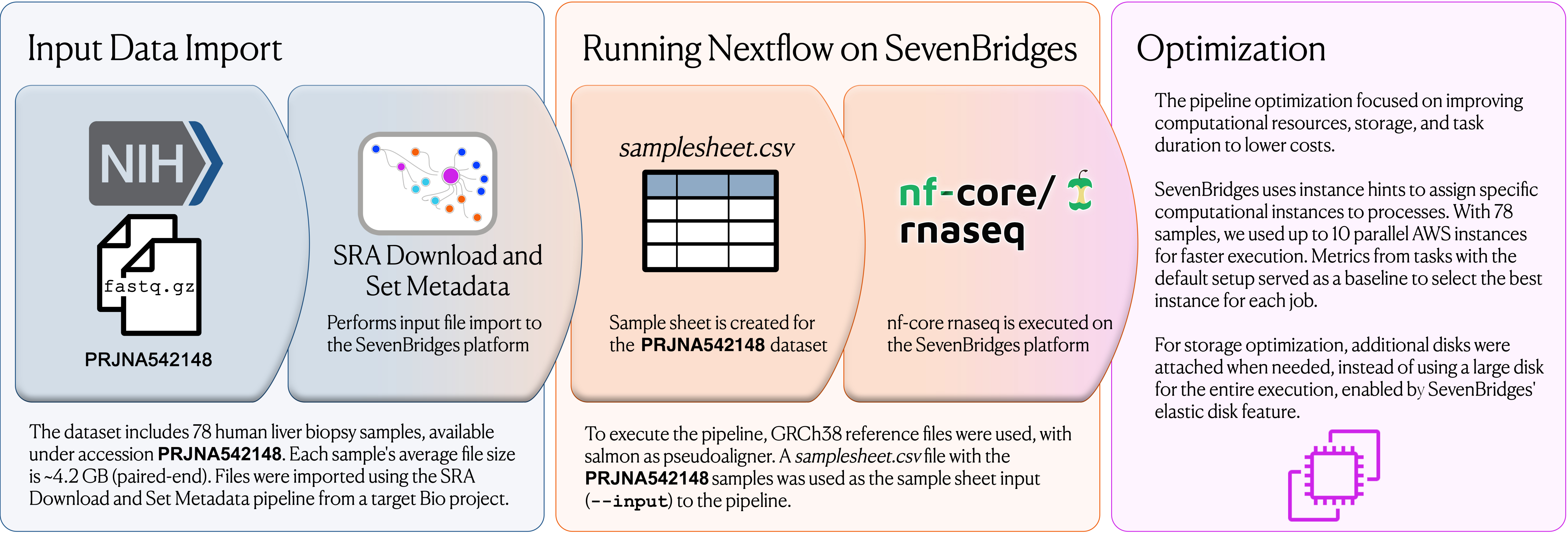


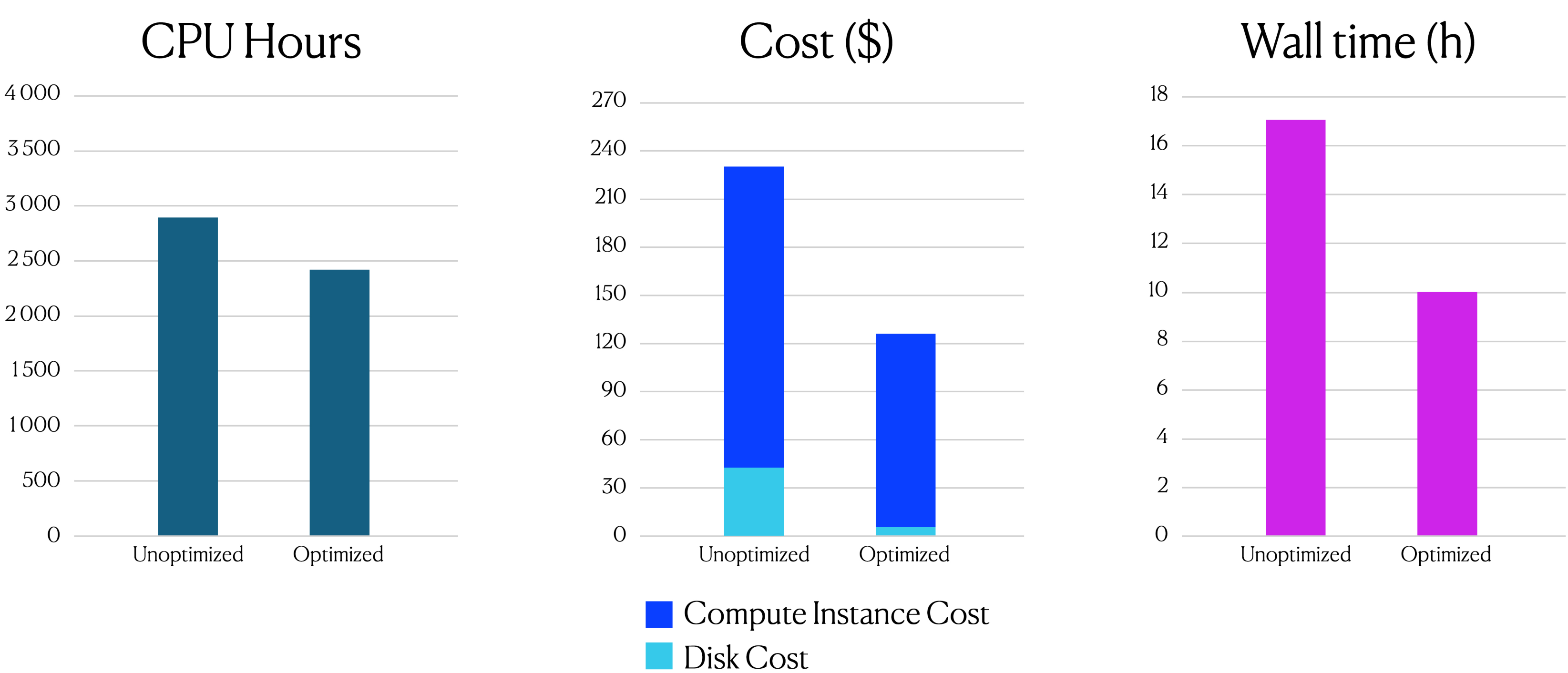
Optimization of **nf-core/rnaseq** pipeline using a large number of samples on the SevenBridges Platform

Darko Čučin (darko.cucin@velsera.com), Pavle Marinković (pavle.marinkovic@velsera.com)
Velsera, 529 Main St, Charlestown, Boston, MA 02129, United States

The nf-core/rnaseq (v3.12.0) pipeline is widely used and has significant community support. It is often benchmarked with the **test_full** dataset (8 samples) to compare bioinformatics cloud providers. However, since RNA-seq experiments usually involve more samples, we optimized the pipeline for larger datasets.



Results



Performance Efficiency

The optimized workflow uses 2420 CPU hours, compared to 2900 CPU hours for the unoptimized version, representing only 83.4% of the CPU hours required by the unoptimized version, significantly improving resource efficiency.

Cost Profile

Disk costs are reduced by 86.5%, and instance costs by 35.9%, reflecting better resource management. The unoptimized pipeline costs \$230.73, while the optimized one costs \$126.28, reducing total costs by 45.3%.

Execution Time

The optimized workflow runs 41.5% faster, completing in 10.00 hours versus 17.08 hours, leading to quicker results and greater efficiency.

The SevenBridges large inputs, and optimized test_full profiles are publically available at the github page found in the QR code to the right.

16.6% reduction in CPU hours.
45.3% reduction in total cost.
41.5% reduction in wall-time.

To establish a baseline and highlight potential cloud cost savings, further analysis was done using the test_full dataset to reduce execution costs.

SevenBridges on-demand was 27.0% more CPU-efficient, 65.6% cheaper (\$34.50 vs. \$100.34) and 11.7% faster than AWS Batch on-demand. For spot instances Seven Bridges was 54.6% cheaper (\$17.05 vs \$37.60), 26.0% more CPU-efficient and 14.0% faster than AWS Batch.

Overall, the Seven Bridges Platform outperformed AWS Batch in achieving better execution times and CPU usage at a lower cost, making it the more efficient cloud-based solution.

