Methylation and Beyond: Accelerating cfDNA Epigenetics and Fragmentomics Profiling with Cloud-Native and GPU Solutions

Nihesh Rathod, Kannadhasan Sekar, Swaraj Basu; **Strand Life Sciences, Bangalore, India**

VISIT OUR **WEBSITE**

Contact

Jaya Singh
Senior Director, Variant Science Solutions

📞 +1 (916) 873-4726 ✉ jaya@strandls.com

Introduction

Our cloud-native genomics pipeline simplifies workflows, boosting biomarker discovery, early disease detection, and clinical assay development to accelerate liquid biopsy-based diagnostics. Some of the applications are:

- **Cancer Genomics:** Develop methylation-based assays for screening and prognosis
- **Disease Diagnostics:** Enhance sensitivity in low-input liquid biopsy samples
- **Population Genomics:** Uncover epigenetic markers at scale
- **Environmental & Exposure Studies:** Link DNA changes to external factors

Analysis Pipeline

Pipeline consists of Fragmentomics, Methylation Profiling, and Variant Calling modules along with quality checks.
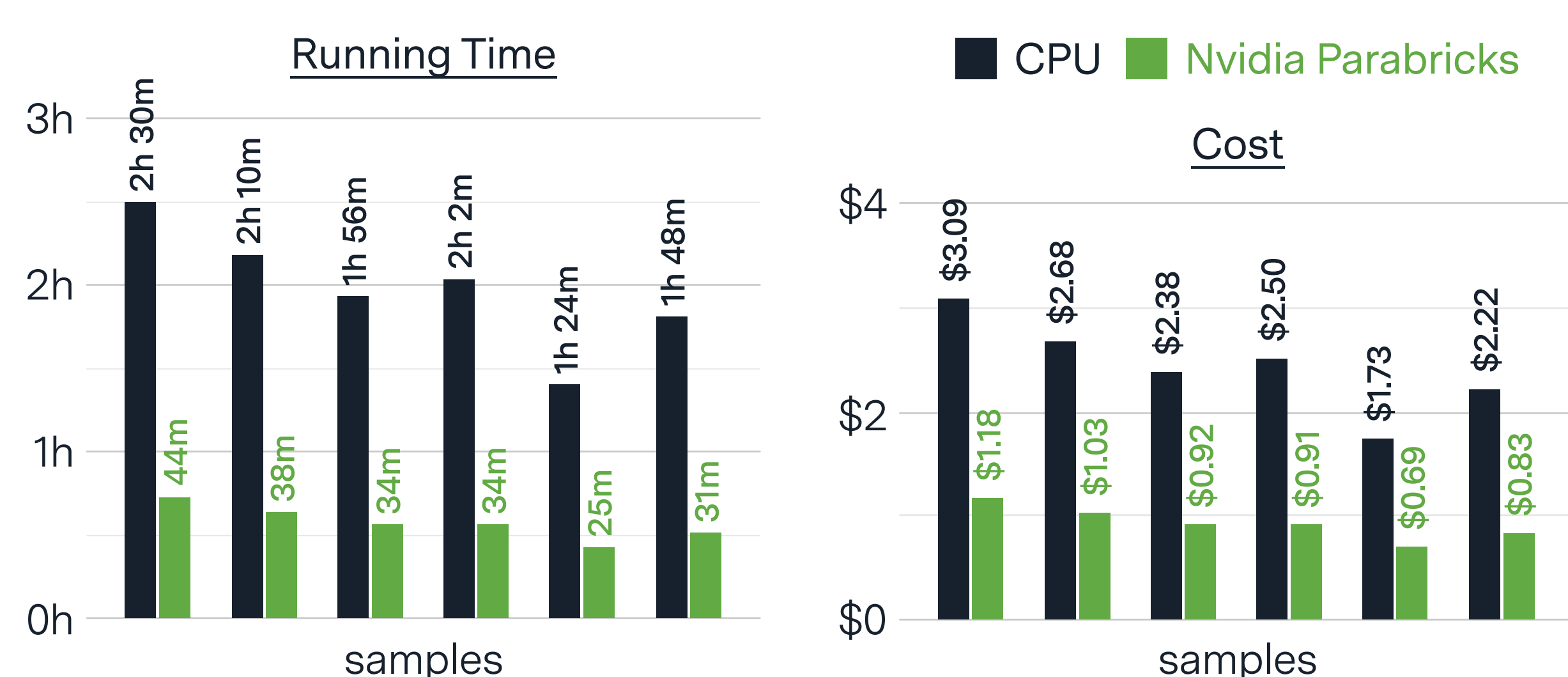
- **Fragmentomics:** Analyzes DNA fragment size, distribution, and patterns.
- **Methylation Profiling:** Maps DNA methylation patterns to assess gene regulation and identifies epigenetic biomarkers linked to diseases like cancer.
- **Variant Calling:** Detects genetic variations (mutations, insertions, deletions).
- **Quality Control (QC) Module:** Monitors sample integrity from collection to sequencing that ensures high-quality, reliable results for research and diagnostics.
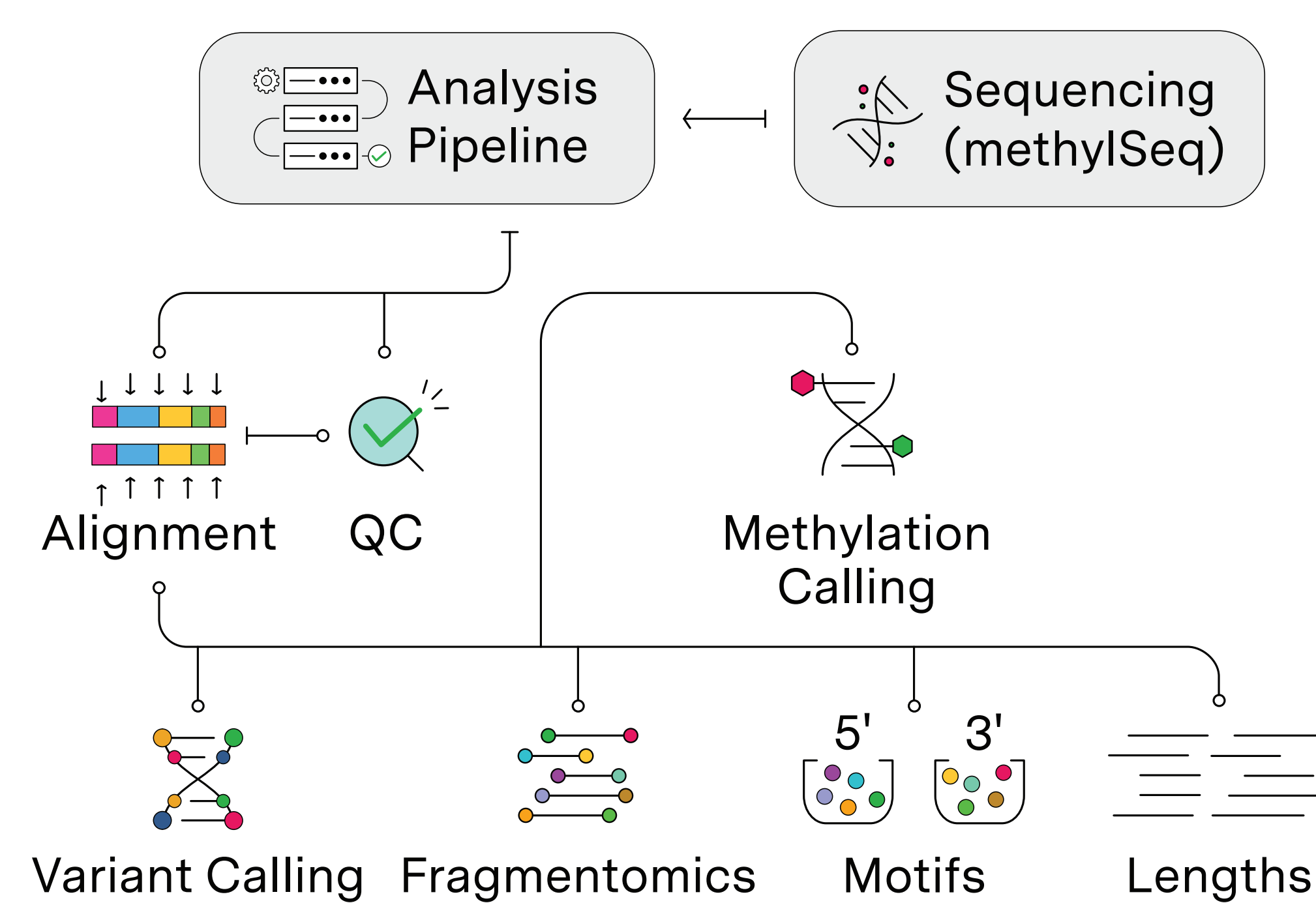
BWAMeth on CPU vs GPU

Machine Specifications
- CPU: c5a.8xlarge instance (32 vCPUs, 64 GB)
- NVIDIA Parabricks: g5.4xlarge instance (16 vCPUs, 64 GB, 1x A10G)

Running Time — CPU / Nvidia Parabricks: 2h 30m / 44m, 2h 10m / 38m, 1h 56m / 34m, 2h 2m / 34m, 1h 24m / 25m, 1h 48m / 31m

Cost — CPU / Nvidia Parabricks: $3.09 / $1.18, $2.68 / $1.03, $2.38 / $0.92, $2.50 / $0.91, $1.73 / $0.69, $2.22 / $0.83

Summary

Cloud Native Pipeline

- Combines fragmentomics, methylation profiling, and variant calling into one streamlined workflow that eliminates the need for multiple tools or pipelines, saving time and reducing complexity.
- Includes a quality control (QC) module at each stage of the pipeline to ensure data integrity and accuracy throughout the analysis process.
- Leverages AWS infrastructure to enable automated workflows for scalability, reliability, and high-performance computing for largescale genomic analyses.
  - Optimized to dynamically use resources as per individual modules, (e.g. RAM, CPU or both heavy) and able to complete its analysis on upto 100 targeted methylSeq samples at 50Gb data per sample (100x coverage) **within 15 hours**.
  - Each module is comprised of tools and methods recommended under best practices for methylSeq and fragmentomics analysis, including those given by EpiQC and FinaleDB studies. Further the pipeline reproduces signatures defined in both EpiQC and FinaleDB studies (on select samples).

BWAMeth on GPU

- Achieves **3.5 times faster** processing using just a single GPU compared to CPU-based implementations makes it ideal for time-sensitive workflows.
- Reduces turn around time (TAT) of samples.
- **2.6 times less expensive** than a CPU instance with a similar price point.
- Matches the accuracy of traditional CPU implementations.

Available as StrandLS Methylation Analysis Platform

- https://strandls-ed-app.us-east-1.elasticbeanstalk.com Registration for the service must be completed by a Strand administrator to ensure security.

AWS Infrastructure