An Automated Framework for Cloud-Based RNA-seq Pipeline Execution and Analysis Ready Outputs

Alissa Cait[1], Haider Hassan[1], Manjot Hunjan[1], Yassine Moumen[1], Anthony Manzella[1], Amy Pulaski[1], Jessica Corrado [1], Dan Ryder [1,2]
[1] Bridge Informatics, Inc., 160 Alewife Brook Pkwy, Cambridge, MA 02138;
[2] Corresponding author and Chief Executive Officer (CEO) of Bridge Informatics, Inc.

Background

RNA-seq analysis workflows rely on well-established tools for QC, alignment, quantification, differential gene expression, and pathway analysis. Yet, their integration is often **ad hoc, analyst dependent, and infrastructure-constrained.** This leads to inconsistent outputs, even for identical datasets.

After workflow execution, users face large collections of intermediate files with no clear path to downstream analysis, requiring substantial hands-on effort to validate metrics, regenerate figures, and debug failed steps.

The Problem

Manual RNA-seq Workflow

~30 manual steps | 4 decision checkpoints | Results vary run-to-run | Hours of active setup per project

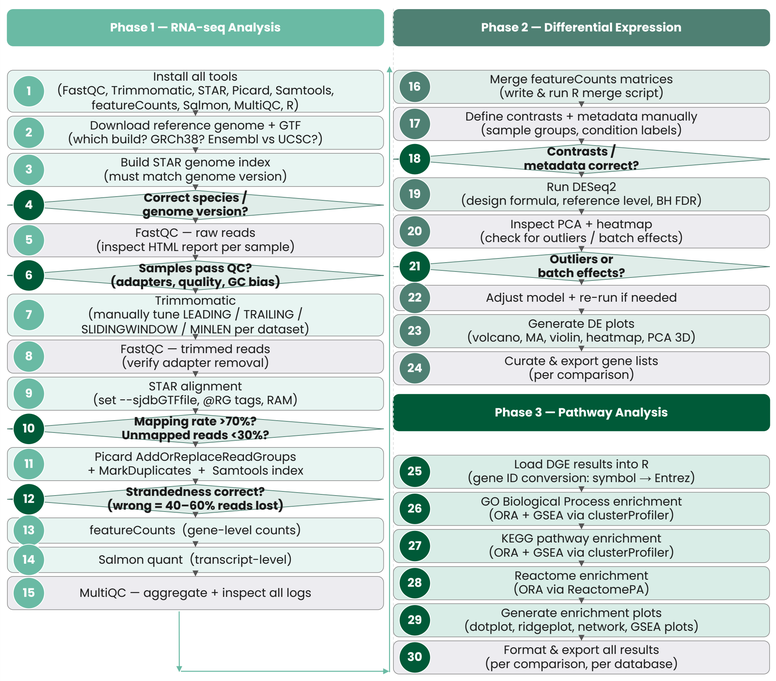
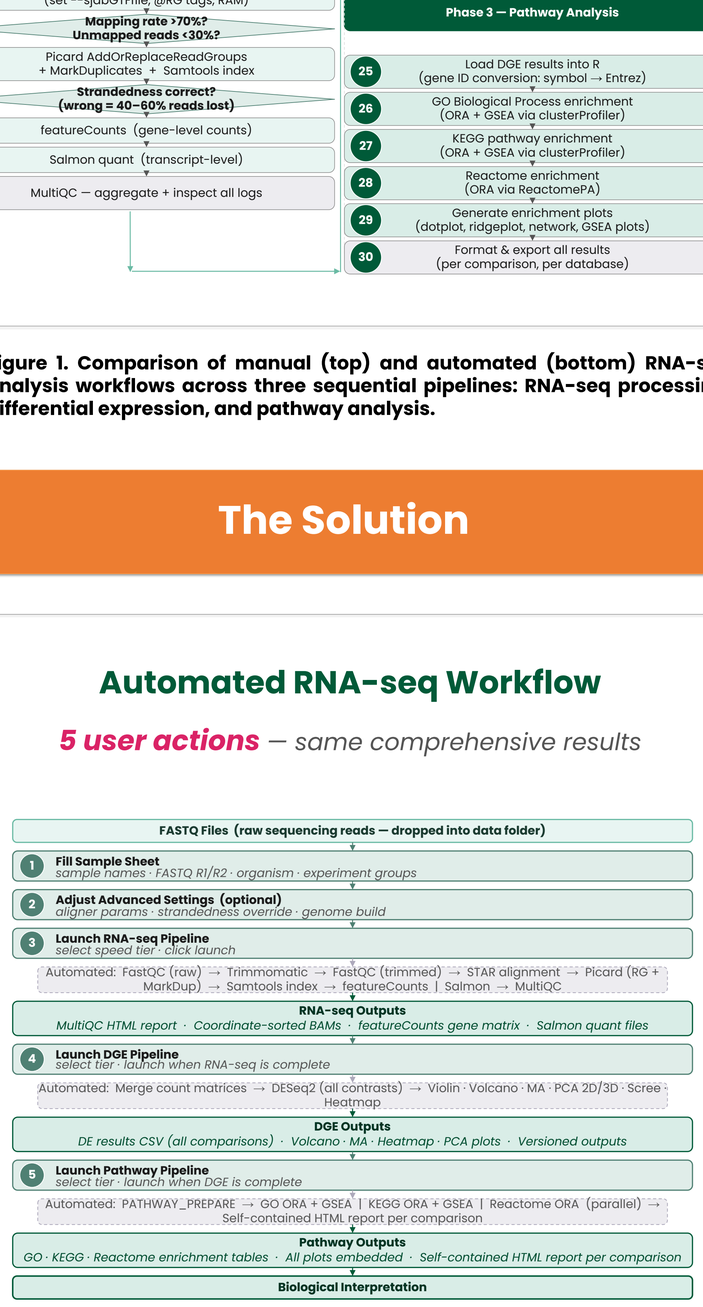


Figure 1. Comparison of manual (top) and automated (bottom) RNA-seq analysis workflows across three sequential pipelines: RNA-seq processing, differential expression, and pathway analysis.

The Solution

Automated RNA-seq Workflow

5 user actions *— same comprehensive results*



Where Does the Time Go?

Across three pipelines, end-to-end pipeline time breaks down into four categories. The automated framework shifts effort away from hands-on and rework time toward idle compute (ie- jobs running without human intervention)

Time allocation: manual vs. automated workflow

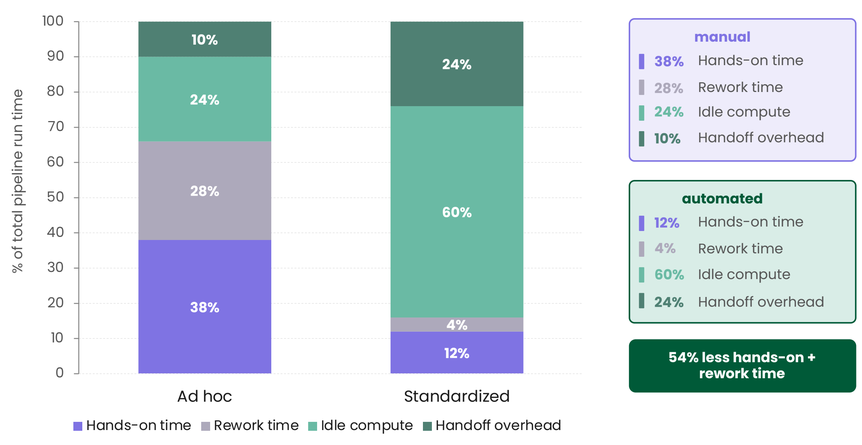


Figure 2. Distribution of total pipeline time by activity type for manual and automated RNA-seq workflows. Values are illustrative estimates; hands on and rework time are substantially reduced under the automated framework, while idle compute increases as jobs run without manual intervention.

The Analyst Effort: Manual vs Automated

The time required to execute a manual RNA-seq workflow varies considerably depending on the analyst's experience, familiarity with individual tools, and local computing environment. Before a single read is processed, the analyst must install and configure each tool independently and obtain the appropriate reference genomes and annotation files. This setup burden doesn't exist in the automated framework, where all dependencies are pre-configured and references are managed centrally. The timelines below should be interpreted as illustrative: an experienced analyst may complete the manual workflow faster, while a less experienced one may take considerably longer.

Three-pipeline RNA-seq workflow: manual vs. automated

Figure 3. Gantt chart comparing manual and automated RNA-seq workflows across three sequential pipelines: RNA-seq processing, differential expression analysis, and pathway analysis. Each step is shown as paired bars (manual = upper, automated = lower) plotted against a shared time axis scaled to the total manual wall time. The automated framework completes all three pipelines in approximately 35% of the manual total, with the remaining time saved indicated by the shaded region. Red dots denote manual intervention and failure points in the manual workflow. Manual pipeline launches between stages are preserved in both workflows. All values are illustrative estimates.

Key Outcomes

2–3X FASTER
ITERATION CYCLES VS STANDARD BIOINFORMATIC WORKFLOWS

↓85%
MANUAL CHECKPOINTS

REMOVES 12+
AD-HOC FAILURE POINTS

ZERO
BACK-AND-FORTH OVERHEAD; ACCESSIBLE TO NON-BIOINFORMATICIANS

Key Deliveries

- **Industry standard workflow:** All tools are selected and validated by PhD level bioinformaticians
- **Validated handoffs:** Pipeline outputs are always compatible with downstream pipelines
- **Reproducibility:** Containerized execution, locked parameters and schema validation across analysts
- **Flexibility:** Analysis parameters, such as p-value thresholds, can be easily adjusted
- **Analysis-ready outputs**: No post-hoc figure regeneration or table mismatches
- **Scalable:** Pilot studies and large cohorts have the same framework
- **Easy handoff:** Bioinformatics to biology via self-documenting reports
- **Share results:** Built in share feature allows you to share results and intermediate files with your team members

Let's Connect

www.bridgeinformatics.com
/company/bridge-informatics
@Bridge_Info_USA
dan.ryder@bridgeinformatics.com