

# Rapid Scale-up of Exome Sequencing Research using a Commercial Lab and Data Management Solution

Johanna Swanson, Data Analyst, Genome Sciences-Genomics Resources Center, University of Washington | Craig Parman, Field Applications Specialist and Mike Sanders, Product Manager, Genologics Life Sciences Software

## Abstract

In mid-2009 The Nickerson Lab in the Department of Genome Sciences at the University of Washington faced the daunting task of rapidly scaling up their next generation sequencing laboratory.

As a participant in a \$64.5 million dollar NHLBI grant for exome sequencing, a laboratory capable of sequencing 2,000 exomes per year was to be developed.

Given the enormous amount of data to be generated, a critical part of this scale-up was picking the right LIMS solution. This system needed to handle data flow from sample submission to results reporting.

Here we describe how the Geneus system was successfully deployed as the backbone LIMS platform for a high throughput DNA sequencing lab with 15 Illumina Genome Analyzers, 6 cluster stations, and more than 30 people.

## Introduction and Objectives

In October of 2009, the National Heart, Lung, and Blood Institute awarded a grant to support sequencing the exomes of more than 8,000 participants in the following 12 long-term studies:

1. The Women's Health Initiative
2. Atherosclerosis Risk in Communities
3. Coronary Artery Risk Development in Young Adults
4. The Cardiovascular Health Study
5. The Jackson Heart Study
6. The Multi-Ethnic Study of Atherosclerosis
7. The Framingham Heart Study
8. Genomic Research on Asthma in the African Diaspora
9. The Lung Health Study
10. A Pulmonary Arterial Hypertension population
11. A Acute Lung Injury cohort
12. A Cystic fibrosis cohort

The goal of the project; NHLBI hopes to identify specific gene variants that are associated with heart, lung, and blood diseases.

As part of the grant, The Nickerson Lab needed to create a high-throughput next generation sequencing lab. A critical component for this lab would be identifying a LIMS capable of supporting the complicated workflows and large amounts of data generated.

To meet the needs of the sequencing lab the following key requirements were indentified:

- Must meet the 6-month deployment schedule
- System must be scalable to meet future growth requirements
- Workflows must be easily adaptable to constantly changing protocols
- Key process QA/QC measurements must be tracked and easily accessible for reporting and analysis
- Integration bar coding and robotics are required
- Samples to be run are very precious, so there must be reliable sample tracking and traceability

Several options for the LIMS system were evaluated against the criteria. The timelines required that this system be fully deployed and capable of operating in a production environment within a few months. This rapid timeline excluded building “in-house” as an option. Based on these criteria GenoLogics’ Geneus system was chosen.

## Methods and Results

### Development

The Geneus LIMS is a server-client based system. The JAVA-based clients are run from any remote computer with network connectivity to the LINUX-based server. Workflows are created by connecting processes that reflect both wet bench and analysis processes. Each process has one or more inputs and outputs, which can be physical samples or data files. The system comes with preconfigured processes for common lab techniques. The system customization is done

through a GUI, which allows users to create new processes and workflows without underlying code changes. This allows changes to be made without spending time designing the server infrastructure, common data objects, or the lab client GUI’s.

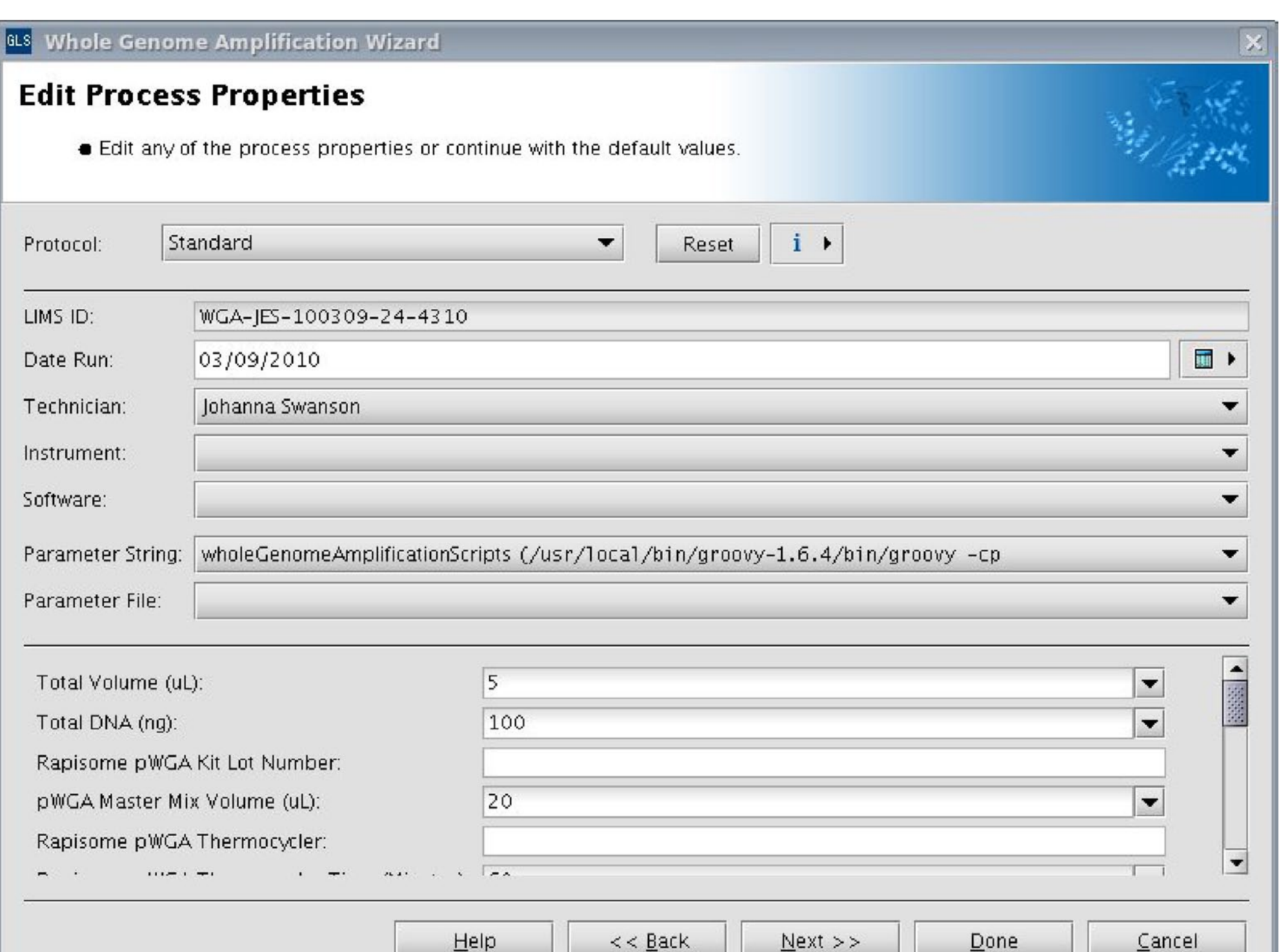


Figure 1: Geneus configuration tool

When configuring processes you have the option of executing a system command on the local server or on remote computers. By executing command line scripts we were able to add automated processes to handle tasks such as processing QC data, capturing output files, and creating robotics work lists. These scripts, written in Groovy, allowed users to increase lab efficiency and eliminate or automate many manual tracking steps. These scripts use the Geneus REST API for automatic data input and output to reduce technician work while ensuring consistent and repeatable tracking. The API exposes data resources such as samples and processes through HTTP web methods in XML format.

### Implementation

To support this project the following basic steps of the workflow steps were implemented in the LIMS:

- QC
- Exome capture
- Sequencing

Using the configuration described above, the LIMS system was configured to provide end-to-end sample tracking. The workflow developed to represent these processes is shown in Fig. 2. The process starts with a template provided to investigators for entering sample information. After this form is filled out and verified it is imported into the LIMS. These sample ID’s along with other sample information are tracked in the system throughout a project. The complete history of a sample along with links to all data files related to the samples is easily viewed in the sample genealogy. At each step, the details regarding the process are captured and associated with the samples.

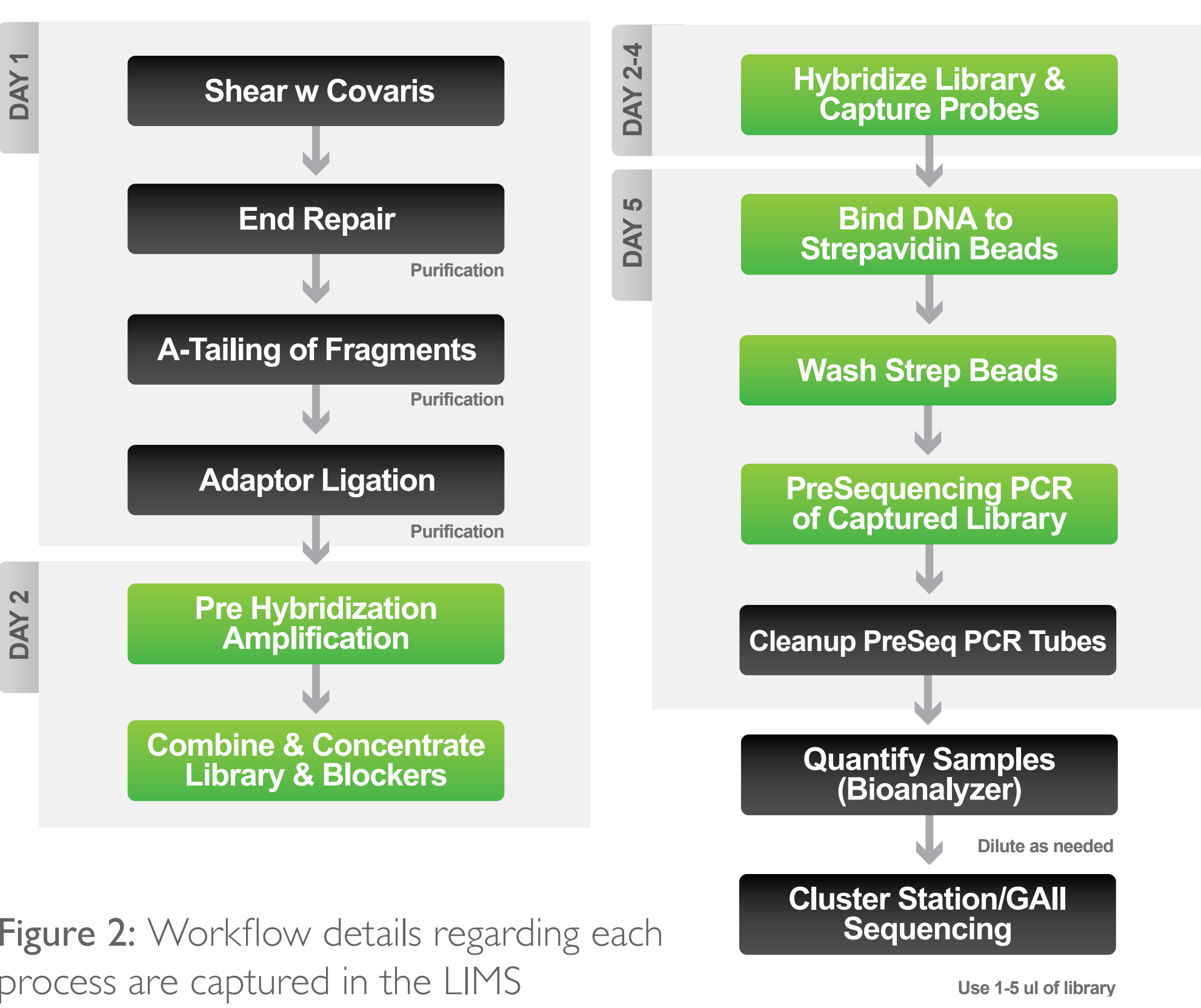


Figure 2: Workflow details regarding each process are captured in the LIMS

A large part of the data capture is focused on capturing the data from extensive QA/QC steps integrated into the process. These measurements provide QC information at various steps in the process and are used to make go / no-go decisions for individual samples, or used to validate sample tracking in the lab.

The key measurements include:

1. Generation of sample manifest sent with a barcode labeled plate to investigator to ensure accurate sample information
2. Picrogreen assay to ensure DNA concentration
3. Gender assay to validate plate sample locations
4. DNA fingerprinting with a 90-plex SNP assay, the data is compared to end sequence results to verify sample tracking
5. qPCR validation to check for sample degradation

All the data generated is captured by the LIMS. Scripts are used to parse the output files from QC tests and then automatically set QC flags in the LIMS. This eliminates the manual and error

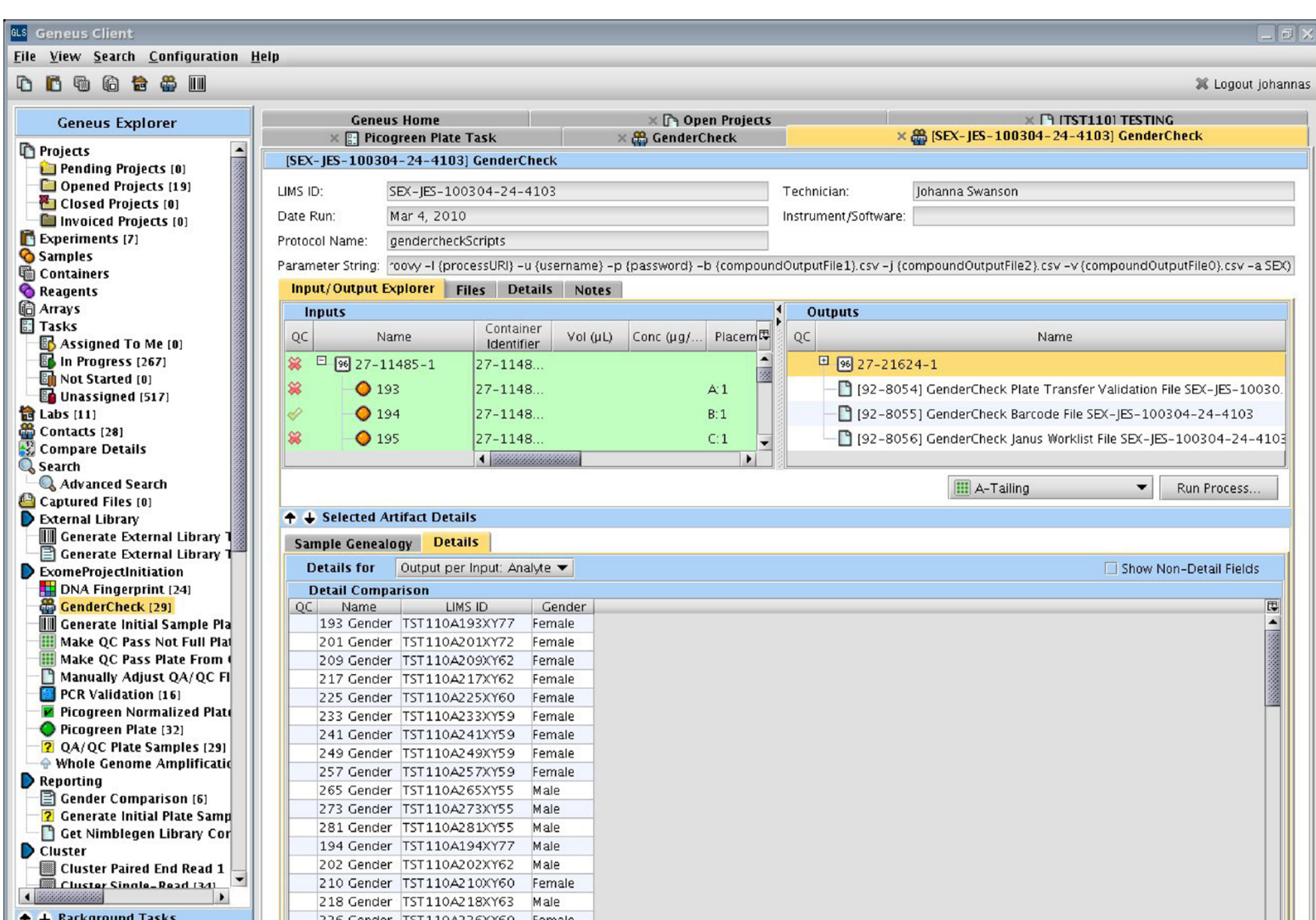


Figure 3: QC steps and sample genealogy with QC flags set

prone step of reviewing output files and typing in QC status.

All of the assays were implemented in a 96-well plate format, which require the automation of several robotic steps. Barcode reading capability was also required to eliminate any tracking errors due to plate swaps or rotations. A single process was configured to select samples that passed QC from multiple plates and build pick plates of passed samples. The system was configured so that the technicians could generate barcodes and robotics pick lists automatically.

## Conclusion

The Nickerson Lab in the Department of Genome Sciences at the University of Washington needed to develop and deploy a LIMS systems for their new Next Generation Sequencing Laboratory. A complete system capable of complete end-to-end sample tracking in a lab with 15 Illumina GA's was implemented and deployed in 6 months. They chose GenoLogics' Geneus system because of its flexibility and scalability. Because the Geneus system allows the user to make configuration changes, add scripts, or modify workflows without underlying code changes, the system was able to adapt to process changes during development. The end product manages all sample data, process information QC results and flagging, and bar coding with a single interface.