illumina®

Five Ways That a LIMS Accelerates NGS Research

How laboratory information management systems (LIMS) can become an essential business tool for modern genomics labs.

Introduction

Before 2010, genomics was an elite discipline, practiced by select labs associated mainly with marquee efforts such as the Human Genome Project. Today, genomics technologies have entered the mainstream. Labs everywhere can adopt high-throughput, next-generation sequencing (NGS) and sequence entire genomes in a matter of days. The bottlenecks often impeding research are now an inability to interpret and exploit NGS data with accuracy and efficiency.

Since the 1970s, research labs have turned to laboratory information and data management systems to streamline data throughput and reporting. The first LIMS were homegrown until commercial systems emerged in 1982. Today, LIMS is a mature class of life sciences software encompassing more than 30 open-source and commercial products capable of integrating instrumentation, automating workflows, and managing samples, users, and lab standards. However, most of these systems were developed primarily for pharmaceutical, chemical, or clinical labs. They have not been built specifically to address the needs of basic biological research and the unprecedented throughput and experimental complexity associated with using NGS.

This paper reviews five ways that a LIMS can accelerate NGS-based research. It focuses on the business and operational reasons that make a well-built, commercial LIMS a smarter investment than building a system from scratch. LIMS specifically built for genomics and NGS can help investigators to:

• Set up a lab quickly

• Manage samples, data, and staff

• Implement technologies and workflows essential to modern sequencing research

• Conduct faster, more collaborative research

• Publish findings sooner

Ways That a Commercial LIMS Accelerates NGS Research

1. Provides Scale to Keep Cutting-Edge Labs on Pace

The pace of change in the NGS market is staggering. In a single decade, sequencing costs have dropped and data volume has increased by several orders of magnitude. The Human Genome Project took 10 years and $2–3 billion to complete. With NGS, the cost of sequencing a single genome is now close to $1000, and the goal at Illumina is to reduce the cost even further.[1] Decreasing costs have been made possible by technological advances on high-throughput instruments. Currently, the NovaSeq™ System boasts a throughput of up to 6000 Gb for a 2×150 bp run, and is capable of reading up to 20 billion clusters in two days.

The increasing rate of genomes entered into public databases demonstrates the speed of change in sequencing technology. Less than a decade after the human genome was completed, over 1000 genomes had been entered into GenBank.[2] Today, sequenced human genomes number in the hundreds of thousands.

The implication for the modern sequencing lab is that it is possible to produce more data than can be managed and analyzed with efficiency and accuracy. Traditional lab and data management infrastructures rely heavily on spreadsheets, emails, and paper lab notebooks to manage and trace samples; and to record and report data. Such methods are inefficient in the current research environment, where data storage, data management, and informatics are potentially the biggest collective hurdle to expanding NGS operations.

NGS results depend on contextual information about research goals, samples, library preparation, sequencing methods, and other parameters prior to data analysis. Scientists and bioinformaticians must access and gather the right experimental information rapidly to answer questions such as:

• What are the research goals?

• What is the taxonomy of the sample?

• Where are all the disparate data located?

• How were the data generated?

• Which analysis tools should be used, and which algorithms and parameters should be applied?

• What quality threshold should be applied to the data set?

Ultimately, too much information can have the same result as no information. Because data and information can only be useful when delivered in a form that can be rapidly interpreted and acted upon, NGS labs should aim to eliminate manual, repetitive processes. A robust LIMS centralizes sample management and workflows. Lab processes, protocols, and analyses can be specified, defined, integrated, automated, and pipelined to create a comprehensive lab workflow (Figure 1). Furthermore, labs using LIMS can readily swap
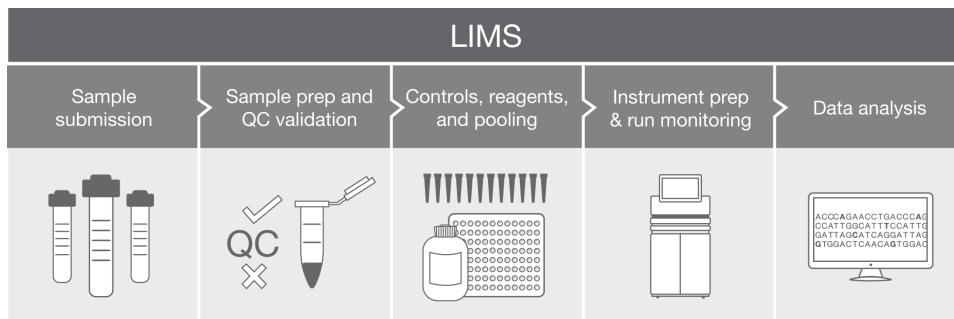
Figure 1: Typical NGS Workflow—LIMS can access and track samples through each phase of the workflow for NGS-based laboratories.

out methodologies, parameters, and instrumentation as technologies and workflows inevitably evolve. Scalability leaves labs well positioned to provide best-in-class service to collaborators and customers now and in the future.

2. Expedites Lab Setup and Production

NGS comes with an expectation of speed. Such an expectation should extend to the initiation, operation, and management of a lab. Grants have deadlines, and meeting those deadlines requires labs to ramp up quickly to accommodate NGS technologies and workflows, thus enabling expected research goals to be met and research results to be published quickly.

In a fast-paced research environment, labs are not likely to have time to build infrastructure software such as a LIMS. Such a project can take between 18–24 months. Furthermore, once the initial system is in place, it must be maintained and updated to accommodate new research objectives, protocols, workflows, and technologies. Most research centers dedicated to modern genomics can not afford the additional burden of becoming an expert in sequencing informatics.

Implementing commercial lab information management software designed specifically for genomics and sequencing can significantly shorten timeframes between lab setup and production. In 2–3 months, a commercial LIMS can be installed and provide real value to researchers. With a commercial LIMS, lab personnel receive support and training to ensure that the system is implemented appropriately to match specific workflows and processes. Training often focuses on system configuration and the sophisticated application program interfaces (APIs) that enable users to customize system capabilities to ensure future needs are met.

A commercial LIMS should also specify best practices and common workflow steps that keep labs on task and producing accurate, traceable data. A self-service web portal captures, organizes, and centralizes sample submission (Figure 2). Centralization eliminates the need to sort through the often confusing collection of emails, conference calls, and cryptically labeled samples received from collaborators.

Within the sequencing lab, the commercial LIMS provides end-to-end workflow management to ensure that samples are easy to trace and tied directly to relevant experimental details (Figure 1). Software interfaces enable lab staff to monitor project progress in real time, and publish specific results and reports rapidly to external collaborators or customers. Robust lab information management software enables a lab to be up and running quickly, and then keeps it running and operating with best-in-class efficiency.
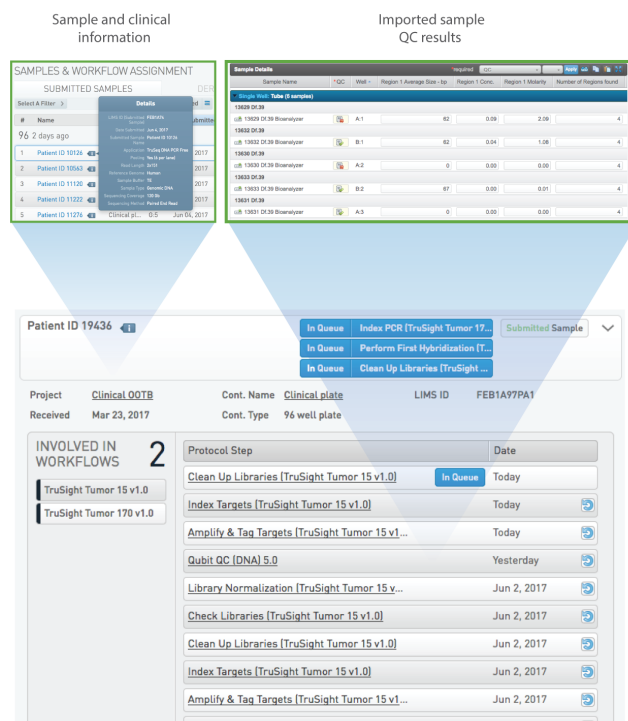

Figure 2: Web Portal Views of NGS Sample and Data Genealogy—Sample information is recorded and organized in a centralized web interface.

3. Improves Results with Sample Traceability

The unprecedented throughput enabled by modern sequencing instruments and multiplexing techniques makes it virtually impossible for labs to rely on manual workflows and disorganized, disparate desktop software or white boards to track samples. Clear sample traceability from the moment a sample enters the lab to the point when the results are ready to be reported and published is a necessity for sequencing labs. In fact, some funding agencies may require that research projects guarantee sample traceability. Sample tracking is essential when dealing with the often limited DNA supplies associated with certain clinical sample cohorts.

A commercial LIMS designed specifically for sequencing can address some of the more onerous aspects of sample tracking, including:

Correlating Wet Lab Sample Information and Experimental Data

The ability to link results to original sample data enables reliable *in silico* work and downstream analysis.

Validating and Setting Up Multiplexed Sequencing Runs

Any mismatch between a DNA barcode, sample, and the loading pattern on a sequencer can invalidate the results of an entire run. A LIMS can be configured to generate sample sheets or run definition files for sequencing instruments automatically. This simplifies and accelerates run times, and eliminates transcription errors from manual spreadsheets.

Organizing, Labeling, and Tracking Samples

The data volumes and experimental complexity associated with NGS can be overwhelming. Traditional barcoding and indexing systems can struggle to trace samples reliably back through complex sample pools to the specific phenotypic information associated with the samples. Lab information management software specifically designed for sequencing can accurately track individual components of complex sample pools from the initiation of a project through to its conclusion, and labs can benefit from their investments in sample management infrastructure.

Making Information Easy to Store, Manage, and Access

Too often, crucial sample and experimental information is spread across paper lab notebooks, emails, electronic documents, and spreadsheets. These disparate pieces of information may in turn be stored on different workstations or lab instrumentation. LIMS for sequencing centralizes this information so that sample preparation details, workflows, QC results, associated experimental details, and data analysis results remain accessible and retrievable years after projects are complete.

By implementing a best-in-class, commercial LIMS, labs can consistently deliver accurate, high-quality results experiment after experiment. LIMS also enable accurate traceability, which preserves lab credibility and reputation.

4. Manages Workflows to Improve Lab Efficiency

The number of NGS instruments and sequencing strategies that can be applied to projects is enough to overwhelm even experienced laboratory personnel. It is possible for lab managers, technicians, analysts, and bioinformaticians to spend the bulk of their day performing repetitive, administrative tasks associated with organizing and managing the data generated by their labs. Busy labs can also struggle to ensure that samples received from clients and collaborators are appropriately labeled so that all vital experimental context is passed on efficiently and accurately to experimentalists. Lack of workflow clarity can lead to miscommunications and inefficient sample handoffs between groups, potentially wasting precious samples, time, and reagents. Moreover, inefficiencies can negatively impact the credibility of a lab, particularly if they result in delays in achieving critical milestones for funded research programs.

A LIMS designed for sequencing centralizes all communications and workflow tasks. It also improves lab efficiency by eliminating many of the manual, tedious, repetitive tasks associated with day-to-day lab operations. A configurable user interface lets lab personnel link together common lab processes quickly and easily to create workflows that mirror various protocols within the lab (ie, sample preparation or data analysis). Time-consuming steps, such as representing and tracking library-pooling interactions, demultiplexing pooled samples, generating sample sheets for sequencers, or tracking critical quality control data and resulting pass/fail decisions, can be automated or assigned to specific personnel. Reporting and communications are also streamlined, because the system tracks and centralizes all the information associated with sequencing projects.

Workflow management ensures sample traceability, standardizes best laboratory practices, and ensures that personnel are always working on high-priority tasks. When codified in commercial LIMS designed for sequencing, workflow management enables labs to focus on their core research objectives and establish their reputations as best-in-class facilities.

5. Supports Faster Publishing

An important goal for principal investigators in academic labs and research centers is publication. Two factors speed publication: (1) how quickly experiments are conducted, analyzed, and reported; and (2) the quality of the data that is produced. These factors are supported by a scalable, adaptable LIMS that centralizes sample information and standardizes workflows. Faster, automated lab procedures lead to more efficient experiments and more reliable results. When reliable results are received quickly, research findings can be published sooner.

To support faster publishing, an informatics solution should be as good at helping scientists extract data from the system as it is at helping scientists enter data. A LIMS optimizes instrument efficiency; automates critical data capture, quality control, and analysis

pipelines; and keeps samples flowing through the lab while tracking associated data and keeping information easily accessible. It should also offer various options for accessing stored data, from on-demand searches to sophisticated reporting frameworks that can also be used to create custom queries and reports. A robust API can also help scientists and internal programmers extract data into the format needed for downstream analysis. After analysis is complete, scientists can choose from a number of ways to summarize and communicate results. Because all project-related information is stored in the system, lab managers can use the same system to generate operational reports for management or clients, and experimental reports to compare key experimental parameters with downstream quality metrics. Internal and external reporting can be streamlined by using a site-branded web portal that enables lab staff and their customers to share and collaborate securely in near real time.

Commercial LIMS is Essential for Modern Sequencing Facilities

The unprecedented data volumes and experimental complexity associated with modern sequencing make a commercial LIMS essential for labs seeking to establish their reputations as best-in-class facilities capable of high-powered research. A commercial LIMS provides the scale labs need to adapt to the changing technologies and methodologies in this field. It enables them to ramp up facilities to produce results quickly. It offers the sample traceability that many funding agencies demand on premiere sequencing projects. It also provides workflow management, data analysis, and reporting tools to ensure that labs run efficiently and collaboratively. All these capabilities can have a tremendous impact on labs, enabling them to:

- Reduce costs by improving reagent use and eliminating the experimental errors and doubt that lead to repeated experiments
- Enable a new sequencing center to be up, running, and processing samples in just a few months
- Accommodate the constant (sometimes daily) changes that are trending in labs conducting NGS
- Produce quality results consistently, speeding the publication process, and raising the "impact factor"
- Communicate better and more effectively with customers and collaborators
- Focus more on research by eliminating manual tasks and implementing best-practice-based workflows
- Conduct more research using current staffing levels and instrumentation, reducing facility overhead cost
- Build a reputation as a best-in-class facility, attracting more business and securing high-profile studies and additional grant money

Learn More

To learn more about BaseSpace Clarity LIMS, visit:
www.illumina.com/products/by-type/informatics-products/basespace-clarity-lims.html

References

1. NIH National Human Genome Research Institute. The Cost of Sequencing a Human Genome. www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/. Accessed July 31, 2017.

2. J.P. Morgan North America Equity Research. Next Gen Sequencing Survey. www.genomicslawreport.com/wp-content/uploads/2011/04/JP-Morgan-NGS-Report.pdf. Published May 12, 2010. Accessed July 28, 2017.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.