

Selecting the Right LIMS for Your Laboratory

A Practical Guide for the Next-Generation Sequencing (NGS) Lab

Introduction

The technologies and methods pioneered since the Human Genome Project have revolutionized the life science industry. The adoption of NGS has given organizations the ability to sequence entire genomes quickly and at low cost. In fact, few other industries have seen processing speeds rise, or costs drop, as dramatically as genomics. However, organizations now face the challenge of handling the massive amounts of data these techniques generate. To find a partner in managing NGS operations, many labs use a laboratory information management system (LIMS) that adapts and scales to meet their changing and complex needs.

LIMSS are a mature class of life science software, and commercial systems that are specifically designed for genomics are available. The ideal system should offer the following advantages to modern sequencing facilities:

- Comprehensive sample traceability from sample accession to results
- Scalability so that labs can modify sample throughput according to changing demands
- Efficient implementation so that labs can set up, run, and produce results quickly
- Adaptability to help labs accommodate changing technologies and methodologies
- Workflow management and operational reporting tools to ensure labs run efficiently and collaboratively

How do users ensure that a LIMS is a good match for their lab? This application note reviews three criteria that NGS-based labs can evaluate when selecting a LIMS:

1. How well does a LIMS integrate and support existing and future instrumentation?
2. How easy is a LIMS to customize and extend?
3. How well does a LIMS support and interact with users in their work environment?

1. Does the LIMS Support and Integrate With Existing and Future Instrumentation?

Most labs recognize the value of lab and data management. To parse and act on the enormous quantities of data produced by NGS, scientists need the ability to assign unique IDs to samples, record associated information, and track this information across the life cycle of the experiment. In modern labs, it is no longer practical to track experiments, manage library preparation, or monitor instrument runs using Excel spreadsheets, Google docs, or paper lab notebooks. Data management and experiment tracking are even more difficult for labs using DNA indexing (also known as bar coding or tagging) to pool and multiplex samples from diverse, unrelated sources on a single flow cell lane. These techniques can create bottlenecks at the library preparation step, where sheer throughput, combined with the need to track sample pooling, delays the rate at which labs can load samples onto sequencers.

A primary advantage of a LIMS is its ability to integrate with laboratory instrumentation. In a recent report, LIMS users reported that instrument integration is a key feature that they wish their LIMS included.¹ In labs using NGS, a LIMS must run and interface with instrumentation, provide a framework to appropriately capture data, and streamline and automate routine tasks. Eliminating bottlenecks can save time with tasks that potentially stall sequencing operations (Figure 1).

Three Phases of Instrument Integration

While each type of NGS instrument comes with vendor-specified kits and protocols to optimize performance, the task of integrating with sequencing instrumentation encompasses three primary phases. Each phase should be supported by a LIMS.

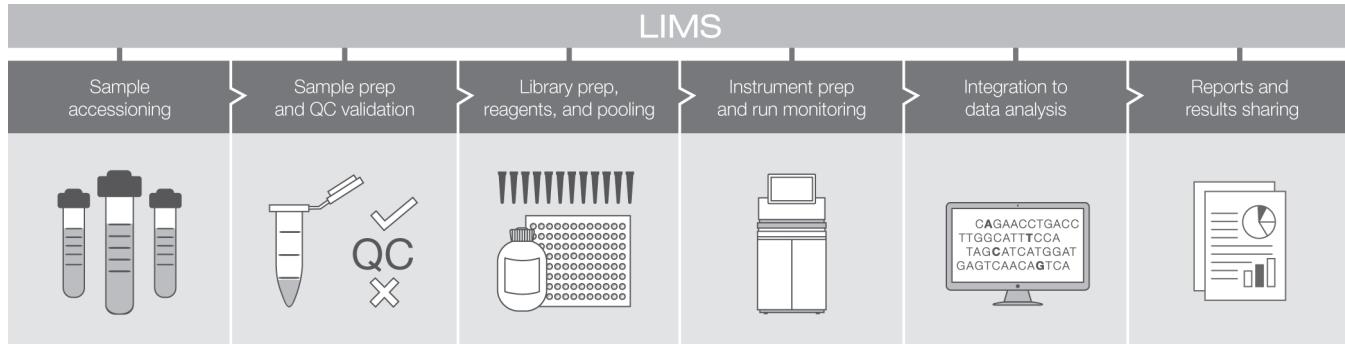


Figure 1: A Typical NGS Lab Workflow—The ideal LIMS helps to save time and avoid bottlenecks at each step between sample accessioning and sharing of data.

Collection and Integration of Sample Information

First, organizations must consider how they collect sample information and associate it with runs. Scientists have traditionally spent hours poring over Excel spreadsheets to check sample preparation and run assignments. However, a LIMS can automate the process of setting up a run. Scientists simply specify the samples they wish to run, and the LIMS automatically generates the appropriate files for the sequencing equipment.² Conversely, NGS instruments can be configured to transfer information from completed runs directly to the LIMS, reducing hands-on time for lab staff. How the LIMS integrates with the instrumentation may differ. Some LIMSS may integrate more tightly with particular instrumentation, and laboratories should verify the connectivity between LIMS and their preferred instrumentation. But any NGS LIMS should provide some level of integration with NGS instrumentation.² A well-designed LIMS provides confidence in sample handling while saving time and automating error-prone tasks in the overall NGS workflow.

Monitoring Quality of Sequence Data

The second phase of integrating instruments with a LIMS is configuring the LIMS to track the quality of sequencing data. Many sequencing instruments run for days, but waiting until runs are complete before evaluating data quality can be inefficient. In addition to monitoring the status of runs in progress, the ideal LIMS should also collect metrics, such as the total bases yielded from a run or the percentage of base calls meeting threshhold quality scores for accuracy.² Over time, these metrics can aid in assessing instrument performance. With data from sample runs archived and searchable in a centralized LIMS, labs can make informed decisions about which samples to reanalyze, whether to request more samples for further experimentation, whether to abort or continue a run, and which instruments may need service.

A LIMS gives organizations confidence in sample handling while saving time and automating tedious tasks in the overall next-gen workflow.

Tracking Results

The final consideration in integrating instrumentation effectively with a LIMS is results tracking. Many labs accumulate massive directories on their local area network dedicated to storing information associated with sequencing runs. Often this granular information surfaces in reports and summaries, while the underlying information is stored for future reference. Unfortunately, locating necessary detail can take staff hours or even days, leading some labs to repeat experiments rather than sift through directories for archived files. Multiplexing can also require an additional data management step because some samples must be unpooled or demultiplexed before the results can be analyzed and interpreted.

A LIMS can eliminate some of the most tedious aspects of NGS for lab managers and bioinformaticians. Intuitive query tools enable labs to collect information quickly for new sequencing runs and archived data. The ideal LIMS also simplifies creation of automated workflows that can handle such tasks as generating sample sheets for the sequencing instrument, demultiplexing reads, or incorporating specific open source and commercial analysis pipelines.² By eliminating the need to search and organize data, lab staff can spend more time on biological interpretation, decision making, publications, and innovation.

2. How Easy is the LIMS to Configure, Customize, and Extend?

In cutting-edge NGS research, constant change is an operational challenge. Methods can quickly become obsolete. Methods might not yet exist for new applications, requiring labs to develop and modify protocols. In this environment, labs succeed by pushing the boundaries of innovation, and cannot afford to be constrained by the software they implement to manage data and workflows.

No software package can effectively meet the needs of every lab—particularly in a field that's evolving as quickly as NGS. As a result, many labs often consider building their own lab and data management systems. Although home-grown systems do enable labs to design and implement exactly the system they want, labs should consider how a system will need to adapt to the inevitable changes in experimental applications. Such adaptation requires a critical investment of time, money, and personnel, and most NGS labs can not afford the additional burden of becoming an expert in software design and development.

While a commercial LIMS is a logical investment for NGS labs wanting to set up and run quickly, adaptability and scalability should be demanded of commercial solutions. Several commercial LIMS have been specifically designed for NGS, but these systems should be flexible about how work proceeds, without prohibiting changes to the configurations.² Labs can also work with broad enterprise LIMS vendors to build tailored systems using a combination of custom components. However, these components usually need to be custom designed for each new customer to create a complete NGS solution. This approach makes the initial implementation costly and slow, and ensures that when a lab's needs inevitably change, the vendor will need to update the system.

An effective commercial LIMS should not require additional support from the vendor every time the lab wants to adapt or improve the software. Software should instead be configurable and customizable by the lab team.² Unfortunately, the terms "configuration" and "customization" are often conflated, particularly in software marketing. In software engineering, configuration refers to changes in existing software that can be made via the user interface without additional programming or changes to the underlying code. In a NGS LIMS, configuration might be used to connect the system to preferred instrumentation, capture results, and set up the system to support general sample preparation and tracking. Customization is when the

actual code must be changed so that the software can do something new or different. Customization can have negative connotations to some users, often implying additional consulting costs or programming assistance that software vendors sell in addition to software. However, customization should be something that labs can undertake themselves when armed with the appropriate expertise (programmers), software tools, and application programming interfaces (APIs) from their LIMS software provider.

While many LIMSS offer programming interfaces, some of these may be built on proprietary code or scripting environments. To avoid future constraints, an API should accommodate both open-source and commercial bioinformatics tools, particularly scripting languages that are familiar to scientific programmers and bioinformaticians such as Groovy, PERL, or Python. Using the API, programmers or bioinformaticians can tailor a LIMS to handle a range of tasks beyond those available out of the box. Scripts can be developed to integrate and automate the system to interface the LIMS with robotics or instruments, collect information into a LIMS, or initiate computational processing tasks. Some of these customizations are exceedingly complicated and powerful, yet can be developed and deployed entirely by programmers or bioinformaticians within a sequencing lab without vendor assistance.³

Using the API, programmers or bioinformaticians can tailor a LIMS to handle a range of tasks beyond those available out of the box.

A flexible and powerful API enables labs to make a commercial LIMS their own, on their own. It also offers flexibility that can enable a vendor to customize the system rapidly should a lab lack the resources to create its own scripts or demand more extensive system engineering. Ultimately, a flexible API benefits both the purchasing lab and the supplying vendor by ensuring that changes can be made swiftly and efficiently regardless of the scope.³

3. Does the LIMS Accommodate Different Users and Workflows?

NGS labs require varied expertise to accomplish their objectives. Principal investigators, lab managers, lab technicians, scientists, bioinformaticians, and scientific programmers all contribute to keep experiments running quickly and efficiently. Many labs also provide sequencing services to other labs or collaborators. This means that labs must also coordinate communications with external collaborators. With distinct responsibilities and priorities, all individuals have different ways that they wish to view and act on data (Figure 2).

For example, lab technicians are interested in finding out what projects have been assigned and what other work is happening in the lab that will impact their routine. They are interested in planning their schedule and quickly recording information pertinent to laboratory tasks. For lab technicians, a LIMS should simplify their work by offering a straightforward user-specific interface that minimizes repetitive tasks.⁴ This enables technicians to spend less time organizing and recording what they have done and more time actually doing experiments.

In a dynamic, leading-edge genomics lab, one user interface does not fit all. To work efficiently, users should only see information that is relevant to their job, rather than struggling with a complicated user interface. Targeted user interfaces in a NGS LIMS can help everyone in a lab work more productively. These interfaces should provide a dashboard of relevant activities while also pulling appropriate data from the larger system and displaying it to those who need to act on it. Intelligently designed, targeted user interfaces can support the following types of users:

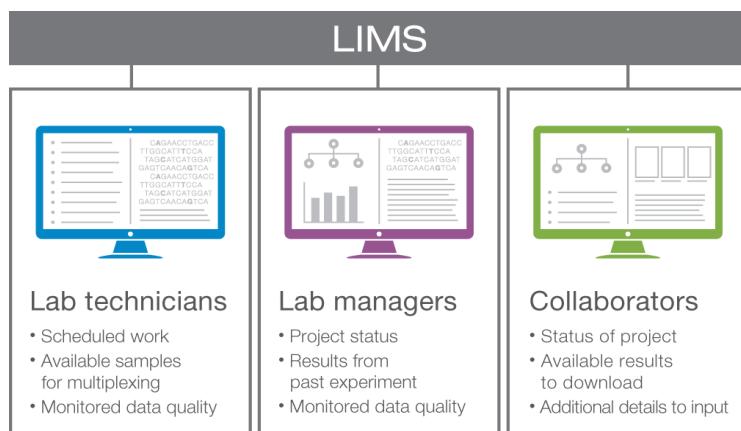


Figure 2: LIMS Should Accommodate Different User Types—With interfaces configured to access only the information that is relevant to each member of the lab, staff can spend less time navigating through unnecessary tasks.

Lab Technicians

Scientists and technical staff require fast, efficient access to information that helps them track sample status, determine which samples can be prepared together, simplify creation of library pools for multiplexed sequencing runs, and access and review past work. Dashboards should help them answer such questions as, “What experiments do I need to carry out today?” “What work is coming my way so I can plan ahead?” “Which libraries can I pool together for a multiplexed run?” or “Am I getting good quality data off that run I just started?”

Lab Managers

It is impractical, if not impossible, for lab managers to keep tabs on every activity occurring in a dynamic NGS lab. Management dashboards should provide a high level summary of everything happening in the lab, including an overview of active project status and instrument performance with the ability to drill down into activities to look at more specific results or metrics. Managers also require reporting and project management tools to manage client communications, invoicing, and administrative reporting. Interfaces should help managers answer such questions as, “What is the quality of data coming off that new sequencer?” “What is the status on the project we’ve been running for our new collaborator?” or “Where are the results from that experiment we did six months ago?”

External Collaborators

The LIMS should provide a secure portal ensuring that outside collaborators have immediate access to data relevant to their projects, while protecting the broader project data accumulated by the servicing lab. The portal should provide a centralized way for collaborators to initiate work requests, inquire about project status, and view project summaries. Through the interface, collaborators should be able to answer such questions as, “Is my project finished yet?” “Are there any results available to download?” or “I’ve got some additional details to provide. How can I get them to you?”

Conclusion

While NGS brings as many data management challenges as it brings scientific opportunities, the software needed to manage and communicate NGS data is mature and well understood. LIMS approaches have proven themselves across a range of industries for over 30 years. Many options exist, and what system is best for a given facility will depend on its size, scope, and research goals. The unique demands of NGS, however, make certain issues imperative:

- Can the LIMS help a lab set up, run, and produce results quickly, with minimal to no vendor intervention?
- How easy is it to adapt and augment the system when lab needs inevitably change?
- Does the LIMS provide relevant information to specific users so that they can do their jobs easily and efficiently?

BaseSpace Clarity LIMS was designed with numerous features to support genomics laboratories. With comprehensive sample traceability, preconfigured and configurable workflows, role-based interfaces, instrument integration, and extensibility through APIs, BaseSpace Clarity LIMS provides flexibility and adaptability to accommodate the changing needs of modern NGS-based laboratories.⁵

Learn More

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

References

1. Atrix Tehnology Group. 2016 LIMS Market Research Survey Report. vertassets.blob.core.windows.net/download/a43a4bdc/a43a4bdc-bfed-4d4b-b8b9-b0ba6d42f786/lims_market_research_survey_summary.pdf. Accessed July 18, 2017.
2. BaseSpace Clarity LIMS. www.illumina.com/products/by-type/informatics-products/basespace-clarity-lims.html. Accessed September 12, 2017.
3. Illumina (2017) BaseSpace Clarity LIMS Integration (support.illumina.com/content/dam/illumina-marketing/documents/products/technotes/basespace-clarity-lims-integrations-technical-note-970-2016-034.pdf).
4. Illumina (2017) Increasing User Adoption Rates and Realizing a Higher Rate of Return on a LIMS Investment. (www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/basespace-clarity-lims-adoption-rates-investment-970-2017-014.pdf).
5. Illumina (2016) Clarity LIMS, Intuitive Lab Management (www.illumina.com/content/dam/illumina-marketing/documents/products/brochures/brochure-clarity-lims-2016-web.pdf).