

Laboratory Information Management System for NGS Genomics Labs

Jitao Yang^{1,2}, Shuai Zhu², Guangpeng Xu² and Chao Han²

¹*School of Information Science, Beijing Language and Culture University, Beijing 100083, China*

²*Annoroad Gene Technology, Beijing 100176, China*

Keywords: LIMS, NGS, Genome Sequencing, Genomics, Sample Tracking.

Abstract: The goal of genome sequencing is to unravel the ordered sequence of nucleic acids that form the DNA or RNA of a given sample. Genome sequencing lab requires the ability to select and track a large amount of samples through many experimental steps. Therefore, laboratory information management system (LIMS) is needed to provide a way of automating the laboratory experimental procedures and track the samples. LIMSs have been proposed and developed for many years, but still remain difficult for labs to implement successfully. In this paper, we demonstrate our genomic next generation sequencing (NGS) LIMS solution. We developed a web-based LIMS with flexible configuration and customization for NGS laboratories, and can help laboratories track samples and optimize experimental procedures and business workflows. We also describe our solution of integrating LIMS with the existing enterprise business information systems. Finally, we share our experience for the implementation of a successful LIMS.

1 INTRODUCTION

Genome next generation sequencing is now commonly adopted, and has a broad areas of applications, such as the Non-invasive Prenatal DNA Testing, ctDNA Testing for Non-invasive Tumor Personalized Therapy, Plant and Animal Molecular Breeding, Genetics and Evolution, Microorganism and Ecological Environment, etc.

To manage tens of thousands of samples that are subject to NGS analysis, it is inevitable to develop adequate laboratory information management system to track and manage the NGS workflows. However, it is extremely difficult to manage the NGS experiment and analysis procedures accurately and efficiently, given the issues of multiple-source samples enrolling, national and international logistical management and tracking, fragmented procedures for assessment and processing of samples, the intricacies of molecular experimental steps, and the complex and multiple pipelines of NGS processing.

LIMS provides many benefits for the users of laboratory, several of the main benefits identified are outlined below:

- brings accuracy and accessibility to the flow of samples and data in laboratory,
- universally accessible data via the web rather than digging through files,

- years of data can be kept and queried conveniently,
- business efficiency improvement,
- data quality control,
- efficient sample tracking and management,
- automated and in-depth customer reports,
- integration with laboratory instruments,
- experimental steps quality control,
- building automated analysis pipelines,
- status and results sharing with collaborators and clinicians,
- financial management,
- access control,
- track and analyze trends,
- error reduction.

In this paper, we demonstrate our genomic NGS LIMS solution. Our LIMS is a web-based system with flexible configurations and customizations for NGS laboratories and can support the integration of multiple NGS instruments, it can help laboratories track samples and optimize experimental procedures and business workflows. We also give solution of integrating LIMS with the existing enterprise business information systems. Finally, we describe the most

likely causes for a failed LIMS and share our experience for the implementation of a successful LIMS.

The rest of the paper is structured as follows: Section 3 introduces the design of LIMS, while Section 4 describes the implementation of LIMS, and Section 5 presents the integration of LIMS with existing information systems. Section 6 discusses the implementation of a successful LIMS. Section 7 concludes the paper and outlines future work.

2 RELATED WORKS

Some commercial and open source LIMSs (Bath et al., 2011; PerkinElmer, 2016; Grimes and Ji, 2014; Progeny, 2016; Illumina, 2016) are available but typically require extensive modification and extension to address the specific needs of NGS genomics labs. We therefore developed a web-based LIMS, which is robust and flexible for managing the samples and the NGS processes.

3 SYSTEM DESIGN

As described in Figure 1, the functionalities of our LIMS can be grouped into the following categories:

- Enrollment of Sample Information. Enrollment of sample information is not a simple form filling process but constituted by two or even three steps, and supports multiple business models. For instance, salesman who stationed at hospitals will input the basic sample information (e.g., sample code, sample type, photo of sample sheet) into system by mobile applications embedded in Wechat. Then, when the sample along with the sample sheet are transported to company, typists in company will finish the full sample information (in sample sheet) enrollment. Finally, the correctness of sample information enrolled in LIMS will be checked by another team in company.
- Sample Logistics Information Tracking. Samples come from different cities distributed in the country, and will be transported to our head quarter to be tested. Each sample logistics information will be tracked by system. When samples are packed and sent to our company, package logistics code will be scanned into system. Then, the system can get the newest logistics information from express companies through their open APIs. We can check the package logistics information status during the whole transportation process. If something unusual (e.g., delayed, destroyed) happened in logistics process, the recipients and senders can response quickly to take steps to minimize the losses.
- Sample Assessment and Processing. Depends on sample attribute, sample sheet, sample number, test type or more factors, the samples arrived at company can be divided to different processing directions including rejection, resampling, storage, or flow to the experiment steps. Sample recipients, experiment operators and genomic analysts can choose the processing directions and operate in the system.
- Experimental Management. Experimental management is the most important and complicated part of LIMS. Based on sample type, test type, previous processing result and operator's subjective judgment, the samples will experience multiple experimental steps such as plasma isolation, nucleic acid extraction, molecular library construction, molecular library quality control, and etc. Our system is also flexible to support the configuration of different experimental processes.
- Sequencing of Samples. After the experimental steps, the samples are ready to be sequenced by NGS instrument (e.g., Illumina HiSeq X Ten). To minimize the cost of using sequencing reagents, generally, one NGS instrument operation will be required to sequencing as many samples as possible, therefore, samples belonging to different test business line will be combined together in the pooling process. After sequencing, the raw .bcl data will be separated into deferent genomic analysis business processes. Abnormal result data will be auto labeled by particular rules configured.
- Genomic Analysis. Before genomic analysis, LIMS will prepare the raw .bcl data for the filter and quality control software to be processed. The generated FASTQ data will be further prepared for the genomic analysis pipelines operated in the cluster computing environment. After genomic analysis, LIMS will retrieve the analysis result set to system so as to wait for the genetic interpretation.
- Genetic Interpretation. Sequencing and analysis results are presented in a neat and orderly manner in LIMS for the genetic interpretation scientists. Additionally, LIMS can provide relevant knowledge base for genetic interpretation such as gene-disease associations from several public data sources (e.g. AutDB (Mindspec, 2016), DisGeNET (Pinerio et al., 2015), OMIM (Hamosh et al., 2015)) and the literatures.

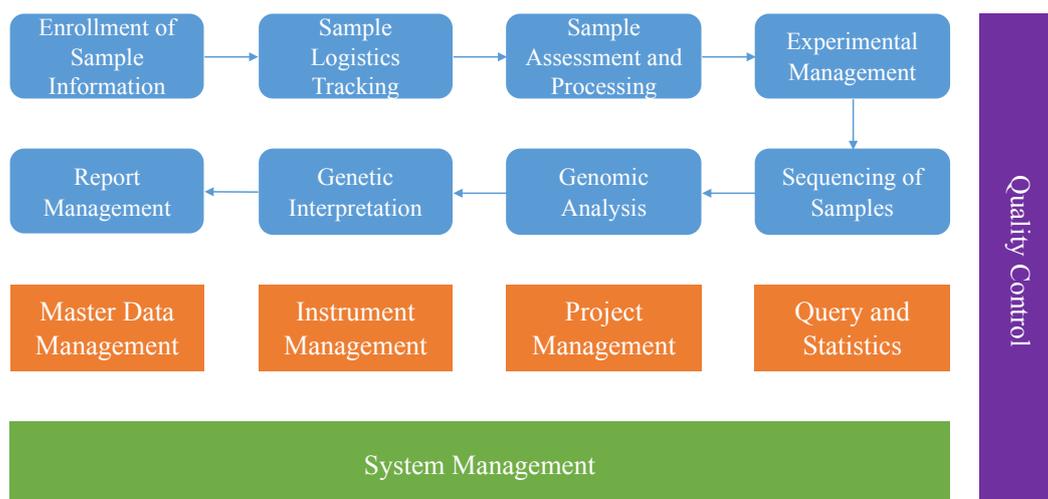


Figure 1: The Application Architecture of LIMS.

- **Report Management.** The genetic interpretation result will be classified by the report management module and transferred into different report generation pipelines of the reporting system.
- **Quality Control Management.** Quality control exists in the whole NGS work flows, including experiment steps quality control, reagents quality control, sequencing data quality control, genomic analysis quality control, genetic interpretation quality control, and etc. Different quality control processes can also be configured in system.
- **Master Data Management.** Master data management module includes tens of thousands of fundamental data for the operation of LIMS, such as the details data for experiment templates, barcode rules, agent information and so on. This module is usually maintained by department manager or system administrator, because master data include very important and sensitive data of the company.
- **Instrument Management.** Instrument management module not only records instrument property information like barcode, type, number, location, status, application methods and so on, but also associates instruments to a certain experiment step. An experiment step including its associated instruments could be exported as a detail guide which is very helpful for the operators.
- **Project Management.** Project management module provides the functions for project managers a convenient way to manage and monitor the projects, including project defining, project status monitoring, process intervening, and so on.
- **Query and Statistics.** Query and statistics mod-

ule provides powerful functions of advanced data query and visualized multidimensional data statistic and analysis. Instead of querying through millions of data items to find meaningful results in experimental workflows, researchers can quickly identify the information of their interests through this module.

- **System Management.** System management module provides the features of user management, role management, privilege management, department management, business management, organization management. A user's role and his privileges could configured flexibly.

4 SYSTEM IMPLEMENTATION

Our LIMS is written in Java using the Spring (Pivotal, 2016), SpringMVC, Mybatis (Goodin et al., 2016) web application framework and implementation is platform-independent. Our web servers run Linux/CentOS and we use the MySQL relational database management system. The system architecture of LIMS is described in Figure 2.

The web interface is designed to handle a variety of functions in a modular format. The left column includes the function category of business process. The top navigation bar provides advanced management and analysis functions.

We have more than one hundred data tables in the system, Figure 3 shows part of our database schema for data tables.

HTTPS (Hypertext Transfer Protocol Secure) is supported and is implemented for the system. User's access to functionality is controlled via user roles

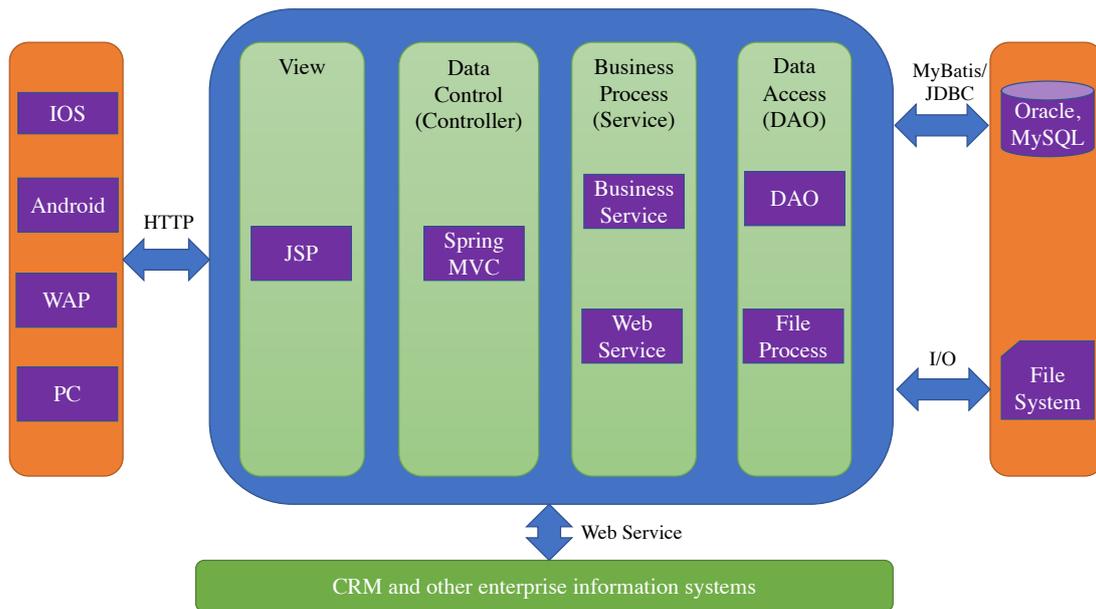


Figure 2: The System Architecture of LIMS.

which are defined and managed by the system management module.

Our LIMS currently supports the following Illumina NGS sequencing platforms: HiSeq X Ten, MiSeq, HiSeq, HiSeq 4000, HiSeq2500, NextSeq 500/550AR, and could be configured for other type of NGS instruments.

5 INTEGRATION WITH EXISTING SYSTEMS

Implementation of a LIMS requires a good degree of integration with the existing business information systems in enterprise. We integrate LIMS to our two existing business system platforms as demonstrated in Figure 4 and Figure 5.

5.1 Integrating with Company Systems

Figure 4 describes the business process in our company.

Customer Relationship Management (CRM) system serves primarily our clinical and health services, including Personal Genome Test, NIPT (Non-Invasive Prenatal DNA Testing), Cancer Gene Therapy, and etc. CRM supports sample information collection, hospital information management, agent management, product management, sales management, sample management, doctor management, financial management, customer reports management,

and etc.

Project Management (PM) system serves primarily our technical services, including the research area of genomics, transcriptomics, epigenetics, and etc. PM includes the functions of sample information enrollment, contract management, project schedule management, project establishment, job order management, quality control, research achievement management, and etc.

Logistics system provides the management functions of express package, express company, alert time, receipt time, samples, logistics tracking, and etc.

All the information related to NGS experiments and analysis will be synchronized into LIMS, therefore, the sample information could be enrolled through:

- CRM web portal, or
- CRM mobile app, or
- PM web portal, or
- PM mobile app, or
- LIMS directly.

The sequencing result data will be stored in Network Attached Storage (NAS) system, and then filtered and classified to be analyzed by hundreds of different genomic analysis pipelines.

The genomic analysis result data will then flow to the reporting system, which can produce hundreds of different areas of professional reports automatically.

The reporting system will finally distribute the user reports to CRM and PM, then the user can receive

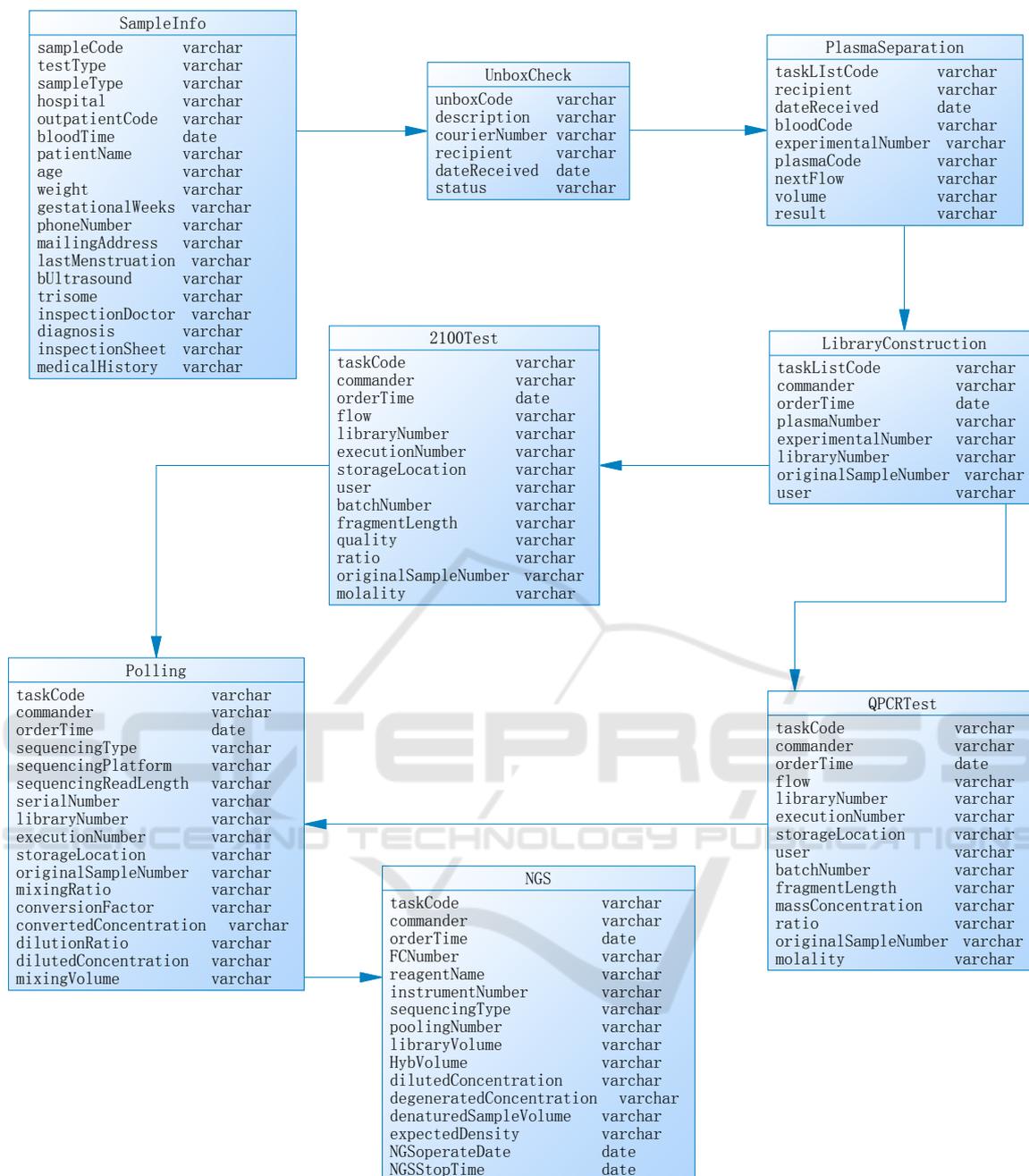


Figure 3: Part of Database Schema for LIMS.

their reports through our mobile app or web portal.

We have also constructed a hybrid cloud computing platform SolarGenomics (SolarGenomics, 2016), because our business can produce around 10 TB sequencing data everyday, both the local computing cluster and the local distributed storage can not provide computing and storage resources sufficiently and elastically.

5.2 Integrating with Hospital Systems

Figure 5 describes the business process in our collaborating hospitals.

We collaborate with more than 2000 hospitals. In some of the hospitals, we set up NGS laboratory and install our NGS instrument there so that the hospitals will have the abilities of NGS experiments, sequenc-

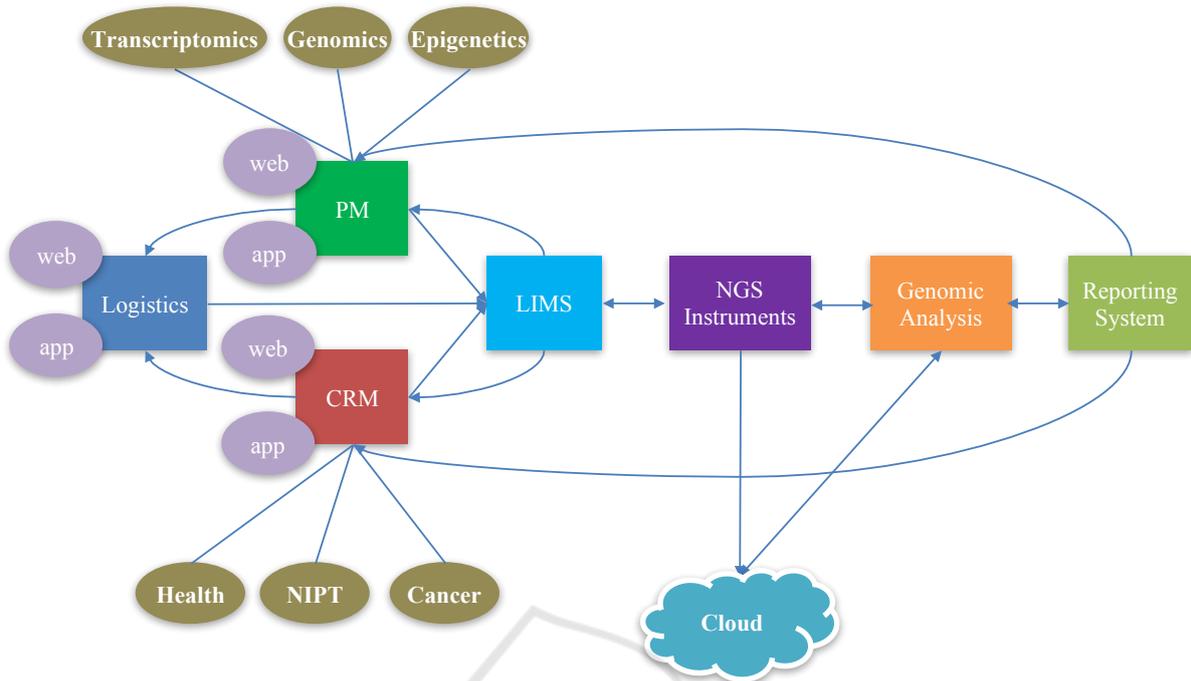


Figure 4: The Business Process System Platform in Enterprise.

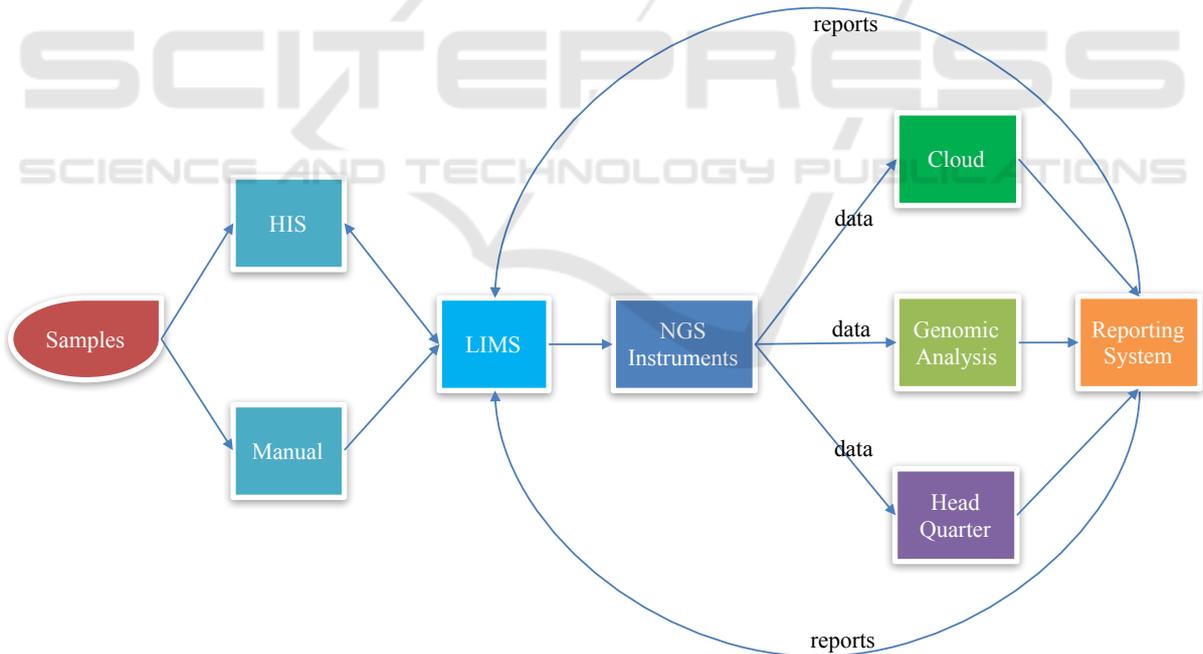


Figure 5: The Business Process System Platform in Hospital.

ing, genomic analysis, and genetic interpretation.

The samples in our collaborating hospitals could be enrolled to LIMS either manually or synchronized from the hospital information system (HIS). The samples will be processed by multiple experimental steps

in the laboratory of hospital and then be sequenced by the NGS instrument placed in the hospital.

Further, the sequencing data will be:

- analyzed by our genomic analysis pipelines installed in the local hospital computing servers, or

- uploaded to our SolarGenomics cloud computing platform then analyzed by our genomic analysis pipelines on cloud, or
- transferred to our head quarter then analyzed by our bioinformatics scientists.

Finally, our reporting system will give professional reports to the clinical doctors or patients.

6 A SUCCESSFUL LIMS IMPLEMENTATION

A large number of LIMS implementations failed to meet the user's initial expectations, and this can be due to the lack of proficient user requirements specifications, the frequent requirements changes, and the technology-based shortcomings.

The absence of adequate requirements is the biggest reason why a LIMS may fail, as generally the software developers are seldom have sufficient knowledge of experiments and NGS, the success of a LIMS relies on the deep understanding of laboratory and genomic analysis business needs.

Frequent requirements changes is another reason causing the fail of LIMS, as the laboratory staffs are normally focused on a specific area of experiment, it is hard for them to identify clearly how a LIMS is going to fit into their laboratory situation and to propose a general function architecture of LIMS. Additionally, the lab staffs lack the IT knowledge and skills, they will think the modification of system functions as a simple task, therefore, the requirements will be modified frequently. This will postpone the deliver of project and even fail to publish LIMS.

Technology-based shortcomings should be emphasized, as the information technology department is often not the first class citizen in gene technology organizations.

Implementation of a LIMS project will also require a good degree of integration with the existing enterprise business information systems, data exchange mode and standard must be adequately addressed.

Mapping out clearly the requirements specifications, controlling the requirements changes effectively, addressing the problem of integrating LIMS into existing enterprise systems, relying on a strong IT team will stand a quite high chance of implementing LIMS successfully.

7 CONCLUSIONS

To meet the needs of managing tens of thousands of samples for genome sequencing, we developed a web-based laboratory information management system that is flexible to be configured to adapt to next generation sequencing technologies. A LIMS system is critical to the accurate and effective management of sample information, experimental data, genome sequencing data, and the reproducible analysis results. Our LIMS addresses all of these needs and seamless integrates with the existing systems in enterprise and in collaborating hospitals.

In our LIMS, all the data are stored into distributed authoritative repositories, samples are traceable from the enrollment, transportation, experiment, a sequencing run, quality control, genomic analysis, genetic interpretation, generation of report, to sample storage, and all the other processing steps in between. In conjunction with a sample identifier (QR code) encoding rules and advanced query and analysis capabilities, LIMS can quickly identify sample and significantly reduce errors in the whole steps of business process. Our LIMS provides a comprehensive and efficient management solution for NGS genomics labs.

LIMS is very complicated and difficult to implement, especially in the NGS research laboratory, we therefore share our experience of implementing a successful LIMS.

Regarding further work, it is promising in our schedule that we will provide a cloud-based LIMS solution in our SolarGenomics genome sequencing big data cloud platform, and open the service for more genomics labs.

ACKNOWLEDGEMENTS

This work was partially supported by the Science Foundation of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (15YJ030001, 14YBB12, 16YJ030001).

REFERENCES

- Bath, T. G., Bozdog, S., Afzal, V., and Crowther, D. (2011). Limsportal and Bonsailims: development of a lab information management system for translational medicine. In *Source Code Biol Med*.
- Goodin, B., Poitras, C., and Begin, C. (2016). Mybatis. In <http://blog.mybatis.org/>.

- Grimes, S. M. and Ji, H. P. (2014). Mendelims: a web-based laboratory information management system for clinical genome sequencing. In *BMC Bioinformatics*.
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., and McKusick, V. A. (2015). Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. In *Nucleic Acids Research*.
- Illumina (2016). Basespace Clarity LIMS. In <http://www.illumina.com/informatics/research/sequencing-data-analysis-management/genomics-lims.html>.
- Mindspec (2016). Autdb: a Genetic Database for Autism Spectrum Disorders. In <http://www.mindspec.org/products/autdb/>.
- PerkinElmer (2016). Genesifter Lab Edition. In <http://www.geospiza.com/Products/LabEdition.shtml>.
- Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., and Furlong, L. I. (2015). Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. In *Database (Oxford)*.
- Pivotal (2016). Spring framework. In <http://projects.spring.io/spring-framework/>.
- Progeny (2016). Progeny LIMS. In <http://www.progenygenetics.com/lims/>.
- SolarGenomics (2016). Solargenomics genome sequencing big data platform. In www.solargenomics.com.

