Databases and ontologies

A Laboratory Information Management System (LIMS) for a high throughput genetic platform aimed at candidate gene mutation screening

C. Voegele1,*, S.V. Tavtigian1, D. de Silva1, S. Cuber1, A. Thomas2 and F. Le Calvez-Kelm1
1International Agency for Research on Cancer (IARC), Lyon, France and 2Department of Medical Informatics, University of Utah, Salt Lake City, USA

Received on April 16, 2007; revised on June 20, 2007; accepted on July 6, 2007

ABSTRACT

Summary: High throughput mutation screening in an automated environment generates large data sets that have to be organized and stored reliably. Complex multistep workflows require strict process management and careful data tracking. We have developed a Laboratory Information Management Systems (LIMS) tailored to high throughput candidate gene mutation scanning and resequencing that respects these requirements. Designed with a client/server architecture, our system is platform independent and based on open-source tools from the database to the web application development strategy. Flexible, expandable and secure, the LIMS, by communicating with most of the laboratory instruments and robots, tracks samples and laboratory information, capturing data at every step of our automated mutation screening workflow. An important feature of our LIMS is that it enables tracking of information through a laboratory workflow where the process at one step is contingent on results from a previous step.

Availability: Script for MySQL database table creation and source code of the whole JSP application are freely available on our website: http://www-gcs.iarc.fr/lims/.

Contact: voegele@iarc.fr

Supplementary information: System server configuration, database structure and additional details on the LIMS and the mutation screening workflow are available on our website: http://www-gcs.iarc.fr/lims/

1 LIMS ARCHITECTURE

1.1 The LIMS components (Fig. 1)

The LIMS is hosted on two secured Linux servers running a Debian Woody Operating System. One server is dedicated to the MySQL relational database and the other to the web application. The latter enables interactions with the laboratory’s instruments, robots and analysis software. The whole system is integrated into an internal network and secured within a strict firewall.

The MySQL database embodies a complex data model that integrates different types of data from sample features to result reports including all plates and reagents that are used along the mutation screening process. The database schema mirrors the workflow and is sufficiently flexible to allow evolution in workflow subprocesses. The content is secured by a combination of different kinds of daily and incremental backups.

1.2 User interface

The LIMS presents a friendly and intuitive user interface; users navigate within the application in the same manner as they are accustomed to navigate web sites, using menus that follow the laboratory workflows. A series of ~250 Java Server Pages (JSPs) enable database management by linking the user interface to the database. The interface includes one or more display screens that are specific for each step in our process and in which the related data are presented in tables that list the queried database content. Each laboratory step has a corresponding LIMS transaction (mostly plate barcode transactions) that triggers addition(s) or update(s) of single or batches of database records. All of these functions are managed through ‘restrictive’ forms that use pull-down menus, whenever possible, to minimize form-filling errors. Specific JavaScript form validations check the data type and integrity with respect to other tables in the database and guide the users through the workflows by prompting them to fill the required fields of one step before moving on to the next step. Plate barcodes are entered using barcode scanners to avoid miswriting, and each database transaction is stored in the database with the user’s name and the date.

Information retrieval is managed through two types of queries. The first type focuses on the content of individual database tables. The second type, much more complex, includes relationships between the tables to retrieve the current content and process history of plates as well as the final experimental results. These queries generate multi-entry tables that are easily converted to tab delimited text for display or analysis with other software.

This web-based approach results in a zero-installation for the users and enables access from any computer within the LAN. The system supports unlimited simultaneous connections;

*To whom correspondence should be addressed.
operates equally well whether the individual user is working in a Windows, Mac OS or UNIX environment, and is compatible with most popular web browsers.

1.3 Creating and deploying new modules

The LIMS is integrated into an Ultimate Bulletin Board (UBB), which is a customized application based on version 6.02 of the UBB from Infopop, Inc. The application contains a framework for database table creation and standard basic web page generation. Each database table generally corresponds to one entity or one step in the workflow and has at least seven associated standard files for listing, adding, updating and deleting entries. In addition, a large number of pages have been created for specific functionalities and for more sophisticated modules such as plate pictures, results requests or scripts for sending files to printers and robots.

User authorization access has been implemented for different levels of functionality through the UBB. LIMS users are identified with personal logins and passwords stored encrypted in the database together with the authorization codes for all activities. Thus, an integrated and safe environment is maintained for the whole application.

2 IMPLEMENTATION

To detect new rare sequence variants, our mutation screening strategy relies on a combination of High Resolution Melt curve analysis (HRM) and resequencing (Chou et al., 2005; Margraf et al., 2006; Tavtigian et al., 1997). All laboratory steps of the process are mirrored in and managed through the LIMS.

Over the next few paragraphs, we outline the workflow displayed in Figure 2. The description of step 4 is expanded to highlight LIMS management of the contingent process occurring at this step. After primary (1) and secondary (2) DNA amplifications, the products are consolidated to a 384-well HRM plate (3). The HRM plate is then queued for formation of hetero-duplexes and collection of HRM profiles on a barcode reader equipped Idaho Technologies 384-position LightScanner.

(4) The HRM curve results are imported in the LIMS database, and based on these results, a subset of samples, including all of those with HRM profiles suggestive of the presence of a sequence variant, are selected for sequencing. In the LIMS, the barcodes for 1–3 384-well HRM plates are specified for re-arraying and a new 96-well sequencing plate is created. The specification process launches the display of the content of the HRM plates color-coded by HRM result into 3 groups:

- Group 1: samples that must be resequenced because their HRM curve was indicative of the presence of a sequence variant—these are already pre-selected
- Group 2: samples that can be resequenced because their curves were at the edge of the distribution of apparently homozygous wild-type samples—(Group 3) samples that are generally not resequenced except to provide a few wild-type controls because their curves were clearly indicative of homozygous wild-type.

In principle, the samples that will be selected for resequencing (all of Group 1 and some of Groups 2 and 3) are randomly distributed across the source HRM plates. Selection by the user of up to 96 samples launches a JSP function that: (a) creates the transfer pattern for re-arraying and consolidating the selected samples from the HRM plates into a 96-well plate (called EXO/SAP because PCR products are subsequently digested with exonuclease I and shrimp alkaline phosphatase). The rearrangement is done automatically in a specific order so that samples derived from an individual amplicon are transferred to consecutive positions on the plate. (b) Inserts the new plate’s features and well content in the database. The identities of each sample queued at this step are thus contingent on results from the mutation scanning, their process history being recorded in the LIMS. (c) Sends a worksheet to the Qiagen BioRobot 9600 to launch the re-arraying program as specified by the LIMS.

PCR samples are then purified and sequenced in both senses (5). The 4 individual (A, C, G, T) dye-primer sequencing reactions are consolidated before being cleaned up and injected on to a 96-capillary electrophoresis sequencer (6). The run produces a pair of forward and reverse SCF format chromatograms for each sample. These files are deposited in the
database linked through their process history to the samples from which they originated and the amplicon for which they were generated. The pairs of forward and reverse chromatograms are then ready for sequence analysis (7).

3 COMMUNICATIONS WITH INSTRUMENTS

The LIMS interacts with most of our laboratory robots and instruments, all of which are run by computers integrated into the internal network. Communications are bidirectional and are mediated by sending and receiving text files. The transfer is ensured by a JSP tag that launches a general loading function, sending specific files to specified computers by calling a Perl script. Below, we describe the interactions with the barcode printer, the LightScanner and the BioRobot. Although not described here, worklists specifying the positions and barcodes of plates to be processed are also sent to the Tecan and Beckman robots and to the sequencer.

Correct plate tracking and therefore barcode printing is fundamental to our workflow. Each creation of plates triggers the creation and sending of a text file to a specific directory on the computer that manages the barcode printer, which launches instantaneous printing. The LIMS generates the barcodes by incrementing the highest plate id of the same type stored in the database, thereby guaranteeing barcode uniqueness. The files sent to the barcode printer contain these barcodes to be printed, encoded in the barcode printer language EPL2.

Communication with the LightScanner is another key element because downstream process decisions are contingent on HRM results. Sample sheets containing the barcode id of the HRM plate to be loaded in the LightScanner, as well as the positions, ids and amplicons of every sample on that plate, are sent to the computer that runs the instrument. Users associate each HRM plate with its sample sheet by barcode. After analysis, the mutation screening results are exported into a new specific sample sheet named with the HRM plate barcode. In the LIMS, the contents of these sample sheets are parsed to transfer each datum to the appropriate tables and fields of the database.

Based on the HRM results, samples are selected for sequencing. A worksheet containing the (1–3) source plate barcode(s) and the one destination plate barcode is sent to the Qiagen BioRobot 9600. This worksheet also contains the source and destination well coordinates that specify the required sample transfers.

4 CONCLUSION

LIMS has now become essential to our laboratory activities. Its use in combination with automation of laboratory processes has improved the efficiency and quality of the work by reducing potential for human errors, accelerating the throughput of analysis, and enabling sample tracking activities that are very difficult to perform without error by hand. The first two projects relying on this system in its entirety have been complete mutation scanning of the promoter, exons and introns of TP53 in a series of 50 subjects and mutation scanning of all of the coding exons of CHEK2 in a series of 1300 subjects. These studies resulted in the generation of ~20 000 chromatograms; because the mutation screening is equivalent to genotyping every base pair screened, the analyzed results correspond to 4.2 million genotypes (Garritano et al., and Le Calvez et al., manuscripts in preparation). We routinely maintain throughput of approximately 50 PCR plates per week.

While several commercial software packages now also provide analogous web-based architecture, phases of specification to adjust these packages to specific laboratory needs are still laborious and suffer limitations. Our LIMS programmed de novo can be flexible and expansible, providing options for continuous improvement and providing a template for the development of new modules governing new workflows.

Conflict of Interest: Dr. De Silva is currently employed by Idaho technology Inc., and holds stock in the company.

REFERENCES

