"Laboratory Assistant Suite

Manual

Version: 1.0
Date: 18/03/2015

Index

Index of figures

Index of tables

1. Introduction

The Laboratory Assistant Suite (LAS) platform assists the researchers in different laboratory activities. Its modular architecture allows managing different kinds of raw data (e.g., biological, molecular) and tracking experimental data. Each LAS module is tailored to handle specific activities or data types, but it is plugged into a broader and uniform framework, allowing effortless integration with the other components of the system. In addition, the data models and procedures integrated in the platform try to comply with best practices and standards widely adopted by the research community at large. User interfaces are designed to be practical in hostile environments, in which researchers should minimize their interactions with the system during data entry procedures (e.g., in sterile conditions). Furthermore, the platform supports the integration of different resources and aids in performing a variety of analyses in order to extract knowledge related to tumors. The LAS platform is the result of a joint effort by both IT and biomedical researchers of the Candiolo Cancer Institute.

Since the laboratory-related procedures can be categorized into different layers according to data complexity and purpose, the LAS architecture has been modeled according to the same rationale. Thus, it has been extensively based on a three-tier design pattern, both at the system-wide and the software module levels. This is a well-established architectural paradigm in software engineering, which targets flexibility and reusability by breaking up an application into tiers. Each tier addresses a specific issue and interacts with the other tiers by means of well-defined interfaces. We modeled the platform in the following tiers: (i) operative, (ii) integration, and (iii) analysis. In addition, a cross-tier software component regulates accesses to the system and enforces user privilege control for all LAS services.

Within the LAS general architecture, each tier includes a set of fully-fledged applications, or modules. While the lower tier is mainly concerned with the collection of experimental data, the modules and data managed by the upper tiers are characterized by an increasing level of abstraction. Lower tiers can serve requests generated by the upper tiers and provide the data needed to carry out complex tasks (e.g., data integration and/or analysis).

The operative tier is responsible for collecting, storing, and tracking raw experimental data. These include data from several sources, such as **tissue collection** and **biobanking**, **molecular experiments** (e.g., sequencing, microarray), **in vivo and in vitro experiments** (i.e., xenografts, cell lines) management, each handled by a specific software module. Modules in this tier are meant to work in close interaction with the researchers in a laboratory setting. Thus, **graphical user interfaces** (GUIs) are explicitly tailored to ease data entry operations and assist the researchers throughout their experiments. The interaction is purposely designed to be lean, with the aid of special input devices, such as touch-screen notepads and barcode readers.

The integration tier is aimed at integrating different types of raw experimental data by means of **complex queries**. Ad-hoc identifiers have been adopted throughout the databases, which allow interlinking different biological entities in a unique network. Integrated data can be browsed or visualized as graphs. In addition, they can be exploited by the analysis tier and enriched by means

of annotations. For instance, a population of samples can be annotated as responsive to a given drug according to statistical analyses or tagged as bearing a genetic alteration based on sequencing data.

The analysis tier (currently a prototype) is designed to define **workflows for the analysis** of integrated data. The main idea is to provide a tool for designing complex analyses by means of a graphical representation. The analysis process will ultimately generate annotations and may optionally export data for visualization with external tools. Finally, predefined analysis flows could be exploited by operative modules to perform analyses on data collected by the user during an experiment.

The access and privilege control system manages all user access to the software modules in each tier, according to their profile. The user profile is defined during user registration and can be updated as needed; it lists the LAS modules accessible by the user, together with the set of functionalities he/she is allowed to use in each module. Moreover, some users with special privileges can create groups of users, based on particular needs (e.g., research studies and/or laboratory activities carried out by a specific group of people). This system also provides a finer-grained control over the data by defining and enforcing user and/or group access privileges with a row-level granularity, in order to guarantee different security levels for confidential information.

In the following the main functionalities of each module currently included in the LAS platform are described.

1.1. System requirements

To ensure full compatibility, we recommend upgrading your browser to the latest version of Google Chrome (https://www.google.com/chrome/).

All operating systems (Windows, Linux, iOS, Android) are supported.

2. Basic concepts

To properly manage all data and experimental procedures, the LAS identifies several types of entities and defines a predefined set of possible relationships among them.

The entity types are the following:

- **Patient.** This entity is a person that signed an Informed Consent to allow the collection of specimens for research purposes.
- **IC.** It represents the signed Informed Consent for a Study. Without a signed IC, no specimens or information whatsoever about the patient may be collected.
- **Study.** The study can be for instance a clinical trial or a research study approved by one or more institutions. Each study has a set of associated rules that define the constraints for collecting biological samples and performing research experiments.
- **Collection.** It represents the collection of several specimens sharing homogeneous characteristics or a common source event. For instance, a collection can pertain to all of the tissues collected from one patient following a surgical intervention. Different surgical interventions on the same patient are usually mapped onto separate collections.
- **Aliquot.** The collected specimens are split into aliquots according to their intrinsic characteristics. The system is capable of managing several kinds of aliquots (e.g. viable, RNALater, SNAP Frozen). According to the characteristics of the aliquots and the purpose of the research study, different operations can be performed: (i) storage in dedicated containers, (ii) extraction of derived aliquots (e.g. DNA, RNA), and (iii) implantation in immunocompromised animals (i.e., xenografts) or (iv) derivation of cell lines.
- **Biomouse.** It represents the biological complex generated by the implant of an aliquot into a given animal. If different aliquots are implanted in the same physical mouse in different sites, different biomice are generated.
- **Cell Line.** This entity represents in vitro experiments. Cell lines can be generated from viable aliquots by means of a generation and thawing process, or from other cell lines by means of expansion procedures.

Each collection and biological entity (i.e., Aliquot, Biomouse, Cell Line) is identified by a unique and mnemonic key, named GenealogyID, that encodes relevant information regarding the history of the entity. This key is automatically generated by the LAS platform through formal rules. In Figure 1 the structure of the GenalogyID is reported. The first part of the GeneaologyID (up to the instance ID field) is focused on summarizing information about the ancestors of the bioentity, while the last part is used to describe some specific features of the current bioentity (e.g., aliquot type for Aliquots, implant site for Biomice).
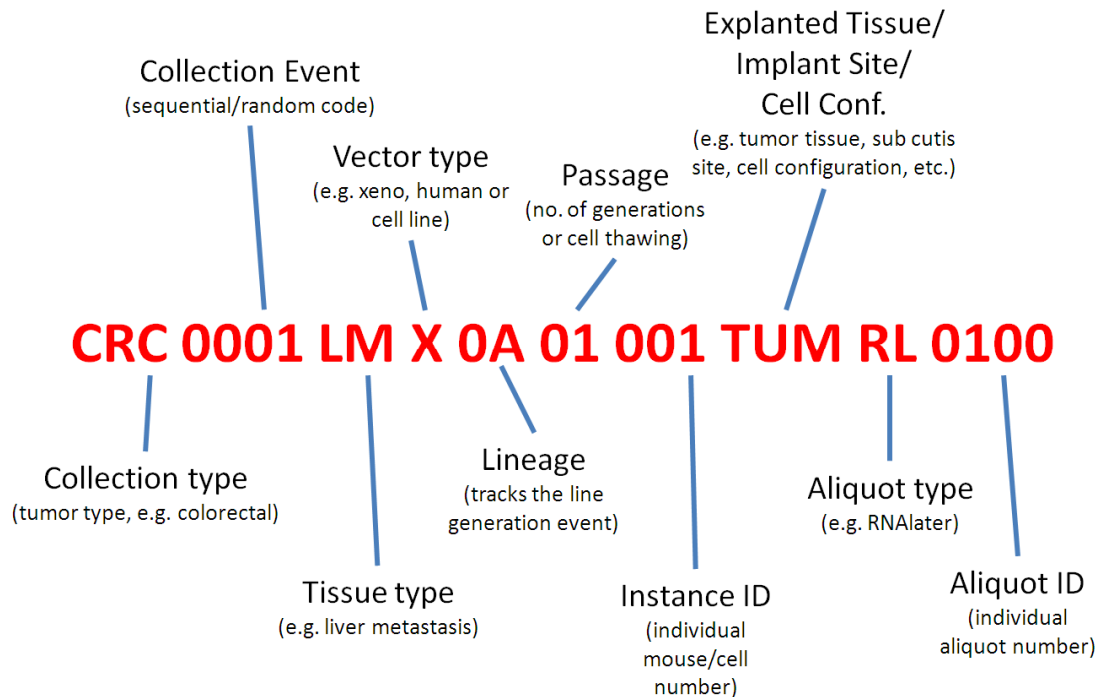
Figure 1 - Genealogy ID structure

In Figure 2 we report a graph representation of the relationships among the LAS entities. It may be noticed that one patient can generate one or more collections by means of the same Informed Consent, or through different ICs related to different studies. Each collection can be composed of different aliquot types and several instances of the same aliquot type. If a viable specimen is implanted into animals, it generates one or more biomice. The biomice of the first passage are also labeled with a different lineage identifier. Each biomouse can generate aliquots that can be implanted into other animals, so as to generate biomice of the second passage. The implantation/explant process can be repeated several times according to the purpose of the research study. Viable aliquots can also be used to generates Cell Lines that can be expanded and/or archived to produce other generations.

All such relationships are stored in a graph database, while detailed information about each entity and the description of the procedures applied are collected in a relational database. A non-relational database is instead exploited to store complex data and files generated by laboratory instruments.
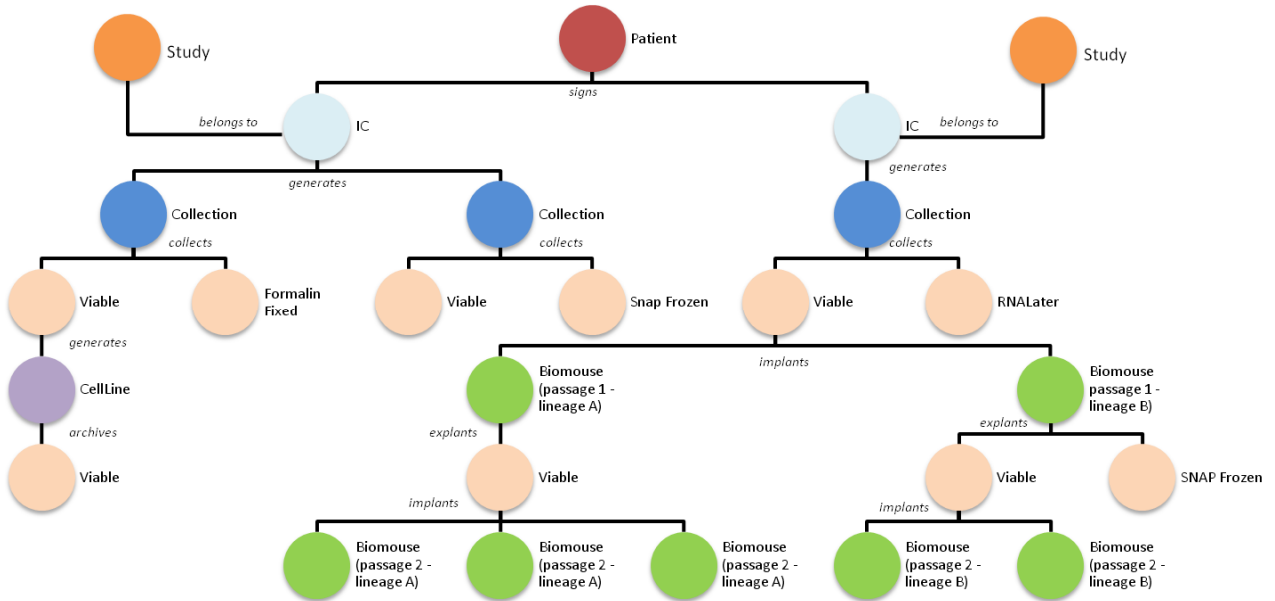
Figure 2 - Graph representation (genealogy tree) of the relationships among LAS entities. The nodes belonging to the same macro-category are represented with the same color. The edges are labeled according to the procedure exploited to generate the children.

3. Authentication

The management of data produced by different people and/or groups requires that access to functionalities and information be limited according to several aspects such as group and/or project membership, and user role. For these reasons, we developed a cross-tier module, named the Authentication Manager, that manages users and their privileges. In the following the main functionalities are described.

Basic concepts
The LAS platform is based on the following concepts.

Working group. A Working Group (WG) is a set of users in the LAS platform that work together towards a specific goal (e.g., project, research activity). The data produced by the users of the same group are private, unless they are intentionally shared with other groups.

User profiles. The users belonging to a WG can be divided into two main categories:

- Principal Investigator (PI): he/she is the manager of one or more WGs. PIs can add or remove users from their WGs, and grant or revoke permission to access functionalities for every user in their WGs. During the registration process, PIs can select one or more activities they wish to handle within the LAS. Their registration will be evaluated by the system administrator.
- LASUser: he/she is a standard user, whose activities depend on the choices of the PI. During the registration process, LASUsers should select the PI they report to. They are also asked to specify one or more roles, according to the task(s) they need to carry out, so as to help the PI in assigning appropriate functionalities during the registration acceptance phase.

The following roles are currently available:
- Biobank Manager
- Collection Manager
- Molecular Biology Technician
- Animal Facility Technician
- Post Doc*
- Research Supervisor*
- PhD Student/Research Assistant
- MicroArray Facility Manager/Technician
- MicroArray Facility User
- Lab Manager/ PI Personal Assistant*

Users of both categories (i.e., PI and LASUser) can only access the LAS pages that correspond to the functionalities they have been enabled to use, and they can only manage data that belong to their WGs. Only roles denoted with a * in the above list are allowed to share data with another WG. Data

sharing can be performed either through a Transfer operation, or directly through the Query Module when the query results include BioEntities such as Aliquots, Biomice or Cell Lines.

The current available activities are listed in Table 1.

Table 1- LAS activities

Manage Samples	Derive
	Experiments
	Kits and Protocols
	Perform QC/QA
	Plan Samples Management
	Retrieve Samples
	Slide Preparation
	Split
	Transfer
Molecular Analysis	Beaming
	DigitalPCR
	IHC
	MicroArray
	NGS
	Real-Time PCR
	Sanger Sequencing
	Sequenom
	Western blots
In Vivo Experimentation	Explants
	Implants
	Manage Mice
	Measures
	Review Experiments
	Treatments
	Xeno Admin
	Xeno Batch
In Vitro Experimentation	Archive
	Cell line generator
	Expansion
	Protocol Manager
	Thawing
Biobanking	Biobank Admin
	Collection
	Patients
Storage	Manage Archive
	Manage Container
	Return Used Samples

Data Querying and Manipulation	Data Sharing
	Querying
	Querying Admin
PI Activities	WG Management
External Projects	Mercuric
Analysis	Randomizer
	Write Formulas

Registration

People who want to gain access to the LAS platform should send a subscription request by filling in the registration form with their personal details (e.g., affiliation, role), to allow the administrators to correctly process the registration.

Home

The authentication module provides access to all LAS modules. On the home page, the LAS modules are grouped into categories. The menu structure is as follows:

- Manage Samples & Archive
    - Biobank
    - Repository
- Biological Experiments
    - Cell Lines
    - Xenografts
- Molecular Experiments
    - Digital PCR
    - Beaming
    - MicroArray
    - Next Generation Sequencing
    - Real-Time PCR
    - Sanger Sequencing
    - Sequenom
- Query & Analysis
    - Query
    - Analysis
- Profiles & Tutorial
    - Working Groups
    - Profile
    - Videos

Each menu item will only be visible if the user has been granted access to the corresponding module and/or functionality.

Management of working groups

A working group is defined as a group of people that work together under some specific rules. For instance, working groups can be matched to research/project groups within a given institution. The

leaders of a working group, also named the principal investigators, can restrict access to some functionalities and define sharing rules on the data. Data of a working group cannot be seen by other groups until an authorization has been approved by the principal investigators. The user privileges determine which functionalities are available to each user, according to their role and the restrictions enforced by the principal investigators of their working group.

Profile

The user can manage his/her account and personalize it. The following actions can be performed:

- *Manage Your Circles.* The LAS platform allows defining Circles. Circles represent an easy way for users to organize the members of their working groups. They allow the user to share the notifications of some experimental activities with the people to whom they are relevant. Once a circle has been created, it will appear in the Manage notifications page.
- *Manage Notifications.* For each functionality that produces a report, the user may define one or more recipients, who will receive a copy of the report by e-mail. This operation must be repeated for each working group on behalf of which the user can execute the functionality. For each functionality, the following predefined recipients may be selected:
  - o Executor: the operator that executes the functionality (i.e., the user him/herself)
  - o Planner: the operator that executes the previous operation (e.g. if the current user is in charge of the derivation operation, the Planner is the user that plans the derivation session)
  - o Assignee: the operator that is in charge of executing the next step of the workflow
  - o Recipient: the operator that will complete the operation, if there are two or more following steps.

  In addition, results can be shared with a circle of LAS users. The user can also freely enter other e-mail addresses, in case he/she wants to share his/her results with more people.
- *Request functionalities.* For each working group, the user can request new functionalities from the Principal Investigator to perform new experimental procedures. The PI may accept or decline the request.
- *Manage Working Groups Invitations.* In this page the list of received invitations to join other WG is shown. The invitation can only be sent by the Principal Investigator and/or the PI's Personal Assistant.
- *Account.* It shows the account details and allows the user to change his/her password.

Video tutorial

Video tutorials illustrating a subset of key LAS functionalities have been made available under the menu item "Videos", in the category "Profiles & Tutorial". The videos are further categorized according to the activities.

Session expiration

The work session in each module expires after 4 hours of inactivity.

4. Biobank

With the rapid advances in biomedical and genetic technologies, collections of biological materials have attracted increasing attention from the research community since they represent a fundamental resource for a deeper understanding of different pathologies, and the study of possible therapeutic applications. Such collections, named biobanks, are commonly divided into tissue and genetic biobanks according to the types of biological materials they store. The BioBanking Management Module addresses both issues.

The scope of this module spans a wide range of activities, including management of biological samples and associated pathological information, as well as support to a number of laboratory-related procedures. In particular, the module can currently handle the following activities and the corresponding data.

Collections

Fresh tissues

The collection of new samples coming from a surgical intervention is the primary source of the biobank. To enable the employment of barcode readers and thus streamline data entry procedures, most pieces of experimental equipment (i.e., clinical folders, plates, tubes) are bar-coded by the institution prior to their use. Sample collection consists of two phases. In the first phase, the researcher inputs preliminary information about the collection event and selects one or more sample types. For instance, it is possible to insert an identifier for the patient. Besides, the "Add clinical parameters" functionality allows the management of several kinds of clinical data, categorized according to an internal ontology. Some of these parameters depend on the selected "Tumor type". For example, "Histological classification" in the "Pathological features" menu is a drop-down list whose values are shown according to the selected "Tumor type".

In the next phase, the researcher can add one or more samples for each of the sample types chosen. Figure 3 depicts the interface for the collection of biological samples, which is designed to closely match the physical working environment of the researcher. The collection plates are organized on screen into different tabs, so that types of aliquots that are usually collected together appear in the same tab. For instance, Figure 3 reports the panel containing the containers for viable, ffpe, RNA later, and snap-frozen aliquots that are usually collected for in-vivo and in-vitro experiments.

The rectangular boxes represent the physical working plates in which sample aliquots are stored. First, the user must choose an option from the "Tube/Plate" selector. Next, he/she can insert the corresponding container barcode. Plates must already be present in the system; for that purpose they must be inserted in LAS through the special interface of the Repository module. Tubes need not be present in the system prior to the collection event. If needed, they can be created together with the aliquot stored in it. For more information about containers, please refer to the description of the Repository module.

Inserting an aliquot in a given plate position is done by simply clicking on the corresponding cell, which can be comfortably achieved on a touch-screen device without resorting to a mouse. This results in a reduced number of lean interactions with the system, allowing the researcher to focus on the experimental task at hand, rather than on data entry operations.
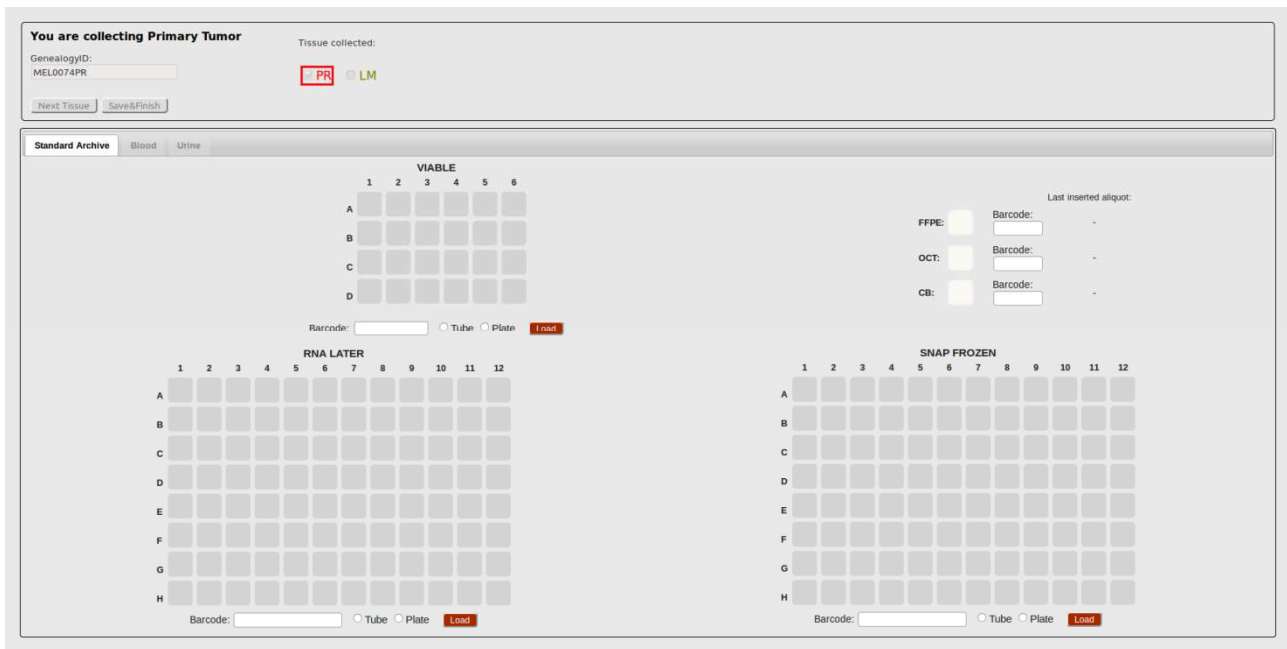
Figure 3- Collection interface

Archive samples

This procedure manages the collection of samples already archived by other researchers, possibly in another hospital. Maybe as a result of a collaboration between research centers, an operator might have to insert some external samples in the storage. In this case, the samples to be archived are not coming directly from a surgical intervention and are not fresh, but they have already been archived in some container. This operation can be done through the "Archive samples" interfaces. The user must select whether he/she is creating a new collection or adding these new samples to an existing one. The first option is useful if new samples come from a new patient whose aliquots were never collected, or if the user wants to independently process these new samples. The second one allows adding new samples to the set of aliquots already collected for a patient.

Batch collection

This procedure also manages the collection of samples but, unlike the previous two, this interface allows uploading a batch of data through a file. A template is available from the interface explaining how this file should be formatted.

Cell lines

The LAS allows the collection of new cell lines. The user must insert the tumor type, the cell line name and, optionally, a datasheet file and an invoice file. In the next interface, other parameters must be defined, as well as the position within the storage in which the line is going to be archived.

Add clinical parameters

The system allows inserting new clinical parameters for an existing collection. Collections can be searched according to patient information, informed consent, protocol, barcode or Geneaolgy ID. When clicking on a collection, a new window will appear, where the user can enter the new parameters.

Aliquots

Aliquot planning

Most of the procedures (derivation, split, experiments, revaluation, transfer, slides preparation) described below have a planning phase in which the user selects the aliquots involved and an operator to whom the work will be assigned. To identify the aliquots, their genealogy ID can be used, as well as the barcode of any container (e.g., plate, rack, freezer), thus scheduling all the aliquots stored in it.

Aliquot validation

For certain procedures a validation phase is required. This occurs before the execution phase and is useful to inform the system about which tubes have been taken from the storage. Since every tube has a unique barcode, entering this code should avoid confusing tubes.

During this validation, in addition to single-tube barcodes, it is possible to enter the barcode of the father container to validate all the aliquots stored in it at once. To take advantage of this feature, it is necessary that all aliquots stored in the validated container have been planned for the current procedure. For instance, if a plate contains 30 samples, but only 25 of them have been planned for derivation, the plate barcode may not be used as a shortcut to validate the aliquots.

Derivation

Through this procedure, the user can derive new biological materials (e.g., DNA, cDNA) from samples stored in the biobank. In particular, one can extract DNA, RNA and proteins from tissue samples, or obtain complementary DNA (cDNA) or RNA (cRNA) from RNA samples. The Procedure has a planning phase followed by an execution phase divided into four steps. In the first step, the user selects the samples to work with, and validates them by entering their tube barcode. Then he/she chooses the derivation protocol to follow. The order in which aliquots will be presented to the user in steps 2, 3 and 4 depends on the order in which they have been validated here. In the second step, a kit must be selected (for more information, please refer to the next section). If the user is extracting proteins, the insertion of a derivation kit is not required. Thus, at the end of step 1 the system will skip to step 3. In the third step, the user must insert the values of the measures made for the samples involved. According to the measurement protocol chosen, the user will insert certain measures; additional measures not expected by the protocol may also be inserted.

The interface of the last step is reported in Figure 4. In the top panel of the interface, information related to the derivation protocol and the output is visualized. In block B, the details for each derivative are reported, to guide the user through the preparation of the aliquots. Some parameters (e.g., volume and concentration) can be changed at will by the user for each aliquot. When clicking on "Recalculate values", the system will automatically update the values of the other parameters to match the protocol rules. Once the technician has prepared the aliquots, he/she should position them in one or more containers (block C). First, an option must be chosen from the "Tube/Plate" selector. Next, the technician can insert the corresponding container barcode. Plates must already be present in the system; for that purpose they must be inserted in LAS through the special interface of the Repository module. Tubes need not be present in the system prior to the collection event. If needed, they can be created together with the aliquot stored in it. For more information about containers, please refer to the description of the Repository module. Newly generated aliquots are then placed

in the available positions using drag and drop. When all aliquots have been processed, a final report is shown.

Figure 4- Derivation interface

Split

It is used to create one or more samples starting from an aliquot. It can be done only on derived (liquid) aliquots. First, the user chooses the aliquots that will be split into parts. Next, in the execution phase, the user sets the number of samples he/she wishes to create and, for each of them, enters the volume and concentration values. Finally, samples must be positioned in a container. First, the user must choose an option from the "Tube/Plate" selector, then he/she can insert the corresponding container barcode. Plates must already be present in the system; for that purpose they must be inserted in LAS through the special interface of the Repository module. Tubes need not be present in the system prior to the collection event. If needed, they can be created together with the aliquot stored in it. For more information about containers, please refer to the description of the Repository module. Then the user can choose to proceed with the splitting of the next planned aliquot, or to end the procedure. In the latter case, a final report is shown.

Slides

The LAS allows the management and tracking of slides (menu item "Slides preparation", under menu category "Aliquots"). Similar to the other procedures, the main steps are planning and execution. In the first step (i.e., planning) the user selects the aliquots to be used and the user who will carry out the procedure. The aliquots can be "Formalin Fixed" or "OCTFrozen". The execution step initially requires the choice of a protocol. Two protocols are currently available: "Microtome" and "Cryostat". The user should then validate the samples and finally create and place the slices on slides. In the latter step, three fields are automatically filled out, based on the selected protocol, but can be modified at will by the user if required. These parameters are: (i) the thickness of the slices, (ii) the number of slices for each slide and (iii) the number of slices obtained from the original aliquot. Positioning is performed automatically: once the code of a slide has been read, several slices are placed in the first available positions according to the settings specified by the user. Additional slices can be placed by clicking on the empty positions of a slide. The user can always cancel each operation by means of the summary table at the bottom of the page.

QC / QA

This procedure permits the insertion of aliquot physical characteristics, such as volume, concentration, purity and quality, and can be done only on derived (liquid) samples. It is used to revalue an aliquot already present in the system and to save who measured which parameter and when. After a planning and a validation phase, the user inserts the measures according to a protocol, and can set the volume used for the revaluation procedure. An "Exhausted" flag can optionally be set on one or more aliquots.

Transfer

This procedure permits the sharing of aliquots between LAS users. The users involved in the sharing procedure are named the "sender" (i.e., the user currently logged in to the system, the one that plans the procedure), the "assignee" (i.e., the operator who will carry out the procedure) and the "addressee" (i.e., the user who will receive the shared aliquots). The latter can be a member of other WGs than the sender's. In this way, the aliquots will be automatically shared with the selected WG. In the planning part, the sender must select the assignee, the addressee, and the aliquots involved. Then the operator in charge executes the shipment, validates the samples and selects the courier. Once the addressee has received the aliquots, he/she can acknowledge it by validating the samples received, and thus terminating the transfer procedure.

Kits

The management of kits is related to the derivation process previously outlined. For example, during the extraction of DNA, a series of reagents and chemical compounds are needed that form the so called derivation kit. Kits are grouped into categories called kit types. Each type has an associated derivation protocol. For instance, to extract RNA three different protocols may be used, and every protocol has its own kit type. The LAS provides an interface to define kit types, and another one to insert a single kit of a specific type. A kit must be present in the system before it can be used during step 2 of the derivation process.

Experiments

The LAS allows the execution of some molecular experiments involving a planning phase and an execution phase. The biobank handles the planning phase for all of them, whereas the execution phase is carried out in the biobank only for some. For the other experiments the execution is managed by the corresponding module. For more information, please refer to section 7 "Molecular Experiments".

In the planning phase the user selects the experiment type, the operator who will carry out the procedure, and the aliquots involved. During the execution phase the operator in charge validates the planned aliquots, thus confirming that the experiment has been executed. An "Exhausted" flag can be optionally set on one or more aliquots.

Patients

A collection is related to surgical interventions performed on a patient. However, the same patient may be subject to several surgical interventions that will produce independent collections. In some experimental procedures, it is very important to identify the aliquots coming from the same patient, to compare possible mutations over time. To retrieve the target aliquots, researchers and technicians may choose one of two alternatives: (i) use the Query Module to build a query that retrieves the aliquots of interest, and (ii) use the patient interface in the BioBank. The first scenario is mainly targeted at more experienced users, who know the relationships among biological entities, as well as which entities contain the information relevant for the search. To enhance the flexibility and the usability of the system, we introduced an ad-hoc interface in the BioBank module to identify aliquots related to a given patient. In this interface, reported in Figure 5, the user can set filtering parameters for the collections, and define security settings. According to the latter, the information in blocks B and C can be displayed in anonymized form. Block B shows the list of patients satisfying the filtering criteria, while block C shows the aliquots related to the selected patient, along with their characteristics. At this point, the user can select one or more aliquots which are moved to the list in block D. When the selection of target aliquots is complete, different actions can be performed (e.g., derivation, molecular experiment). The system will automatically check if the aliquot types are consistent with the selected action.

Figure 5- Patient-based filter interface

Admin

Through this procedure new items can be added to certain predefined lists used by the system in some interfaces. For instance, a new tumor type or a new tissue type can be added to the list that is shown in the collection interface. Similarly, it is possible to insert a courier to the list that is shown during the transfer procedure.

5. Repository

Research laboratories make use of several types of containers to store biological material. Within the LAS, a "container" is defined as anything that is able to contain biological entities (i.e., samples) or other containers. Thus, we can think of the repository as a tree, where root containers contain some objects, which in turn contain more objects. The nesting continues until the leaves of the tree are reached, representing biological samples. All LAS containers are divided into four categories (generic types), listed from the root of the container hierarchical tree: freezer/cabinet, rack/drawer, plate/box, tube/bio cassette. Although this order reflects the actual containment relationships (e.g. a freezer contains a rack), a few exceptions are also possible. For example, a plate may directly contain some biological entities (rather than tubes). Every category is further divided into more container types. For instance, the freezer/cabinet category includes -196°C freezers, -80°C freezers and room-temperature cabinets. For every container type, a series of mutual interactions (i.e., which container type can host another one) has been defined, which can vary according to characteristics such as the layout or the laboratory procedure. For instance, a plate of a given manufacture and model may be able to host only some kind of tubes.

For every container type, several container instances exist, univocally identified by a barcode. Every container has a geometry, defined by the number of its rows and columns. The geometry is used to visually render the container in the user interfaces. Furthermore, every container can contain one or more types of biological entities (e.g. viable, RNA later, snap-frozen), which are tracked by the system.

In the LAS, containers can be managed both as independent entities and as grouped objects. For instance, a plate and its tubes can be handled as a single entity: in some LAS interfaces, by simply loading the plate, both the plate and its tubes are loaded simultaneously (e.g. when positioning new samples on a plate).

The Repository Management Module allows managing any kind of containers by defining and applying different rules to them. In the following the main functionalities are described.

New container

Containers can be managed by accessing the "Container" menu from the main screen.

The insertion of new container instances can be done individually or in batch mode. In the former case, the user must select "Single container" from the "New container" sub-menu and enter the features of each new container he/she wishes to insert. This procedure can turn out useful, for example, in the creation of empty plates (i.e., without any tubes).

In the second case, the "Load batch" interface allows uploading one or more files describing the container(s). By clicking "File Format", a template will be downloaded, showing how this file should be formatted. After clicking "Next step", the system asks for features to be assigned to every new container. To assign the same features in batch to multiple containers, the containers of interest must be selected by means of the checkboxes positioned to the right of each container; next, the appropriate values must be chosen from the menu on the right side of the screen.

New container type

New container types can be easily inserted into the LAS. The user must select the name, the generic type (i.e., freezer, rack, plate, tube), the geometry, and the relationships with other container types

(i.e., the type(s) of container it can contain, and the type of container in which it is contained). In addition, other optional parameters can be specified.

Change container features

This procedure allows changing some features of an existing container. In particular, the user can change the generic type, the container type, the biological entities contained, and the geometry of a container. Users can also choose whether to make a container disposable (i.e., once the biological content has been exhausted, the container must be thrown away). Container features can be changed only if the container is empty. For example, in order to be updated, a plate should not contain any tubes.

Archive container

This interface can be used both to navigate containers and to archive them. For the first purpose, users can navigate the container hierarchy starting from one container and going to the children or the father through the "zoom +" and "zoom -" buttons.

To move/archive containers, two options are available to users.

In the first one, users can drag one container from the source panel over the destination panel.

In the second one, users can select a single container or a block of containers in the source panel, and next click on an empty place in the destination panel. To select a block of containers, users must click on the starting container of the block in the source panel and then on the final container of the block. All containers included within the rectangular area that has been drawn will be selected. For example, by clicking on A1 and then on B2, the containers in positions A1, A2, B1, and B2 will be selected. When moving a container the system verifies the archiving date of all contained samples. If the archive date is not available, it is set to the current date.

Automatic placement of containers is also available. When loading a source container, if the destination container has already been defined, the system checks if the first one can fit into the second one and, if so, it executes the automatic movement to a free position. For example, if the user wants to place a series of tubes in an empty plate, he/she must load the plate as the destination and then just enter the barcode of the tubes. If the user is unhappy with the position assigned automatically, the block can be moved manually. If this automatic procedure is not desired, click on "Deselect container".

The system provides an area called "Temporary place" where some containers can be moved temporarily. This is useful when working with a completely full container, to rearrange the blocks inside of it.

Return

In a laboratory environment, when a researcher takes a sample from the storage to execute an experiment, he/she will need to return it once his/her activity is finished. The LAS tracks such operations and, whenever the validation of an aliquot occurs in one of the biobank interfaces, it records that the aliquot has been taken by a certain operator. When the operator wants to hand back the tube, he/she has to perform a "return" procedure in the LAS. The interface shows all containers taken by that operator and, to return them, the operator must enter their barcode. Once the return operation has been confirmed, a final report is shown.

6. Xenografts

In vivo experiments (e.g. xenografts) allow testing different drug therapies and expanding the collection of biological samples. This kind of experiments in the IRCCS institution is based on the model described in Bertotti et al. (2011)[1]. Each aliquot implantation defines a biological entity named Biomouse. Since multiple tumor aliquots may be implanted at different sites within the same animal, the physical animal can host several Biomice. Biomice can be grouped into Experimental Groups according to the purpose of the research. In this way, researchers can study tumor growth and treatment response on a subset of animals. The physical animals can assume different statuses which affect the life cycle of the hosted Biomice. For instance, if the status of an animal is set to "dead accidentally", it will no longer be possible to perform measures and explant activities on all of its hosted Biomice. The Xenografts Management Module manages the immunocompromised animals and monitors the xenograft life cycle, from their acquisition by the research institute to their death. The main functionalities provided by this module are described in the following.

Animals Management

The set of functionalities related to the management of animals can be found under the "Animals" menu item.

Animal registration

Under the "Register New" interface, the user can register the acquisition of new animals and fill in different features (e.g., status, strain, age, source) to describe them. The system promotes the use of barcode readers when animals are equipped with RFid tags, in order to speed up the identification of the animal and the retrieval of related information.

Change animal status

Each animal is characterized by a status, describing its current condition within the experimental platform (e.g., breeding, experimental, dead accidentally, transferred, sacrificed).

While some procedures carried out within the Xenografts Management Module will automatically cause a transition from one status to another (e.g. the implant operation will take the newly implanted animal from the "experimental" status to the "implanted" one), it may be necessary in some cases to manually set the status of a specific animal (e.g. when a mouse is found to be accidentally dead).

Through this interface, the user may select a new target status for the animal. According to the selected status, additional information may be requested. As the animal barcodes are inserted, the system checks if any conflicts arise due to the change of status, and inserts the corresponding BioEntities (e.g. Biomice) in a list. BioEntities can be removed from the list at any time by clicking on the red cross button. By clicking "Finish", the new status for the selected animals will be permanently saved.

---

[1]Bertotti, A., Migliardi, G., Galimi, F., Sassi, F., Torti, D., Isella, C., & Trusolino, L. (2011). A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. Cancer Discovery, 1(6), 508-523.

Manage Experimental Groups

This set of functionalities allows the management of experimental groups of Biomice. The user can perform the following actions:

- Create a new experimental group: a name and the expected experimental protocol are requested
- Load existing/recent groups: by selecting the groups of interest from the listed shown by the system, the user can visualize a description of the group and the set of animals belonging to it
- Edit group: the user can rename the group and change the expected protocol
- Move animals: after the selection/creation of two groups, the user can move the animals belonging to one group to the other one

Measurements

Xenopatients can be regularly measured to monitor tumor growth. The Xenografts Management Module keeps track of measurement events of several types, in the corresponding "Measurements" interfaces. There are 4 modes of measuring xenopatients:

- *Observation.* The tumor dimension can be evaluated by means of a categorical measure (i.e., None, Very Small, Small, Medium, Large), optionally associated with the weight of the animal.
- *Weight.* This functionality allows measuring only the weight of the animal
- *Tumor volume.* This functionality allows measuring the tumor mass by means of instruments (e.g. caliper) and optionally tracking the weight of the animal
- *Tumor volume + weight.* It requires weighing the animal and measuring the volume of the implanted tumor mass

In each case, the user can contextually execute a number of actions, such as plan treatments, interrupt the treatments in progress, plan the sacrifice of the animal or the explant without sacrifice. In the former two modes, the user can optionally request a review from the Principal Investigator, while in the latter two all actions are always sent to the Principal Investigator for a review.

Surgery

Surgical operations, which include the implant of tumor tissue aliquots into animals and the explant of tissues from an implanted animal, are tracked by corresponding functionalities grouped under the "Surgery" menu item.

Implant Xenografts

The system can handle tumor tissue implants into the available recipient mice. The user should select an animal, the site of implantation and the source aliquot. The same animal can be implanted at different sites. Both the animal and the aliquot are retrieved by means of their barcodes (i.e., RFid barcode for the animals, container barcode for the aliquots). The system automatically generates a GenealogyID for the newly generated Bioentity. Additional information such as data of the

implantation, expected protocol and notes can be edited. Once the implantation session is finished (click on "Next step" button), the system asks if the containers used for the session should be emptied. If the user confirms, the containers are emptied and all the aliquots stored in these containers will no longer be available. Conversely, if the user declines, possible leftover aliquot pieces are not discarded and remain available in the biobank. In any case, the quantity of each aliquot used during the implantation is updated accordingly. Then the system automatically defines new experimental groups according to a set of default parameters. The user can redefine these parameters, create new groups and/or move the animals among the groups.

Explants Xenografts

The tumor explant marks the end of the xenograft life cycle. The user should select the date of the explant (the current date is automatically selected by the system), the set of tissues that will be explanted, and the animals that have been programmed for this operation. Animals in the "waste" status can also be explanted. By inserting the animal barcode and the implantation site, the system retrieves its associated information. The interface for the collection of biological samples through xenograft explant is similar to the one of the Biobank collection (see Figure 3). Similarlto the collection procedure of the Biobank, plates or tubes can be loaded or created in the application by means of their barcode. Inserting an aliquot in a given plate position is done by simply clicking on the corresponding cell. By clicking on the "Submit" button, the explant session is closed and the aliquots are stored in the corresponding containers. The status of each explanted animal is set to "explanted".

Treatments

Treatments can be created, browsed and finalized through the "Treatments" functionalities.

Finalize Treatments

A list of experimental groups for which a treatment has been programmed is shown. By selecting the group of interest, the system shows the list of programmed treatments and, for each treatment arm, the list of animals selected for the drug administration. By inserting the barcode of the animal, the system checks if it is programmed for the programmed treatment and, if there is no error, records the start of treatment.

Browse/Create Treatments

Using this functionality the user can see the details of the treatment protocols recorded by the system. In particular, the list of steps for each treatment arm can be visualized. Each step defines the used drug, the dosage, the administration mode and the number of times in the time unit (e.g., day, hour). Moreover, the user can define a new treatment protocol, for which he/she should assign a unique name and a description. The protocol will be composed of several arms that can be created and/or selected among the existing ones. When a new arm is created, some additional information, i.e., the set of steps and their composition should be defined.

Experiments

New measurements check

Immunocompromised animals that have been implanted with a tumor mass are regularly measured by laboratory technicians to monitor the growth of the tumor and assess the effectiveness of treatments (see Measure paragraph). During this operation, technicians record the observed values and may propose a given course of action for each mouse (e.g., start/stop a treatment, explant). The new measurements and new actions should be reviewed by the supervisor through the "New Measurements" review screen (see Figure 6). In the left side of the window (block A) a list of all experimental groups for which new measurements exist is shown. When an experimental group is selected, its measurements and all associated pending actions are shown in the bottom table (block D). To ease the supervisor's task, any pending action awaiting approval is highlighted, while further information is available in the other tabs (e.g., historical measurement record for each mouse). Moreover, a plot showing the average variation of the tumor mass over time for each treatment arm (block C) provides a useful tool to support the supervisor in the decision-making process. The supervisor may accept the actions proposed by the technician for any measurement, or he/she may modify or reject them, by means of a set of buttons (block B). He/she may also add comments, such as explanations or future directions. Once the review is complete, a report is displayed, including all actions (both approved and rejected or modified ones), and sent to both the supervisor and the technician.

Ongoing experiments

This functionality is similar to the New measurements check, but all experiments in which at least one animal is alive are listed. The user can select the experimental group of interest and perform the same actions available in the New measurements check functionality.

Archive experiments

In this screen all closed experiments (i.e., experiments in which all of the animals have been explanted) are listed. The user can only display the information of the experimental group of interest, see the graphical representation of tumor volume measurements and the history of the data collected for each animal. No further action can be applied to the animals.
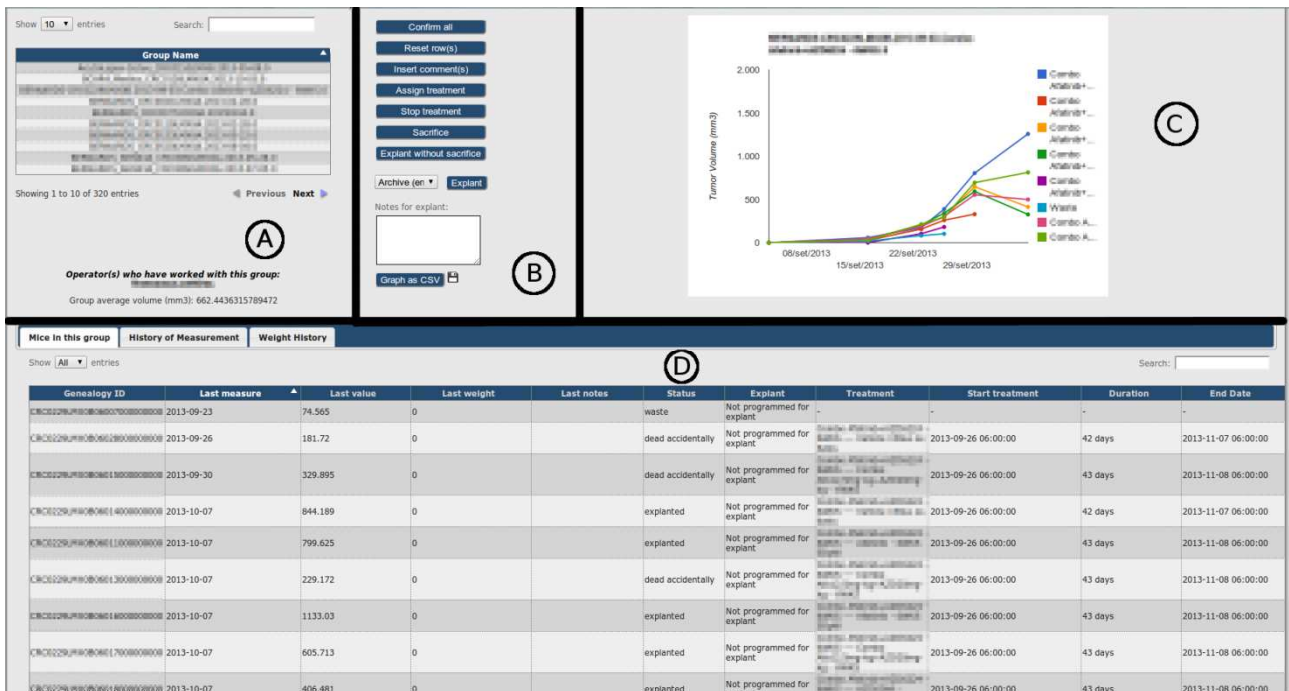
Figure 6- Experiment review interface

Admin

Through this procedure, new elements may be added to certain predefined lists used by the system throughout its interfaces.

The system allows inserting new records for the following objects:
- Methods of drug administration
- Explant scope details
- Sites of implant
- Drugs
- Tissue types
- Mice strains
- Mice suppliers

For each record some attributes are required, while other are optional.

Batch

To speed up the loading of large batches of data, the system allows uploading a text file containing the data for certain types of procedures. In particular, the user can load data corresponding to the following actions:
- register new animals
- implant animals
- explants animals
- make observations of tumor mass

- record tumor volume and animal weight

A template file for each operation can be downloaded after selecting an action type, by clicking on the "Template" link. The template lists all required fields and includes some sample data, which must be replaced with the actual content.

7. Cell Lines

In vitro experiments (e.g., cell lines) allow the manipulation of tumor cells in a controlled, artificial environment and the test of the effects of different drug therapies. The LAS system defines a Cell Line as the set of bioentities that are generated from the same source and are under the same experimental conditions. The experimental conditions are defined by the protocols that describe the type of process (i.e., adherent, suspend, organoid) and the set of culturing conditions applied. The culturing conditions are defined by a set of items categorized in the following groups: (i) Nutrients and chemical, (ii) Hormones/growth factors, (iii) Antibiotics, (iv) Serum, and (v) Media. The Cell Line Management Module aims at tracking the life cycle of cell lines and monitoring the experimental conditions applied to each cell. The main functionalities provided by this module include the following.

Protocol Manager

To properly manage Cell Line entities, the user should define protocols for both generation and expansion procedures. Each protocol defines on which procedures it can be applied (i.e., generation or expansion) and the type of process (i.e., adherent, suspend, organoid) exploited on cell lines. The system requires a unique name for each protocol. Moreover, users can describe the items (i.e., culturing conditions) used by the protocols in manipulating the cell lines. The culturing conditions are categorized in the following groups: (i) Nutrients and chemical, (ii) Hormones/growth factors, (iii) Antibiotics, (iv) Serum, and (v) Media. The user can select one or more culturing conditions and define associated parameters (e.g., concentration, dosage), where it is required.

Generation

Cell Lines entities can only be generated from viable aliquots. In this procedure, the user can generate new cell lines by selecting one or more generation protocols. The system automatically generates the new Cell Line entities according to the information of the selected protocols. The user can schedule a cell line generation for a set of aliquots from the Biobank module. This set of aliquots is visualized in the "Pending Generation" page. The user can also generate aliquots without a schedule using the "New Generation" interface. In both pages, the user should select the generation protocols of interest and then select the aliquots from which he/she wishes to generate new Cell Lines. Similar to the implantation of xenografts, in the New Generation page, viable aliquots can be loaded by means of the container barcode.

Expansion

The user can visualize and manage the currently available cell lines owned by his/her WGs. The user can perform a set of operations on each cell line: (i) trash one or more plates, (ii) expand the cell line by defining dilution parameters, expansion protocols and the number of output plates for each protocol, (iii) send a set of plates to experiments (fake viable aliquots are created in the Biobank), and (iv) plan an archive procedure. It is possible to filter the list of shown cell lines according to their GenealogyID or the last user who manipulated the lines.

Archive

In the expansion interface the user can schedule one or more cell lines for the archive procedure. All scheduled cells are listed in the archive page. For each cell line, the user can archive different aliquot types with their associated parameters (e.g., volume [ml], count [cell/ml]) and position them in the destination container.

Thawing

Similar to the generation procedure, the user can create new cell lines from the viable aliquots generated from archived cell lines. The user can schedule a cell line thawing for a set of aliquots from the Biobank module and then process them from the "Pending Thawing" interface. As an alternative, he/she can load containers by means of their barcode and select the target aliquots from the "New Thawing" page. In both cases, the user should select the generation protocols exploited for the operation and then select the aliquots from which he/she wishes to generate new Cell Lines.

8. Molecular Experiments

8.1. Digital PCR

Digital polymerase chain reaction (i.e., dPCR) is a refinement of conventional polymerase chain reaction methods that can be used to directly quantify and clonally amplify nucleic acids including DNA, cDNA or RNA. Unlike traditional PCR, dPCR separates the sample into a large number of partitions (i.e., droplets) and the reaction is carried out in each partition individually. This separation allows a more reliable collection and sensitive measurement of nucleic acid amounts. The method has been demonstrated useful for studying variations in gene sequences (e.g., Copy Number Variation), quantifying cancer biomarkers, and studying gene expressions. The Digital PCR module manages and tracks the samples analyzed using this technology. The following activities are available.

Validate samples

The user must first plan a set of aliquots for a Digital PCR experiment from the Biobank module. During the experiment planning phase, the user also defines a title and a description (optional). Before performing the molecular experiment, the user should validate the scheduled aliquots and confirm the used quantity for each aliquot. The barcode of each aliquot must be entered to perform the validation procedure.

Experiment layout definition

All validated requests are visualized. After selecting an experiment, the user should define the instrument and the layout of the support used in the experiment. Once these parameters have been defined, the system shows the samples from the selected experiment and the layout of the support. The mutation targets used by the user in other experiments are shown in the "Targets" table. It is possible to add new targets by performing a search according to the gene symbol and the mutation syntax. The samples and the targets can be associated with the positions of the support by clicking on a single cell or the column/row header. The positions of the support will be updated with the sample label and/or target color. A position may contain only one sample and one target. Only the positions containing both a sample and a target will be saved into the system.

Measurement collection

All experiments for which a layout has been defined are visualized. By selecting an experiment, the user can upload a file with the digital PCR results. A table with the measurements is shown, according to the contents of the file. In addition to the measurements, the user can also associate a set of files with an experiment.

8.2.Beaming

The Beaming technology can be used to detect mutations having occurred in cancerous cells without the need for a physical biopsy. This technology can be used to detect mutations in circulating tumor DNA (ctDNA), which can be found at very low concentrations in peripheral blood samples. Using this type of analysis, researchers and clinicians can obtain vital information about cancer initiation, development, metastasis, and response to therapy without the need for invasive surgery. The Beaming module manages and tracks the samples analyzed using this technology. The following activities are available.

Validate samples
The user must initially plan a set of aliquots for an experiment using the Beaming technology from the Biobank module. During the experiment planning the user also defines a title and a description (optional). Before performing the molecular experiment, the user should validate the scheduled aliquots and confirm the used quantity for each aliquot. To perform the validation procedure, the user should insert the barcode of each aliquot.

Experiment layout definition
All validated requests are visualized. After selecting an experiment, the user should define the instrument and the layout of the support used in the experiment. Once these parameters have been defined, the system shows the samples from the selected experiment and the layout of the support. The mutation targets used by the user in other experiments are shown in the "Targets" table. It is possible to add new targets by performing a search according to the gene symbol and/or the mutation syntax. The samples and the targets can be associated with the positions of the support by clicking on a single cell or the column/row header. The positions of the support will be updated with the sample label and/or target color. A position may contain only one sample and one target. Only the positions containing both a sample and a target will be saved into the system.

Measurement collection
All experiments for which a layout has been defined are visualized. By selecting an experiment, the user can upload a file with the liquid biopsy results. A table with the measurements is shown, according to the contents of the file. In addition to the measurements, the user may also associate a set of files with an experiment.

8.3.Microarray

With the advent of microarray technology, thousands of gene expression values can be analyzed with a single experiment to identify markers of a target phenotype. Thus, to assess the quality of the analyses and improve further integration tasks, it is fundamental to track both the biological information related to the sample source and the entire experimental process. The Microarray Module aims to track all the steps of a microarray experiment. The main functionalities it provides are listed in the following.

External requests

Microarray experiments can be planned for both internal and external aliquots. All of the aliquots stored in the Biobank module are considered "internal aliquots". By using the "Experiment" functionality in the Biobank module, the user can define the set of aliquots on which the microarray experiment will be performed, the title and a description (optional) for this request. In order to also offer the management of aliquots collected only for the microarray facility, the system allows the insertion of external requests. For this kind of requests, the user should provide a unique identifier and a minimum set of information (i.e., volume and concentration) for each aliquot. It is possible to upload a file containing the description of these aliquots.

Plan hybridization

The first step of a microarray experiment is the hybridization of samples positioned on chips. The user should define the instrument and the hybridization protocol that will be used. Then, he/she can define a number of (virtual) chips, even with different layouts (i.e., number of positions), and position the aliquots belonging to several requests. The requests visualized are both those assigned to the current user, and those not assigned to any user; they can be external requests (see External request paragraph) as well as those defined in the Experiment procedure of the Biobank. The system advises the user on how to prepare the aliquots positioned on the virtual chips.

Retrieve aliquots

Before applying the hybridization procedure, the user should validate all of the planned aliquots and confirm the used quantity. After selecting a hybridization plan, the user should insert the barcode of each aliquot to perform the validation procedure. The system automatically updates the volume of each aliquot according to the hybridization protocol, although the user can still modify the used volume.

Hybridization

The hybridization procedure is used to prepare the chips before the scan, by positioning the aliquots on them. In this functionality all the hybridization plans for which the aliquots have been validated are visualized. After selecting a plan, the user should enter the barcode of the chips. The system will automatically move the planned aliquots onto the chip according to the chip layout and the virtual hybridization plan. The user can modify the position of each aliquot by using the drag & drop function.

Scan

The hybridized chips are ready to be inserted into the scanner. The user selects the scan protocol which defines the scanner and its settings. By inserting the chip barcode the system shows the chip contents (i.e., aliquot identifiers), and automatically positions the chip in the first available position of the scanner. The chip location can be changed by using the drag & drop function.

Pending scans

Some scan procedures require a Q/C evaluation. For the chips scanned with this kind of protocols, the user should specify for each aliquot if the scan was successful or not. After selecting a scan, the system shows the chips belonging to the experiment of interest. For each aliquot in each chip, a "successful scan" flag is automatically set by the system, but the user can change this status by simply clicking on the target aliquot.

New chip

Chips used in the experiments should be described by the following set of information: (i) barcode, (ii) type, (iii) owner, (iv) expiration date , (v) lot number, (vi) link to Dmap file.

Chips Management

Each chip inherits some characteristics from its chip type. For instance, all chips with the same type have the same layout. During the definition of new chip type, the user defines the number of positions, while the system suggests an identifier for each position.

Hybridization protocol

The hybridization protocols are used to define the rules applied to prepare the aliquots and the chip before the scan. A hybridization protocol includes the following features: (i) unique name, (ii) load quantity (ng), (iii) hybridization buffer (ul), (iv) hybridization temperature (°C), (v) total volume (ul), (vi) hybridization time (hours) (vii) denaturation temperature (°C), and (viii) denaturation time (minutes).

Scan protocol

A scan protocol describes the main characteristics of the process by means of the following parameters: (i) unique name, (ii) used software, (iii) used instrument, and (iv) whether the Q/C evaluation of the result is required or not. According to the instrument information, a set of scanner parameters are visualized and can be modified.

Experiment

Using the chip barcode, the user can browse the files associated with the scan event and with each aliquot positioned on the chip. The file can be downloaded for further analyses. The microarray raw data coming from the scans are remotely loaded into the Document Storage Manager and linked to the corresponding sample by the provided software.

8.4. Next Generation Sequencing

The Next Generation Sequencing module manages and tracks samples analyzed with NGS technology. The following activities are available.

Validate samples

The user can plan a set of aliquots for an experiment using the Next Generation Sequencing (NGS) technology from the Biobank module. During the experiment planning the user also defines a title and a description (optional). Before performing the molecular experiment, the user should validate the scheduled aliquots and confirm the used quantity for each aliquot. The user should insert the barcode of each aliquot to perform the validation procedure. Moreover, the instrument, the assay and the kit used in the experiment must be defined before validating the samples and concluding the procedure.

Upload results

All validated requests are visualized. After selecting an experiment, the user can upload a set of files for one or more samples. The files associated with each sample can be visualized and removed if necessary. Every time a set of files is associated with one or more samples, the previous association is overwritten. Moreover, the user can update the used volume for each sample, mark whether the aliquot has been exhausted and/or whether the experiment on the sample has failed. No files can be associated with failed samples.

Download results

To retrieve the files associated with analyzed samples, the user can exploit the following filters: (i) Patient code, (ii) Informed consent, (iii) Collection protocol, (iv) Barcode, and (v) Genealogy ID. Filters are evaluated by performing a logical AND of all conditions. For each sample, the system visualizes the associated files for each experiment. The user can download the files of interest, which will be saved as a zipped folder with the same structure as the one visualized on screen. Each search resets the set of visualized samples.

8.5. Real Time PCR

The real-time polymerase chain reaction (RT-PCR) is a laboratory technique of molecular biology based on the polymerase chain reaction (PCR). It is used to amplify and simultaneously quantify a target sequence. For one or more specific sequences in a DNA sample, quantitative PCR enables both detection and quantification. The RT-PCR module manages and tracks this kind of experiments. A single experiment can include many samples analyzed for several amplicons. The module stores the results for each analyzed sample, using the LAS genomic annotation facility. The experiment description can be enriched with the raw data extracted by the instrument.

Validate samples

The user can plan a set of aliquots for an experiment using the Real Time PCR technology from the Biobank module. During the experiment planning the user also defines a title and a description (optional). Before performing the molecular experiment, the user should validate the scheduled aliquots and confirm the used quantity for each aliquot. The user should insert the barcode of each aliquot to perform the validation procedure.

Upload results

All experiments for which the samples have been validated are visualized. By selecting an experiment, the user can upload a file with the Real Time PCR results. A table with the measurements is shown, according to the contents of the file. In addition to the measurements, the user may also associate a set of files with an experiment.

8.6. Sanger Sequencing

Several methods can be exploited for DNA sequencing. Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides, named primers, by DNA polymerase during in vitro DNA replication. Differently, Next Generation Sequencing (NGS) enables rapid sequencing of large stretches of DNA base pairs spanning entire genomes, with some instruments capable of producing hundreds of gigabases of data in a single sequencing run. The Sanger Sequencing module is aimed to track the experiments performed with both technologies. In particular, for each sample included in an experiment, the gene and the mutation identified by external analyses are saved using the genomic annotation facilities provided by the LAS. Additional resources such as the electropherograms produced by the instruments can also be included to enrich the description of the experiment and reevaluate it in the future. These documents and the NGS raw data are managed by the repository module described in the next section.

Validate samples

The user can plan a set of aliquots for an experiment using the Sanger Sequening technology from the Biobank module. During the experiment planning the user also defines a title and a description (optional). Before performing the molecular experiment, the user should validate the scheduled aliquots and confirm the used quantity for each aliquot. The user should insert the barcode of each aliquot to perform the validation procedure.

Upload results

All experiments for which the samples have been validated are visualized. By selecting an experiment, the user can upload a file with the Sanger Sequencing results. A table with the measurements is shown, according to the contents of the file. In addition to the measurements, the user may also associate a set of files with an experiment.

8.7. Sequenom

The Sequenom SNP genotyping uses a bead-less and label-free primer-extension chemistry to generate allele-specific products with distinct masses. Allele discrimination is conferred by primer extension with a high-fidelity polymerase enzyme across the SNP site, leading to differences in mass of extended products. Differentiation of genotypes for each of the SNPs multiplexed in one assay is a result of unique mass ranges for the extension primers. The Sequenom module stores the results for each analyzed sample, using the LAS genomic annotation facility.

Upload results

The user can plan a set of aliquots for an experiment using the Sequenom technology from the Biobank module. During the experiment planning phase the user also defines a title and a description (optional). After the experiment definition, the user can upload a file with the Sequenom results. A table with the measurements for the aliquot scheduled for this kind of experiment is shown, according to the contents of the file. In addition to the measurements, the user may also associate a set of files with an experiment.

9. Query

Tracking experimental procedures and recording all the data related to biological samples is fundamental to monitoring molecular and in vivo experiments. However, researchers also aim at integrating such heterogeneous information to generate new knowledge. The Query Module can extract all information of interest from the operative tier in a uniform way by exploiting graphical tools. To date, the module includes the following functionalities.

Query Generator

The Query generator aims at performing queries and/or defining templates representing frequent queries. The query set is defined as a workflow composed of blocks, named query blocks. Each query block defines the object that will be retrieved (e.g., aliquot, xenopatient), its related information of interest and the filtering conditions. Before retrieving the data from the corresponding modules, the workflow is analyzed to detect improperly defined operations (e.g. intersections among disjoint sets of objects) and to define an optimal execution plan on the distributed databases. The templates and/or translators defined by the user will only be shared with users belonging to the WGs of the owner. The query blocks are shown on the left hand side of the editor (block A in Figure 7) and categorized according to the source from which the data are drawn (e.g. the flask icon for biobank data, the mouse icon for xenopatient data). The user can drag and drop the query blocks on the workflow editor (block B) and build the workflow of the query by connecting them. Filtering conditions may be specified for each block by clicking the funnel icon. Set operators (union, intersection, difference) and special operators (group-count, extend, template blocks), listed in block C, can also be used in the workflow. Once the workflow has been defined, the user can assign a title and a description to the query (block D). Finally, he/she can run the query or save it in the system, to be reused in the future for different purposes.
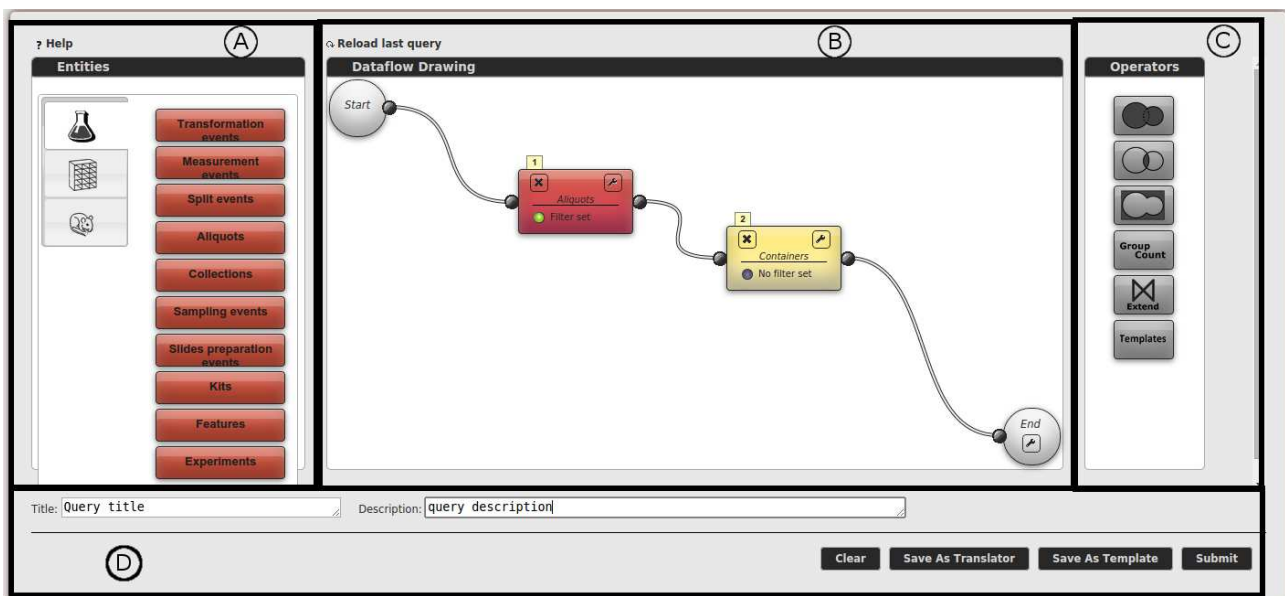

Figure 7- Query generator interface

Sharing

When the results of a query include shareable Bioentities (e.g., Aliquots, Biomice, Cell Lines), enabled users (see "User profiles" in the "Authentication" section) can share a set of records with one or more WGs.

History

Each user can visualize a chronological list of his/her own launched queries, with some additional information. Given a query, the user can (i) visualize the query workflow (the query generator will be opened), (ii) directly re-run the query, or (iii) visualize the results from past runs.

Edit templates

Templates can be visualized and modified by users within the same WGs as the template owner. When an edited template is submitted, if the template structure (in terms of query blocks) has changed, a new template will be saved, and no modification will be made to the original template. On the other hand, if only filter constraints and/or filter names have been modified, the original template will be updated accordingly.

Edit translators

Translators can be visualized and modified by users within the same WGs as the translator owner. When an edited translator is submitted, if the query structure (in terms of query blocks) has changed, a new translator will be saved, and no modification will be made to the original translator. On the other hand, if only filter constraints have been modified, the original translator will be updated accordingly.

10.   Genomic Annotation

Thanks to a plethora of independent projects, maintained by universities and research institutions, nowadays genomic information is largely available in a number of public, freely accessible databases. On the other hand, the lack of universally recognized standards for data representation can make it difficult to select a definitive and unique source of information. In addition, these databases are not always interchangeable, since data from one source may be missing, incomplete, or simply different in the other, so that finding a one-to-one correspondence is sometimes impossible. To address these issues, the Genomic Annotation Manager has been developed, which aims to integrate heterogeneous data sources and provide a uniform and non-redundant knowledge base for genomic annotation. For instance, genes, mutations and SNPs (single-nucleotide polymorphisms) can be unambiguously identified by all modules within the LAS system. So far, sequence mutations from COSMIC: Catalogue of somatic mutations in cancer (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/), and gene/transcript/exon information from assembled by the Gencode Project (http://www.gencodegenes.org/) have been imported into the Genomic Annotation Manager. Furthermore, Blat (http://genome.ucsc.edu/cgi-bin/hgBlat) has been integrated to support different molecular experiments and annotation procedures.

The Genomic Annotation Manager is currently under development. Despite providing some essential services to other operative modules behind the curtain, it currently does not yet offer any user-oriented interfaces.

11. Analysis

The analysis module provides a tool to group a set of objects according to their associated measure (see the template file to correctly format your data). The groups defined by the algorithm will have the same number of objects, a similar mean value, and minimum standard deviation. The user can define a set of parameters to discard objects from the analysis, define the expected mean value and its acceptable range, the number of groups and the number of objects per group. The system will find a random categorization to fit all the requirements.

12. Citation

The LAS platform is the result of a joint effort by both IT and biomedical researchers of the Candiolo Cancer Institute. If you wish to publish data and/or results obtained with the support of the LAS platform, please cite at least one of the following papers:

1. Fiori, A., Grand, A., Alberto, P., Geda, E., Brundu, F. G., Schioppa, D., & Bertotti, A. (2014). A Case Study of a Laboratory Information System Developed at the Institute for Cancer Research at Candiolo. Laboratory Management Information Systems: Current Requirements and Future Perspectives: Current Requirements and Future Perspectives, 252.
2. Baralis, E., Bertotti, A., Fiori, A., & Grand, A. (2012). LAS: a software platform to support oncological data management. Journal of medical systems, 36(1), 81-90.