# CytomicsDB: A Metadata-Based Storage and Retrieval Approach for High-Throughput Screening Experiments

E. Larios[1], Y. Zhang[2], L. Cao[1], and F.J. Verbeek[1]

[1] Section Imaging and Bioinformatics, LIACS, Leiden University, Leiden, The Netherlands
{e.larios.vargas,l.cao,f.j.verbeek}@liacs.leidenuniv.nl
[2] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
{Ying.Zhang}@cwi.nl

**Abstract.** In Cytomics, the study of cellular systems at the single cell level, High-Throughput Screening (HTS) techniques have been developed to implement the testing of hundreds to thousands of conditions applied to several or up to millions of cells in a single experiment.

Recent technological developments of imaging systems and robotics have lead to an exponential increase in data volumes generated in HTS-experiments. This is pushing forward the need for a semantically oriented bioinformatics approach capable of storing large volume of linked metadata, handling a diversity of data formats, and querying data in order to extract meaning from the experiments performed.

This paper describes our research in developing CytomicsDB, a modern RDBMS based platform, designed to provide an architecture capable of dealing with the computational requirements involved in high-throughput content. CytomicsDB supports web services and collaborative infrastructure in order to perform further exploration of linked information generated in each experiment.

The objective of this system is to build a semantic layer over the data so as to enable querying metadata and at the same time allowing scientists to integrate new tools and APIs taking care of the image and data analysis. The results will become part of the metadata of the whole HTS experiment and will be available for semantic post analysis.

## 1 Introduction

High-Throughput Screening (HTS) is a well-established process in drug discovery for pharma and biotechnology companies and is now also being set up for basic and applied research in academia and some research hospitals [10]. Recent developments in microscopy systems and robotics enabled large-scale screening of cellular systems. A popular screen setup is automated time-lapse confocal image acquisition which enables capturing of e.g. high content subcellular information (derived as features) or dynamic aspects like cell migration. Cells are exposed to hundreds and even thousands of different conditions using one or several multiwell (96, 384, 1536) plates. This typically results in 20-40 GB of data consisting of in the order of 100,000 - 200,000 images in an overnight experiment.

In cytometry, HTS-experiments are usually employed in the context of functional analysis, closing the gap between genomics-proteomics and functional responses on the

M. Comin et al. (Eds.): PRIB 2014, LNBI 8626, pp. 72–84, 2014.

cellular level. Examples are genome wide siRNA screens, where all existing genes are lowered in activity one at a time using siRNA mediated knock down followed by some cellular-level phenotypic readout, e.g., cell migration speed, focal adhesion dynamics, subcellular morphological changes, cell death.

A next step in the HTS-experiment pipeline is image quantification using image analysis software tools. In this manner, biological hypothesis can be statistically tested using the quantification results from the image analysis stage, and can depict an objective understanding of the cell response to various treatments or exposures.

In a typical HTS workflow, spreadsheet applications are commonly used for bookkeeping all information related to the design of the multiwell-imaging plates, image analysis quantification results and even statistical analysis results. This approach has many drawbacks. Firstly, it is extremely difficult to link the data produced during the different stages of an HTS experiments, such as linking the images generated in the HTS experiment and the metadata collected during the design of the plate layout. Secondly, it is highly prone to man made errors. The lack of standards, formats and a centralized place for storing the information makes it difficult to promote a collaborative environment with or between research groups. Finally, spreadsheet applications are not suitable for knowledge discovery, as they do not allow to combine sophisticated visualization and querying of the (meta)data previously stored.

In our previous work [8], we presented the initial design of a platform for managing and analyzing HTS images resulting from cytomics screens taking the automated HTS workflow as a starting point. This platform *seamlessly* integrates the whole HTS workflow into a single system. The platform relies on a modern relational database system to store user data and process user requests, while providing a convenient web interface to end-users. Using this platform, the overall workload of HTS experiments, from experiment design to data analysis, can be significantly reduced. Additionally, the platform provides the potential for data integration to accomplish genotype-to-phenotype modeling studies. In this paper, the initial design, particularly, the database model, has been rigorously revised and generalised to manage all kinds of metadata produced by automated HTS systems. We call our system *CytomicsDB*, which is designed as a user oriented platform but considers the HTS workflow as a template for managing, visualizing and querying the metadata.

Current software and architectures for HTS are mostly based on generic Lab Information Management Systems (LIMS) [12], which face significant challenges to accessing, analyzing, and sharing the data required to drive day-to-day processes within the laboratory. Furthermore, the limited connectivity to other legacy systems and poor visualization of the data is an obstacle to extract new insights from the data stored, and cause a deep impact in the efficiency of the HTS experiment. Comparing with the existing LIMS systems, CytomicsDB has a number of important advantages:

1. Ease of promoting scientific collaborations. Since all data in CytomicsDB are centralized, granting access to collaborators or sharing information has been made simple;
2. Flexibility for integration with other legacy systems. It it common to use external APIs for performing image and data analysis results, such as Weka, PRTools. In the

design of the architecture of CytomicsDB, special care has been taken to assure the possibility of invoking external API through web services.
3. The web-based architecture allows its users to easily access to their experiments data from wherever and at any time. The architecture also allows the whole or parts of the system to be smoothly moved to a Cloud based environment.
4. The capability to drill-down through experiments' metadata due to the metadata-based approach.
5. A single interface for visualization of all experiments data, include raw images, metadata and analysis results.
6. Pattern recognition (PR) within an experiment and PR across HTS experiments.

To sum up, the contributions of this work include:

1. Metadata organization in an HTS experiment (Section 2).
2. Data modeling and storage (Section 3).
3. A case study in endocytosis of EGFR, describing how a Metadata-based RDBMS approach can facilitate the identification of EGFR dynamics and classification of EGFR phenotype stages (Section 4).

Finally, in Section 5 we discuss related work and in Section 6 present our conclusions.

## 2 Metadata Organization in an HTS Experiment

The metadata of an HTP experiment consists of a variety of types and formats and has been grouped in five levels as showed in Figure 1: Project, Experiment, Plate - Wells, Video/Images and Measurements. These levels contain each other in a cascade fashion, for instance: [1] Project contains [1..n] Experiments, [1] Experiment contains [1..n] Plates, [1] Plate contains [24,48,96,384] Wells, [1] Well contains [1..n] Video/Images and finally [1] Well contains [1..n] measurements.

**Project.** This level contains a title which describes the aim of the project, the duration, the author, etc. When a project is created, its creator becomes its administrator and is possible to grant access to another scientist in order to promote a collaborative environment.

**Experiment.** Figure 2 shows the structure of the metadata contained in the Experiment level. This level is divided in Hardware and Type of Experiment. Firstly, the metadata associated to the hardware correspond to the microscope and the imaging technique used. Depending on which microscope is used, the set of imaging techniques differs. For instance, the imaging techniques available for a Becton Dickinson (BD) Pathway microscope are EPI, Spinning disk or Bright Field, but in a Nikon TE 2000-e microscope it is possible to use: FRAP, FRET, EPI, Confocal, Spectral or DIC. Secondly, the metadata associated to the type of experiment can be separated in four groups: (1) Fixed or Live experiment including a 2D or 3D option for each case; (2) Assay type, in this case there are the following options: migration/invasion, proliferation, primary tumor, apoptosis and sub cellular perturbations; (3) Species, the options available are: human, rat, mouse and zebrafish; and (4) Cell / Tissue origin, considering in this area: primary, cell line, iPSC, stem cel, biopsy, etc.

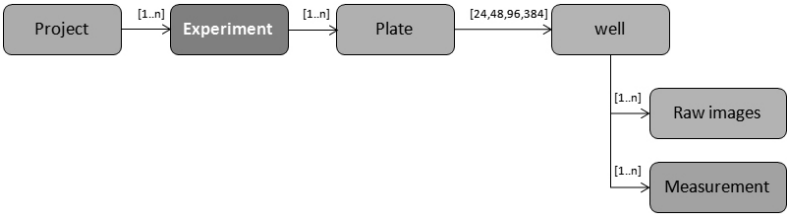


**Fig. 1.** Structure of the metadata in an HTS experiment

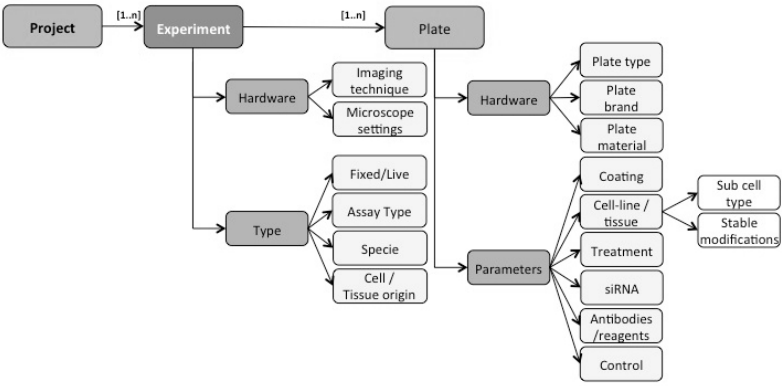


**Fig. 2.** Structure of the Experiment and Plate metadata

**Plate-Wells.** This level is also divided into two groups: metadata about the Hardware and about the Parameters used. Figure 2 shows an UML diagram of the structure of these two groups. The Hardware sub level includes information about the plate type, the brand and the fabrication material. The level of Parameters includes information about (1) Coating, (2) Cell-line / tissue, (3) treatment, (4) siRNA, (5) Antibodies / reagents and (6) Parameters of control or comments. The metadata of Wells is a subset of metadata of the plate level. For instance, in a 8x12 wells plate, different wells can have a subset of the parameters assigned to the whole plate. This level is also associated with the output of the HTS process (Raw Video / Images) and with the results of the image and data analysis phase which is also called measurements.

Part of the metadata at this level is critical information that should be verified and validated when it is uploaded. For instance, The parental cell line/tissue, or the treatment and its concentration are just two cases which the entry is verified in a first instance (obligatory data) and then they are validated with the information pre loaded in the imaging database. In order to keep the consistency of the metadata it is necessary to validate each entry and when a new value is detected the administrator of the platform is in charge of accepting this new entry as valid or correct to the right value if it is necessary. The consistency in the metadata is a key task in the imaging database because the obligatory data will be further used as a controlled vocabulary for querying.

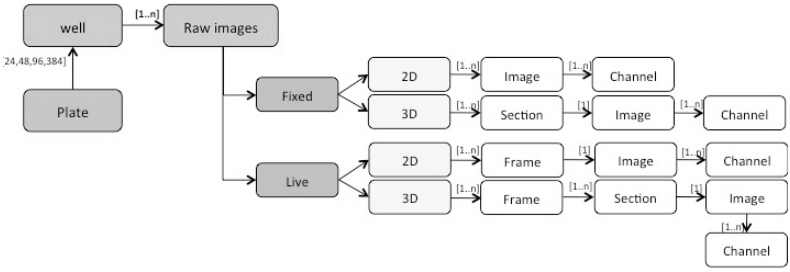


**Fig. 3.** Structure of the Raw Images metadata

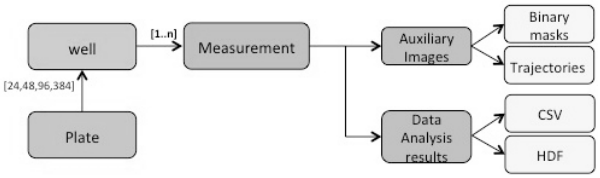


**Fig. 4.** Structure of the Measurements metadata

**Raw Images.** Raw images are obtained after image acquisition with automated microscopy systems. These images are the basis for the image analysis which results in quantitative data used for hypothesis testing. The response of the cells is recorded through time-lapse microscopy imaging and the resulting image sequences are the basis for the image analysis. The structure of an image file depends on the type of experiment (Fixed/Live) and the microscopy technique used in the experiment. Currently, four types of structures are supported (cf. Figure 3) [8]:

1. 2D (XY): this structure corresponds to one frame containing one image which is composed of multiple channels ([1]Frame - [1]Image - [1..n]Channels).
2. 2D+T (XY+T): this structure corresponds to one video with multiple frames. Each frame contains one image composed of multiple channels ([1]Video - [1..n]Frame - [1]Image - [1..n]Channels).
3. 3D (XYZ): this structure corresponds to one frame with multiple sections. Each section contains one image composed of multiple channels ([1]Frame - [1..n] Sections - [1]Image - [1..n]Channels).
4. 3D+T(XYZ+T):this structure corresponds to one video with multiple frames. Each frame can have multiple sections and each section contains one image composed of multiple channels ([1]Video - [1..n]Frame - [1..n]Sections - [1]Image - [1..n] Channels).

**Measurements.** This level contains the results of the Image and Data Analysis process (cf. Figure 4):

*Results of Image Analysis:* The results of image analysis are auxiliary images which are usually binary masks or trajectories. These images are results of the application of quality enhancing filters and segmentation algorithms employed to extract regions of

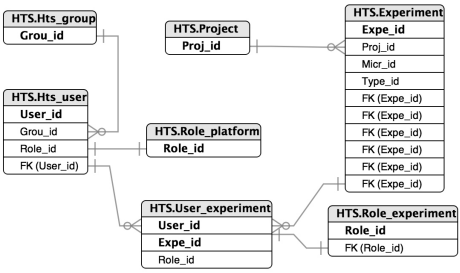


**Fig. 5.** Database schema for Project Metadata

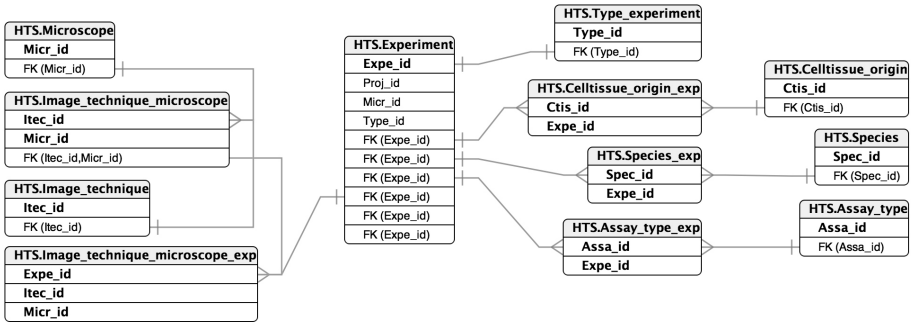


**Fig. 6.** Database schema for Experiment Metadata

interests (ROIs). These metadata are also linked to the raw video image file, on which the image analysis has been applied.

*Results of Data Analysis:*
Measurements extracted from the image analysis are further analyzed using pattern recognition tools. After applying operations such as feature selection, clustering and classification, a CSV file is generated with the results accompanied by a HDF file with information of the structure of the CSV file (features).

## 3 Data Modeling and Storage

The relational database schema designed to store the metadata in an HTS experiment is divided in 5 schema views according to the structure described in Section 2. Figure 5 shows the key components for the project metadata and the possibility to create groups and grant 4 different levels of access to our experiments (Author, Expert User, Analyst User and Guest).

Figure 6 describes the entity Experiment and how the metadata is stored according to the type of experiment performed, the microscope used and the image technique associated. Furthermore, other key components of metadata are mandatory for creating an experiment, such as the specie, assay type and cell/tissue origin.

The most critical part of the metadata corresponds to the Plate-Well metadata shown in figure 7. It requires a validation and verification process before registering new entries

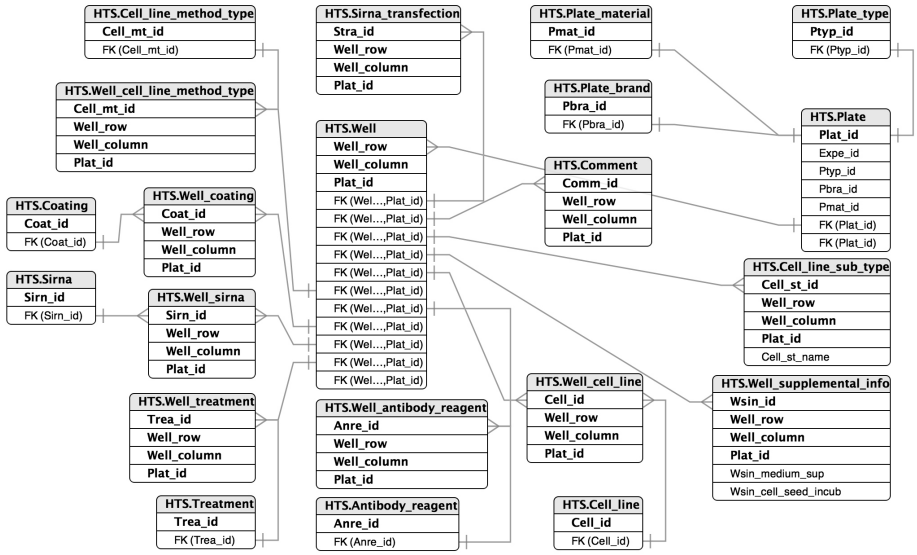


**Fig. 7.** Database schema for Plate-Well Metadata

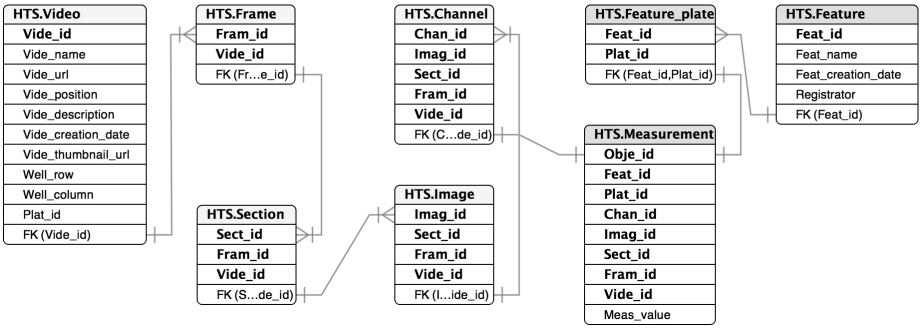


**Fig. 8.** Database schema for Raw Images and Measurement Metadata

to these entities. The author of an experiment uploads the metadata associated to a plate upon completion of the plate layout design and every entry is validated with the master entities for siRNAs, cell lines, antibodies, reagents, coatings and treatments in order to ensure consistency with the metadata being uploaded.

The main output in an HTS experiment are the raw images or time lapse sequence of images. The image dataset is located on a file server and the URLs to access them are stored in the database (cf. Figure 8). Those images are uploaded to the CytomicsDB through a web interface, which represents the web plate layout interface. During uploading time, using the open source Bio-Formats library [9], a new dataset of images (thumbnails) is generated, and also linked to the raw images in the database. These images are used to have a preview visualization of the plate layout in the web interface.

Another key component to consider in the database schema are the image and data analysis results. The information obtained after the image analysis process is parsed to the database using the entities Features and Measurements (cf. Figure 8). These two entities store the information required by another API such as PRTools [6] to perform the pattern recognition and statistical analysis.

## 4 Case Study in Endocytosis of EGFR: Identification of EGFR Dynamics and Classification of EGFR Phenotype Stages

In this section we describe a case study on how the structure of the metadata and RDBMS are applied in order to identify the EGFR dynamics and classify the different EGFR phenotypes.

Endocytosis is regarded as a mechanism of attenuating epidermal growth factor receptor (EGFR) signaling and of receptor degradation. Increasingly, evidence becomes available showing that cancer progression is associated with a defect in EGFR endocytosis [5]. Functional genomics technologies combine high-throughput RNA interference with automated fluorescence microscopy imaging and multi-parametric image analysis, thereby enabling detailed insight into complex biological processes, like EGFR endocytosis. The experiments produce over half a million images. Such a volume of images is beyond the capacity of manual processing and therefore, image processing and machine learning are required to provide an automated analysis solution for HTS experiments [2]. The total size in average can vary between 500 Mb to 20 Gb of raw images per experiment and CytomicsDB is designed to cope with the growing data size due to the scalable architecture for storing the images in a File Server and the metadata of the entire experiment in the database.

According to the methodology described in [2], three stages are identified: (1) Image Acquisition, (2) Image Analysis and (3) Data Analysis. We describe each stage as follows:

*Image Acquisition:* The experiment "Endocytosis of EGFR" is created in the CytomicsDB platform and its respectively plates. The type of metadata required for creating an experiment and the plates in our platform is described in Section 2. The respective values associated to each type of metadata have been detailed in [2]. After designing the plate in the platform, the wet-lab experiment is initiated, which includes the following steps: (1) cell culturing, siRNA transfection and EGF exposure, (2) fluorescent staining of proteins of interest and (3) image acquisition. Upon completion of the acquisition process 960 images are uploaded to the platform which size in total is 767 Mbytes. These images correspond to a 96 wells plate (cf. Figure 9) and for each well, images are captured from ten randomly selected locations. However, an experiment can consist of more than one plate and the number of samples per well can differ per case.

*Image Analysis:* The API in charge of the image analysis, request from the database the location of each image to process. The query executed is:

```
SELECT v.vide_id, v.vide_name, v.vide_url, v.vide_position, v.well_row, v.well_column
FROM HTS.Video v
WHERE v.plat_id = 17;
```

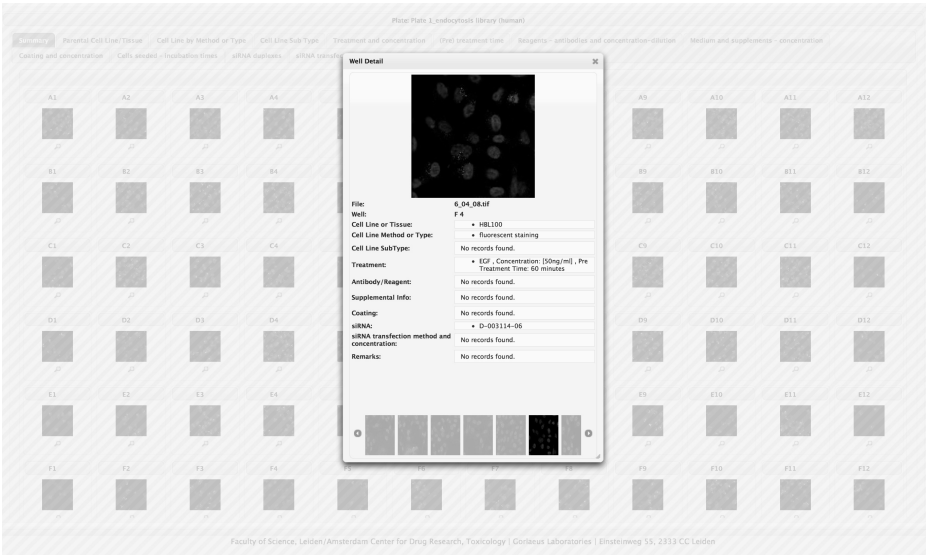


**Fig. 9.** Web plate layout

The value of column *plat_id* is in this case 17 and it was assigned after selecting *the plate for endocytosis* in the web interface. Three steps are performed by this API: (1) noise suppression, (2) image segmentation and (3) phenotype measurement. The algorithms and process details are described in [2]. Upon completion of the image analysis process, the API returns two outputs: (1) The location in the database of a new set of images and (2) a CSV file containing the features and the phenotype measurement respectively. The set of images generated are: (a) Original image: PERK (red), EFGR (green) and nucleus (blue) (cf. Figure 10), (b) Component definition: artificial cell border (red) and binary mask of protein expression (green) (cf. Figure 11), (c) Cell border reconstruction: artificial cell border (W-V) (cf. Figure 12), (d) Image segmentation: binary mask of EFGR channel by WMC (cf. Figure 13) [13].

The phenotype measurements (CSV file) are parsed first and then stored in the database by a web service executing the following query:

```
INSERT INTO HTS.Measurement
(Obje_id, Feat_id, Plat_id, Chan_id, Imag_id, Sect_id, Fram_id, Vide_id)
VALUES (0,1,17,1,1,1,1,1,14.0);
```

In this example, the column *Feat_id=1* corresponds to *Area* in the entity Feature and the measurement obtained for this feature is 14.0. The column *Plat_id* is still 17 because we refer to the same plate.

The measurements are categorized in two subgroups: (1) basic measurements of the phenotypes covering shape descriptors and (2) the localization phenotype describing the assessment of the correlation between two information channels. The basic phenotype measurement includes a series of shape parameters such as: size, perimeter, extension,

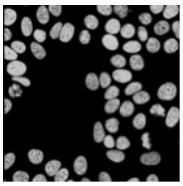

**Fig. 10.** Original image

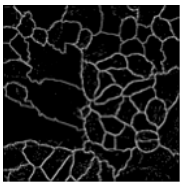

**Fig. 11.** Component definition

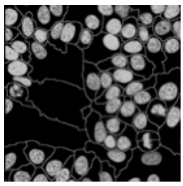

**Fig. 12.** Cell border reconstruction

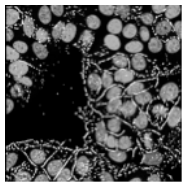

**Fig. 13.** Image segmentation

dispersion, elongation, orientation, intensity, circularity, semi-major axis length, semi-minor axis length, closest object distance and in nucleus, these can be extended as the experiments so dictates. In addition to the basic phenotype measurement, localization measurements can be derived for a specific experimental hypothesis. The localization phenotypes are quantifications of comparative measurement between information channels such as relative structure-to-nucleus distance or structure-to-border distance. The features in EGFR-screen based localization phenotypes used are: nucleus distance, border distance and intactness. On the basis of the phenotype measurements, objects are classified into phenotypic stages. For the assessment of significance statistical analysis is performed [2]. Upon completion of the image analysis, it is possible to visualize the results in a web plate layout and export the measurements to files.

*Data Analysis:* The aim of the endocytosis study is to quantify the process of EGF-induced EGFR endocytosis in human breast cells and to identify proteins that may regulate this process. The EGFR endocytosis process can roughly be divided into three characteristic episodes: i.e. (1) at the onset EGFR is present at the plasma-membrane; (2) subsequently, small vesicles containing EGFR will be formed and transported from the plasma-membrane into the cytoplasm; and (3) finally, vesicles are gradually merging near the nuclear region forming larger structures or clusters. The characteristic episodes are the read-out for HTS. Based on this model it is believed that EGFR endocytosis regulators may be potential drug targets for EGFR-induced breast cancer. Studying each of the stages, i.e. plasma-membrane, vesicle and cluster, may provide a deeper understanding of the EGFR endocytosis process [2].

When the data analysis process is triggered, a web service request to the database (entities feature and measurement) the results from the image analysis process. The output of this web service is the location of a file which contains the results of the test for each siRNA regulator. This file will be requested for the API PRTools for generating classifications and graphs with the comparison of the results, such as: (1) Weighted classification error curve, which represents a combination of a feature selection/extraction method and a classifier algorithm, (2) Results of the feature extraction and (3) Average number of plasma-membrane (a) and vesicle (b) per nucleus [2]. Consolidating in CytomicsDB the experiment's metadata, raw images and images/data analysis results, facilitates further comparison with the result of other HTS experiments.

## 5 Related Work

In the current area of -omics research, various systems/tools have emerged to try to solve the problem that the existing practice of keeping meta data does not allow for effective

data searching and mining. They are generally referred to as Laboratory Information Management System (LIMS).

The work proposed by Colmsee et. al. [4] is probably the closest to CytomicsDB. The authors defined central requirements for a primary lab data management and aspects of best practices to realise those requirements. As a proof of concept, the authors implemented a pipeline to manage primary lab data of crop plants. The pipeline consists of i) data storages including a Hierarchical Storage Management system, an RDBMS and a BFiler package to store primary lab data and their meta information; ii) the Virtual Private Database for the realisation of data security and the LIMS Light application to iii) upload and iv) retrieve stored primary lab data. Compared with this work, CytomicsDB has a more sophisticated data model to cope with different types of data (i.e., images, videos, and data produced in different steps in an HTS experiment), pays special attention to the extensibility of the architecture to enable adding new tools.

In [11], the authors presented three open-source, platform independent software tools for genomic data: a next generation sequencing / microarray LIMS and analysis project center (GNomEx); an application for annotating and programmatically distributing genomic data using the DAS/2 data exchange protocol (GenoPub); and a standalone Java Swing application (GWrap) that provides a GUI for the command line analysis tools. CytomicsDB provides similar functionalities as these tools, but focuses on dealing with Cytomic data. Moreover, for the design of CytomicsDB, we have deliberately chosen for a single integrated system to include all features required for conduction HTS experiments and analysis, instead of individual tools and enabling high profile pattern recognition.

In [12], the authors describe a general modeling framework for laboratory data. The model utilises several abstraction techniques, with focus on the concepts of inheritance and meta-data. In this model, distinct regular entity and event schemas can be defined and fully integrated via a standardized interface. The design allows definition of a processing pipeline as a sequence of events. A layer above the event-oriented schema integrates events into a workflow by defining processing directives, which act as automated project managers of items in the system. This LIMS is built on the Oracle RDBMS, and is maintained by multiple database administrators (DBAs). While with CytomicsDB, our goal is to meet the needs of HTS experiments with a more light-weight, flexible system. By adapting modern web and database technologies, CytomicsDB is easy to maintain (i.e., no DBAs required) and extend (i.e., allowing integrating new tools naturally).

The work by Chan et al. [3] focuses on interactive visualization methods for data generated by HTS experiments. The visualization methods might be adapted by CytomicsDB. However, CytomicsDB is a much more comprehensive information system for HTS data, because it integrates both experiments and analysis data into a single system, and allows various types of users and groups to be defined.

Based on the Golm Plant Database System, Köhl et. al. [7] devised a data management system based on a classical LIMS combined with web-based user interfaces for data entry and retrieval to collect this information in an academic environment. This system stores plant cultivation units in an MS ACCESS database, which would quickly run into scalability issues as the data size grows.

## 6 Conclusions and Future Work

In this paper, we have presented a semantic approach for organizing metadata and an RDBMS for metadata management in High-Throughput Screening experiments. Our goal is to facilitate the exploration process in the HTS workflow, scientist are aware of semantics and they are pushing forward the need for new approaches in organizing the metadata according to which queries are mostly applied on the data. In HTS, images by itself do not have any meaning, but linking images to their respective metadata allows researchers to learn from their experience and help them in mentalizing semantic structures of the metadata. The RDBMS schema has been designed to support the acquisition, visualization and integration stages using a metadata-based approach. Furthermore, CytomicsDB uses a database engine suitable for applications which demands intensive data mining tasks. Finally, we plan to extend this architecture to a more sophisticated interdisciplinary platform for cytomics. The structure of the metadata proposed in this paper will further evolve to an ontology based framework. A new layer to the architecture will be added in order to perform semantic queries, turning the architecture to a web based interactive semantic platform for cytomics [1].

**Acknowledgements.** We thank Dr Marjo De Graauw for providing us the data of the experiment included in our case study. This work is partially supported by the Erasmus BAPE program, Cyttron II project (EL), the European FET Flagship Programme the Human Brain Project (`www.humanbrainproject.eu/`) and the Dutch national project COMMIT (`www.commit-nl.nl/`).

## References

1. Bertens, L.M.F., Slob, J., Verbeek, F.: A generic organ based ontology system, applied to vertebrate heart anatomy, development and physiology. J. Integrative Bioinformatics 8(2) (2011)
2. Cao, L., Yan, K., Winkel, L., de Graauw, M., Verbeek, F.J.: Pattern recogntion in high-content cytomics screens for target discovery - case studies in endocytosis. In: Loog, M., Wessels, L., Reinders, M.J.T., de Ridder, D. (eds.) PRIB 2011. LNCS, vol. 7036, pp. 330–342. Springer, Heidelberg (2011)
3. Chan, T., Malik, P., Singh, R.: An interactive visualization-based approach for high throughput screening information management in drug discovery. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 5794–5797 (August 2006)
4. Colmsee, C., Flemming, S., Klapperstuck, M., Lange, M., Scholz, U.: A case study for efficient management of high throughput primary lab data. BMC Research Notes 4(1), 413 (2011)
5. de Graauw, M., Cao, L., Winkel, L., van Miltenburg, M.H.A.M., LeDévédec, S., Klop, M., Yan, K., Pont, C., Rogkoti, V.-M., Tijsma, A., Chaudhuri, A., Lalai, R., Price, L., Verbeek, F., van de Water, B.: Annexin a2 depletion delays egfr endocytic trafficking via cofilin activation and enhances egfr signaling and metastasis formation. In: Oncogene (2013)
6. Duin, R.P.W.: Prtools - version 3.0 - a matlab toolbox for pattern recognition. In: Proc. of SPIE, p. 1331 (2000)
7. Kohl, K., Basler, G., Ludemann, A., Selbig, J., Walther, D.: A plant resource and experiment management system based on the golm plant database as a basic tool for omics research. Plant Methods 4(1), 11 (2008)

8. Larios, E., Zhang, Y., Yan, K., Di, Z., LeDévédec, S., Groffen, F., Verbeek, F.J.: Automation in cytomics: A modern RDBMS based platform for image analysis and management in high-throughput screening experiments. In: He, J., Liu, X., Krupinski, E.A., Xu, G. (eds.) HIS 2012. LNCS, vol. 7231, pp. 76–87. Springer, Heidelberg (2012)
9. Linkert, M., Rueden, C.T., Allan, C., Burel, J., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K.W., Swedlow, J.R.: Metadata matters: access to image data in the real world. The Journal of Cell Biology 189, 1 (2010)
10. Mayr, L., Fuerst, P.: The future of high-throughput screening. Journal of Biomolecular Screening (2008)
11. Nix, D., Sera, T.D., Dalley, B., Milash, B., Cundick, R., Quinn, K., Courdy, S.: Next generation tools for genomic data generation, distribution, and visualization. BMC Bioinformatics 11(1), 455 (2010)
12. Wendl, M., Smith, S., Pohl, C., Dooling, D., Chinwalla, A., Crouse, K., Hepler, T., Leong, S., Carmichael, L., Nhan, M., Oberkfell, B., Mardis, E., Hillier, L., Wilson, R.: Design and implementation of a generalized laboratory data model. BMC Bioinformatics 8(1), 362 (2007)
13. Yan, K., Verbeek, F.J.: Segmentation for high-throughput image analysis: Watershed masked clustering. In: Margaria, T., Steffen, B. (eds.) ISoLA 2012, Part II. LNCS, vol. 7610, pp. 25–41. Springer, Heidelberg (2012)