BMC Genomics

BioMed Central

Software

Open Access

MicroArray Facility: a laboratory information management system with extended support for Nylon based technologies

Paul Honoré†[1], Samuel Granjeaud†[2], Rebecca Tagett[1], Stéphane Deraco[1,5], Emmanuel Beaudoing[2,3], Jacques Rougemont[2,3], Stéphane Debono†[1] and Pascal Hingamp*†[2,4]

Address: [1]IPSOGEN SAS, Luminy Biotech Entreprises, 163 avenue de Luminy, Case 923, 13009 Marseille, France, [2]TAGC, INSERM ERM206, Parc Scientifique de Luminy, Case 928, 13288 Marseille Cedex 09, France, [3]Now at Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland, [4]Now at IGS, CNRS UPR 2589, 163 Avenue de Luminy Case 934, 13288 Marseille Cedex 09, France and [5]Now at CNRS – DSI, Tour Gaïa, rue Pierre-Gilles de Gennes, BP 21902, 31319 LABEGE CEDEX, France

Email: Paul Honoré - honore@ipsogen.com; Samuel Granjeaud - granjeaud@tagc.univ-mrs.fr; Rebecca Tagett - tagett@ipsogen.com; Stéphane Deraco - Stephane.Deraco@dsi.cnrs.fr; Emmanuel Beaudoing - emmanuel.beaudoing@isb-sib.ch; Jacques Rougemont - Jacques.Rougemont@isb-sib.ch; Stéphane Debono - debono@ipsogen.com; Pascal Hingamp* - pascal.hingamp@igs.cnrs-mrs.fr

* Corresponding author   †Equal contributors

Published: 20 September 2006

BMC Genomics 2006, **7**:240   doi:10.1186/1471-2164-7-240

Received: 29 March 2006
Accepted: 20 September 2006

This article is available from: http://www.biomedcentral.com/1471-2164/7/240

Abstract

Background: High throughput gene expression profiling (GEP) is becoming a routine technique in life science laboratories. With experimental designs that repeatedly span thousands of genes and hundreds of samples, relying on a dedicated database infrastructure is no longer an option.

GEP technology is a fast moving target, with new approaches constantly broadening the field diversity. This technology heterogeneity, compounded by the informatics complexity of GEP databases, means that software developments have so far focused on mainstream techniques, leaving less typical yet established techniques such as Nylon microarrays at best partially supported.

Results: MAF (MicroArray Facility) is the laboratory database system we have developed for managing the design, production and hybridization of spotted microarrays. Although it can support the widely used glass microarrays and oligo-chips, MAF was designed with the specific idiosyncrasies of Nylon based microarrays in mind. Notably single channel radioactive probes, microarray stripping and reuse, vector control hybridizations and spike-in controls are all natively supported by the software suite. MicroArray Facility is MIAME supportive and dynamically provides feedback on missing annotations to help users estimate effective MIAME compliance. Genomic data such as clone identifiers and gene symbols are also directly annotated by MAF software using standard public resources. The MAGE-ML data format is implemented for full data export. Journalized database operations (audit tracking), data anonymization, material traceability and user/project level confidentiality policies are also managed by MAF.

Conclusion: MicroArray Facility is a complete data management system for microarray producers and end-users. Particular care has been devoted to adequately model Nylon based microarrays. The MAF system, developed and implemented in both private and academic environments, has proved a robust solution for shared facilities and industry service providers alike.

Background

Transcriptome surveying using microarrays has become an established and widespread technique[1]. Although glass based microarrays and oligonucleotide chips are common in the gene expression profiling (GEP) landscape, Nylon supported microarrays coupled with radioactive detection, either home made [2-9] or from industrial suppliers [10-14], are an alternative still favored by some researchers[15], including the NIH's National Institute on Aging [16-19]. The resilience of this platform can be explained by the easy setup of this technical combination, its high sensitivity achieved without target amplification[20], and its cost effectiveness[16]. Technological development for this platform is ongoing, as demonstrated by a two 'isotope' dual channel variant that uses real time emission integration[21].

Regular increase in microarray reporter densities together with falls in unitary costs have meant experiments routinely generate tens of millions of data pieces to store, search and analyze[22]. For all but occasional microarray users, dedicated laboratory information management systems (LIMS) are now a requirement. Arguably one reason bench scientists are yielding to laboratory databases is increasing pressure from journal editors to have data appropriately submitted to international repositories as a prerequisite for publication[23].

Amongst the gene expression profiling LIMS that have been reported [24-34], glass dual channel Cy5/Cy3 and oligonucleotide chips are extremely well catered for. However, Nylon based microarrays are at best marginally supported by these software, with more substantial Nylon support only to be found in commercial products[35]. Drawing on experience with a previous LIMS[36], the specific functionalities that we have found critical for comprehensive Nylon based microarrays modeling include the ability to record so called 'vector' hybridizations (quantitations of spotted reporter amounts), the stripping and recycling of microarrays[37] and to model controls spiked into RNA samples [20,38]. Finally, the pharmaceutical industry manufacturing and regulatory rigor required for the development of diagnostic applications are typically not of major concern in freely available LIMS.

Here we report the development of MAF (MicroArray Facility), a LIMS designed to accommodate the following desiderata:

-stringent quality control, traceability and audit tracking to meet industrial requirements

-multi-platform support (i.e. Nylon, glass, and oligo-based microarrays)

-rich data annotation for MIAME standard compatibility[39]

-support for oncogenomic projects (i.e. clinical data)

-dynamic cDNA reporter annotation using public data banks (notably UNIGENE[40])

-MAGE-ML enabled[41]

-import robotic equipment files in native formats to reduce error-prone reformatting

-multi-user privilege environment to promote data sharing and ensure confidentiality.

The scope of MAF (as defined in [42]) is a local LIMS serving a community of researchers, such as found around academic shared facilities or commercial array and service providers. The MAF user interface is entirely web browser contained, thus there is no technical contra-indication to using MAF as a means to publish datasets over the Internet. However one-stop shops for gene expression data have innumerable advantages for public data mining[43]; thus we strongly recommend uploading public data to international archives such as ArrayExpress[44], CIBEX[45] or GEO[22] as a mechanism for publishing data.

MAF is a LIMS *sensu stricto* in that it records, tracks, structures, searches and reports all information required to establish gene expression profiles. High level downstream data analyses are carried out by exporting selected data to any of the myriad of dedicated analysis packages such as Cluster[46], BioConductor[47], MeV[26] or ProfileSoftware[48].

Implementation

MAF follows a client-server architecture implemented as a web-based application, allowing simultaneous multi-user access to a central database. Client browsers connect to an Apache server in a Unix/Linux environment. The application is entirely written in Perl; since the Perl packages rely on the abstract DBI module, switching between different Relational Database Management systems (RDBMS) is as simple as changing a single line in the MAF configuration file. MAF has been implemented, tested and validated for the Oracle 8i and the PostgreSQL RDBMS. With Oracle 8i, certain Oracle-specific features such as backup tools and transportable table spaces can also be used. Having deliberately avoided RDBMS specific SQL syntax, we believe that MAF is easily portable to other SQL database platforms that have a Perl DBD driver.

The relational database underlying MAF – composed of 2258 fields held in 215 tables – is called ELOGE. Its schema extends the conceptual ArrayExpress design[44] which integrates microarray design and manufacturing, sample description, hybridization and data acquisition. Through a 5-year development cycle, ELOGE was considerably expanded to integrate fine grain modelling of wet lab routine procedures such as plate management, PCR quality control annotations, sequencing results, analytic validation, and GLP (Good Laboratory Practices) compliant protocols. The database scheme is designed to avoid computation-intensive queries and optimize user interface responsiveness.

MAF is accompanied by two complementary software modules (Fig. 1): the Gene Finder (GF) and Clone Chooser (CC) which respectively manage gene and clone lists, as well as provide effective mapping from one to the other. Genomic data relevant to all three modules are imported into ELOGE every two months from several commonly used public databases (GENBANK EST, UNIGENE, SWISSPROT, ENTREZ GENE, GO and REFSEQ), allowing automatic and thorough annotation of genes and cDNA clones. Together, the MAF, GF and CC modules constitute a package called Discovery Software.

MAF user data are collected through web form cascades (Fig. 2). Where appropriate (e.g. array layouts or image quantitations) data files are uploaded using background queue processing to avoid tying-up the interface. Many instrument and third party software data files are thus directly imported using a large set of data formatting drivers (Table 1). MAF web forms can be used to add, update or view data. However, browsing MAF data is best achieved through a collection of hyper linked data reports specific for each step of GEP processing (Table 2).

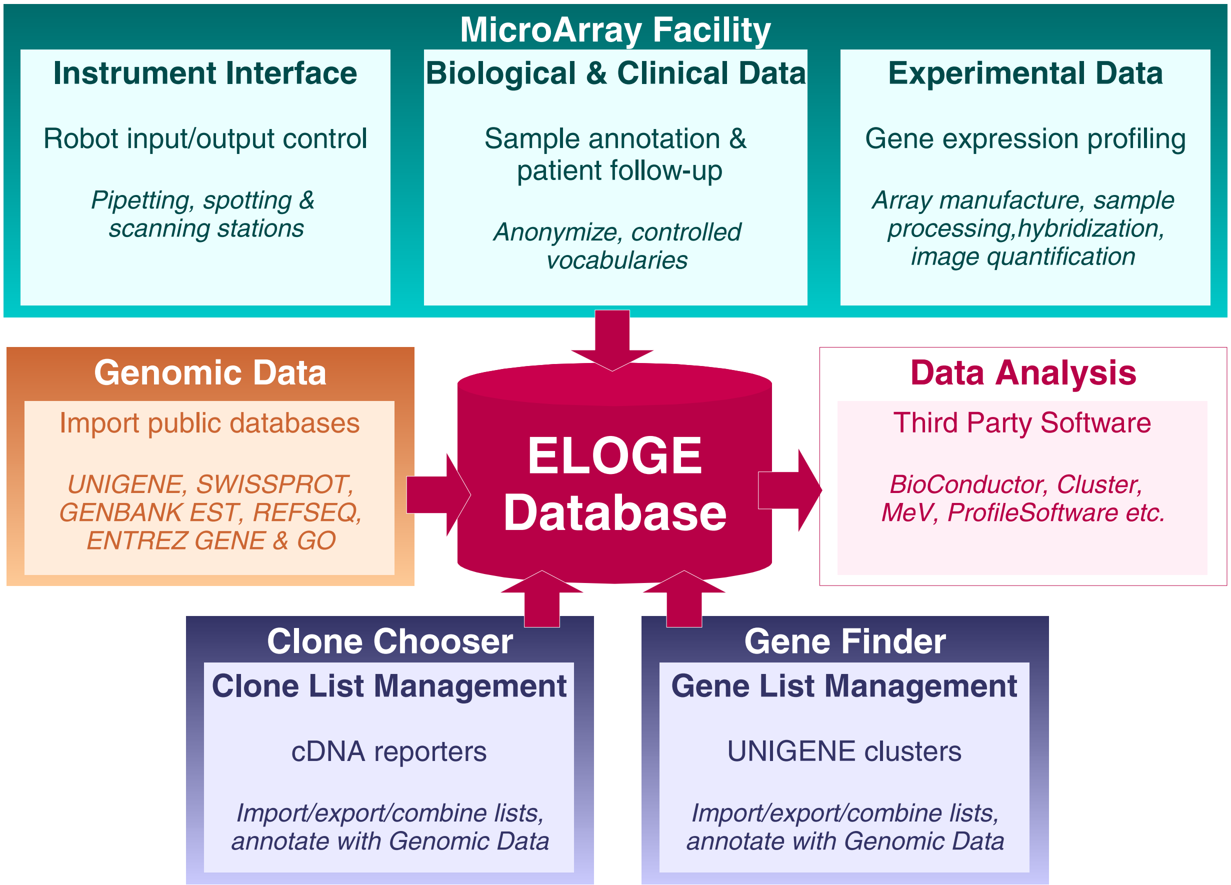
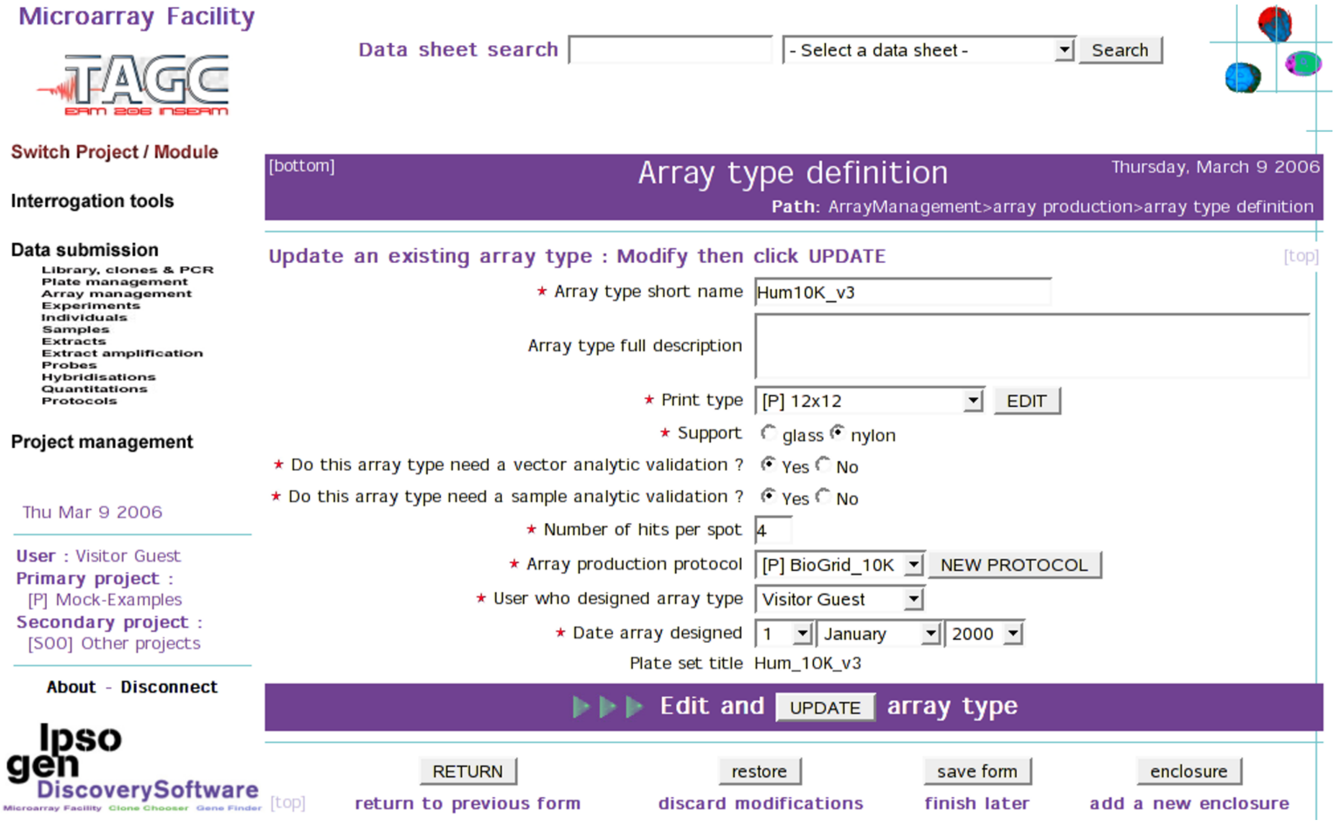

MicroArray Facility
Instrument Interface
Robot input/output control
Pipetting, spotting & scanning stations
Biological & Clinical Data
Sample annotation & patient follow-up
Anonymize, controlled vocabularies
Experimental Data
Gene expression profiling
Array manufacture, sample processing,hybridization, image quantification
Genomic Data
Import public databases
UNIGENE, SWISSPROT, GENBANK EST, REFSEQ, ENTREZ GENE & GO
ELOGE Database
Data Analysis
Third Party Software
BioConductor, Cluster, MeV, ProfileSoftware etc.
Clone Chooser
Clone List Management
cDNA reporters
Import/export/combine lists, annotate with Genomic Data
Gene Finder
Gene List Management
UNIGENE clusters
Import/export/combine lists, annotate with Genomic Data

Figure 1
The Laboratory Information Management System (LIMS) is articulated around a relational database (ELOGE). Users manage gene lists, clone lists and microarray experiments through three software modules (Gene Finder, Clone Chooser and Microarray Facility respectively). Reporter annotation is achieved by importing genomic context data from public databases. Stored data can be queried and exported for analysis by appropriate third party software.



Data sheet search [ ] - Select a data sheet - ▾ Search

Thu Mar 9 2006

User : Visitor Guest
Primary project : [P] Mock-Examples
Secondary project : [S00] Other projects

About - Disconnect

[bottom] Array type definition Thursday, March 9 2006

Path: ArrayManagement>array production>array type definition

Update an existing array type : Modify then click UPDATE [top]

★ Array type short name	Hum10K_v3
Array type full description	
★ Print type	[P] 12x12 ▾ EDIT
★ Support	○ glass ● nylon
★ Do this array type need a vector analytic validation ?	● Yes ○ No
★ Do this array type need a sample analytic validation ?	● Yes ○ No
★ Number of hits per spot	4
★ Array production protocol	[P] BioGrid_10K ▾ NEW PROTOCOL
★ User who designed array type	Visitor Guest ▾
★ Date array designed	1 ▾ January ▾ 2000 ▾
Plate set title	Hum_10K_v3

▶▶▶ Edit and UPDATE array type

RETURN return to previous form
restore discard modifications
save form finish later
enclosure add a new enclosure

Figure 2
User data is collected through dedicated web form cascades. This example shows the Array Type definition form, which includes a "vector analytic validation" field specific to Nylon filters.

Confidentiality is ensured through user login/password authentication and project containers. Access to a project's objects (arrays, plates, hybridizations, quantitations etc.) is controlled by the project owner through read, write, update, and delete independent privileges. Unique transaction ID's for each request, inactivity time-outs, enforced single user sessions and SSL network encryption complete the MAF security strategy.

Results and discussion

Clone library management

MAF clone management models the laboratory procedures for clone handling through wells, plates and plate sets. The LIMS precisely tracks clones from the resource plate to the array through replication, reorganisation, amplification and spotting steps.

Since plate handling is a cornerstone of GEP experimental work, much effort has been devoted to MAF's ergonomics and the minimization of manual data entry. For instance figure 3 shows the "PCR run" form used to enter quality control data from gel migrations of PCR amplifications. This synoptic side by side visualization of the gel picture with the colour coded annotated plate has allowed systematic verification of at least 10% of the plates before every plate set manipulation, without adversely affecting efficiency. Following this PCR quality control annotation, MAF can not only directly produce a reorganisation work list for the plate handling workstation (e.g. Tecan), but also verify and validate the reorganisation by comparing its original work-list with the trace file summarizing the work actually carried out by the work station.

Another example of MAF's routine quality controls is the update of sequence verified clone identities by monthly BLAST analysis and checking of clone to UNIGENE cluster associations.

Microarray production

MAF manages every step of array production from abstract print type and array type definitions to batch production runs of physical arrays. A custom produced array design can be created by directly uploading layout definition files from spotting robots (e.g. Microgrid II or GMS), hence avoiding error-prone manual entry of the array design. Bypassing printing steps, microarrays or even oligo-chips

Table 1: Equipment and third party software which are directly compatible with Microarray Facility.

Operations	Software/Workstation	File format
Plate handling	Tecan	Text
Arraying	Biogrid	MAF web form
	GeneMachines	Text
	Affymetrix 417 (ex GMS 417)	Text
	Microgrid II (Genomic Solutions)	CSV
	Generic	Tabulated text
Image acquisition	Raytest TINA (agarose gels)	PCB + BMP
	Generic	BMP/TIF/IMG+INF
Quantitation	ArrayGauge 2.1 (Fuji)	Text
	BZScan 1.0 (INSERM)	Text
	GenePix 3.0 (Axon)	Text
	Imagene 4.0 (BioDiscovery)	Text
	ProfileSoftware Corporate/Cancer (Ipsogen)	Text
	Generic	Tabulated text
Normalization	ProfileSoftware Corporate (Ipsogen)	Text
	Generic	Tabulated text
Enclosures	Generic	Any (PDF, Text, DOC etc.)

Interfacing is through upload of output files in their native formats (except for the Biogrid arrayer which is modeled through a web form). Enclosures allow users to attach arbitrary files to any of the MicroArray Facility objects (such as spreadsheet results, publications, images etc.).

Table 2: List of objects that can be stored and queried from the MicroArray Facility web interface.

Clone annotation
Clones
Genes
Library
Clone sequencing
Clone history
PCR primer
PCR run
Purified PCR run
Plate management
Plate
Plateset
Plate history
Plateset history
Worklist
Freezer
Array management
Array
Array batch
Array analytic validation
Array batch analytic validation
Strip
Unstock
Production Follow-Up

Vector hybridization
Probe
Spike
Oligo
Hybridisation
Quantification result
Extracts
Sample
Individual organism
Parameter
Extract
Extract amplification
Sample hybridization
Experiment
Probe
Spike
Hybridisation
Quantification result
Project
Project
User
Process
Enclosure
Protocol
Protocol

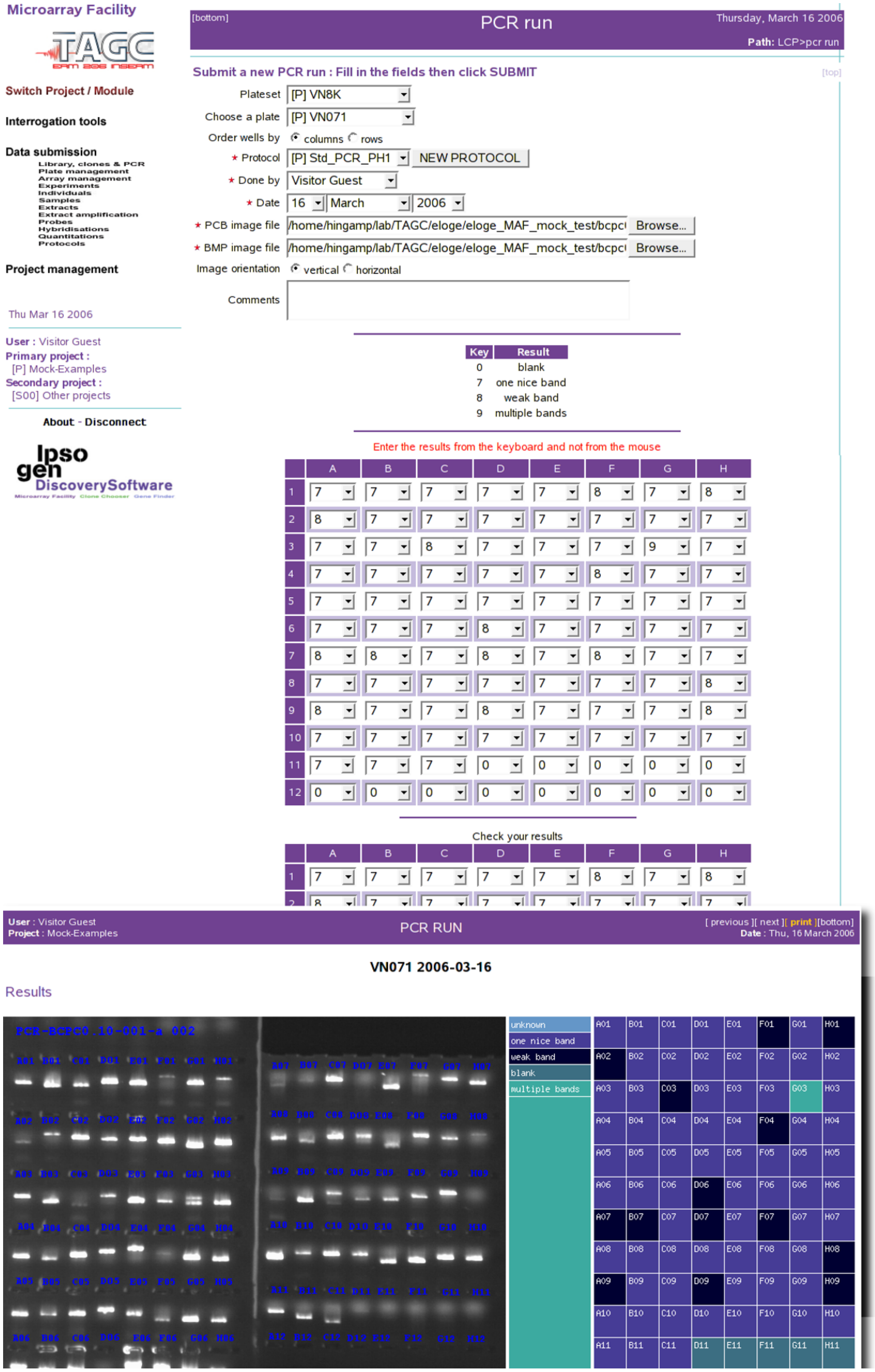
from third party providers can also be loaded into MAF, albeit with less detailed annotation at the array design level.

In the case of Nylon based arrays, MAF manages the "vector probe hybridisation" post-production quality control. This step measures the quantity of reporter material bound at each spot through batch hybridisation of filters with a labelled oligonucleotide (the sequence of which is common to every spotted clone), followed by filter stripping.

Every array produced can be tracked down to the projects and hybridizations in which it has been used.

Expression profiling data user submission

MAF data is partitioned into projects containing one or more experiments. An experiment ties together any number of hybridizations, usually all undertaken as part of the same experimental design. Experimental data submission follows the flow leading from biological samples to RNA extraction, labelling, hybridization, scanning, image feature extraction, and finally normalisation of measurements.



Thu Mar 16 2006

User : Visitor Guest
Primary project :
[P] Mock-Examples
Secondary project :
[S00] Other projects

PCR run

Thursday, March 16 2006
Path: LCP>pcr run

Submit a new PCR run : Fill in the fields then click SUBMIT

Plateset [P] VN8K
Choose a plate [P] VN071
Order wells by ● columns ○ rows
* Protocol [P] Std_PCR_PH1
* Done by Visitor Guest
* Date 16 March 2006
* PCB image file /home/hingamp/lab/TAGC/eloge/eloge_MAF_mock_test/bcpc
* BMP image file /home/hingamp/lab/TAGC/eloge/eloge_MAF_mock_test/bcpc
Image orientation ● vertical ○ horizontal
Comments

Key	Result
0	blank
7	one nice band
8	weak band
9	multiple bands

Enter the results from the keyboard and not from the mouse

	A	B	C	D	E	F	G	H
1	7	7	7	7	7	8	7	8
2	8	7	7	7	7	7	7	7
3	7	7	8	7	7	7	9	7
4	7	7	7	7	7	8	7	7
5	7	7	7	7	7	7	7	7
6	7	7	7	8	7	7	7	7
7	8	8	7	8	7	8	7	7
8	7	7	7	7	7	7	7	8
9	8	7	7	8	7	7	7	8
10	7	7	7	7	7	7	7	7
11	7	7	7	0	0	0	0	0
12	0	0	0	0	0	0	0	0

Check your results

	A	B	C	D	E	F	G	H
1	7	7	7	7	7	8	7	8
2	8	7	7	7	7	7	7	7

User : Visitor Guest
Project : Mock-Examples

PCR RUN

[previous][next][print][bottom]
Date : Thu, 16 March 2006

VN071 2006-03-16

Results

PCR-BCPC0.10-001-a 002
A01 B01 C01 D01 E01 F01 G01 H01
A07 B07 C07 D07 E07 F07 G07 H07

unknown
one nice band
weak band
blank
multiple bands

A01	B01	C01	D01	E01	F01	G01	H01
A02	B02	C02	D02	E02	F02	G02	H02
A03	B03	C03	D03	E03	F03	G03	H03
A04	B04	C04	D04	E04	F04	G04	H04
A05	B05	C05	D05	E05	F05	G05	H05
A06	B06	C06	D06	E06	F06	G06	H06
A07	B07	C07	D07	E07	F07	G07	H07
A08	B08	C08	D08	E08	F08	G08	H08
A09	B09	C09	D09	E09	F09	G09	H09
A10	B10	C10	D10	E10	F10	G10	H10
A11	B11	C11	D11	E11	F11	G11	H11

Figure 3

PCR run form used to verify cDNA clone amplifications on agarose gels. Double data submission (second table) is part of MicroArray Facility's quality control measures. Band readings are entered through keyboard strokes for better ergonomy and increased throughput.

Feature extraction (image quantitation) results can be imported in a number of common software formats such as BZscan[49], ArrayGauge[50], Genepix[51], Imagene[52], and ProfileSoftware[48]. Where a format is not supported MAF provides a simple generic tab delimited format that can easily be produced with a spreadsheet program.

External controls added to labelled samples, such as RNA spike-ins which are commonly used with Nylon filters, are also quantitatively represented by MAF for a more accurate assessment of quality control and for spike based data set normalization.

MAF representation of biological samples links individuals, samples, RNA extracts and labelled extracts, each with a many-to-many cardinality. Rich quantitative and qualitative annotation of both individuals and samples is supported (e.g. code, age, mass, sex, tumor grade etc.), using either standard nomenclatures (such as oncology terms) or user defined parameters. This controlled vocabulary annotation of samples is of paramount importance for effective downstream correlation of expression profiles with experimental, biological and clinical factors.

Data annotation and MIAME compliance

Results from high throughput gene expression profiling experiments differ from single gene measurements in that the effects of many more experimental parameters are likely to be observed. The proper correlation of expression signatures with biological parameters therefore requires careful recording of all known experimental variables.

This long recognized specificity of transcriptome analysis has led the Microarray Gene Expression Data Society (MGED[53]) to draw up MIAME, a set of minimal annotation guidelines for microarray based experiments[39]. All three international gene expression archives support MIAME standard data annotation, and an increasing number of scientific editors are requiring MIAME grade data for publication in their journals[23].

Thus MIAME compatibility has been a design ambition for the development of MicroArray Facility since its inception. This has directly impacted the underlying MAF data scheme as can still be seen in the [individual > sample > extract > labeled extract] part of the model. Defined name spaces for MIAME annotation are reserved in all relevant parts of the database relations. The annotation is either collected through web forms from the user (e.g. hybridization protocol), or generated automatically by MAF using imported public data (e.g. gene symbols from ENTREZ GENE or SWISSPROT).

Pivotal in MIAME is the requirement to attach laboratory protocols for all experimental steps. MAF has 15 protocol categories which are user supplied documents (e.g. text or PDF files) supplemented with optional or obligatory variable parameters, e.g. exposure time for "Image Acquisition" protocols.

Since software can only be 'MIAME supportive', i.e. potentially able to store the required annotation, MAF provides a MIAME check-list to help researchers make sure their data is actually MIAME compliant. The check list, accessible at any time, reports any missing annotation, in particular required protocols currently undefined in the project.

Data export and interoperability

Data collected through the forms and their associated annotations can be viewed at any time through specific web reports launched from the permanent search box in the form header. All reports are dynamically hyper linked facilitating navigation across object categories. Displayed data can also be directly downloaded as tabulated text files. A more substantial reporting tool is also provided for more transversal data searches, such as finding all samples verifying specific criteria, for instance tumor grade or patient age. Search results are exported as classical flat file datasets including comprehensive sample and reporter annotations as well as the expression measurements. This text file format is compatible with most downstream analysis tools such as BioConductor, Cluster or ProfileSoftware [46-48].

MAF also currently provides an experimental MAGE-ML complete experiment export functionality comprising all MAGE packages (ArrayDesign, DesignElement, Experiment, BioAssay, BioMaterial, BioAssayData) suitable for exporting to data archives or to MAGE-ML enabled data analysis tools[47]. The MAF produced MAGE-ML has been validated and a test experiment was successfully pipelined into ArrayExpress.

Regulatory compliance

Regulatory agencies are currently working to define a proper regulatory environment for GEP use in drug development and market approval processes. A guidance on Pharmacogenomic Data Submissions was issued by the FDA in March 2005[54]. This document defines the rules to be followed in order to ensure that GEP data submitted for drug approval will have the quality level required by the FDA.

Compliance with FDA 21 CFR part 11 regulations is audited at least once a year. The production version of MAF is currently 80% compliant with 21 CFR part 11. Ongoing developments aim to reach full MAF compliance with Good Laboratory Practices (GLPs). Guidelines for

Table 3: Comparison of microarray LIMS with regard to licencing, specific technology support, regulatory compliance, MIAME compliance and Mage-ML export.

Name	Supplier	Specific Support for			License Type	Regulatory Compliance	MIAME	Mage-ML export
		Nylon filters	2-color slides	Oligo chips		21 CFR part 11		
ArrayDB 2.1	National Human Genome Research Institute (NHGRI)	*	✓	*	OS	*	*	*
BASE 1.2	Lund University	±	✓	✓	OS	*	✓	✓
GeneX Va/GEOSS 2.5.2	University of Virginia	*	±	✓	OS	*	✓	*
LIMaS 2	Medical Research Council (MRC) Harwell	*	✓	✓	RF	*	✓	✓
MADAM	The Institute for Genomic Research (TIGR)	±	✓	✓	OS	*	✓	✓
MADGE	University of Florida	*	✓	*	RF	*	*	*
MARS 1.1.1	Graz University of Technology	*	✓	*	RF	*	✓	✓
maxdLoad2	University of Manchester	*	✓	✓	OS	*	✓	✓
MicroArray Facility	Ipsogen/Institut National de la Santé et de la Recherche Médicale (INSERM)	✓	✓	±	RF	✓	✓	✓
MicroGen	Politecnico di Milano	*	✓	*	OS	*	✓	*
MiMiR	Imperial College London	*	*	✓	RF	*	✓	✓
QuickLIMS	German Cancer Research Institute	*	✓	*	RF	*	*	*
SMD/Longhorn Array Database 1.5	Stanford University	*	✓	*	OS	*	✓	✓
Acuity 4.0	Molecular Devices	*	✓	✓	Com	*	✓	✓
Expressionist	GeneData	*	✓	✓	Com	✓	✓	✓
GeneDirector 3	BioDiscovery	*	✓	✓	Com	*	✓	*
GeneSpring GX Workgroup	Agilent technologies	✓	✓	✓	Com	✓	✓	✓
GeneTraffic(Multi)	Stratagene	✓	✓	✓	Com	*	✓	✓
PARTISAN arrayLIMS	Clondiag	✓	✓	✓	Com	*	✓	✓
Resolver 6.0	Rosetta Biosoftware	✓	✓	✓	Com	✓	✓	✓

OS: Open Source, RF: Royalty Free licence for academics, Com: Commercial Licence, 21 CFR part 11: aimed to comply with U. S. Food and Drug Administration Guidance for Industry [54], ± : partially supported feature, *: unsupported feature or feature not described in software associated documentation, website or publications.

archiving records and standard operating procedures (SOPs) are distributed with the commercial version of the MAF (Discovery Software, see licensing).

Conclusion

We have developed MicroArray Facility, a software tool for the management of microarrays which offers extended Nylon functionalities not found in other freely available LIMS (Table 3). All gene expression profiling steps from cDNA clone management to spot measurements are represented in the MAF database with annotation granularity compatible with MIAME.

Importantly, the MAF system has been tried and tested in both academic shared facilities and industrial environments managing cDNA and Affymetrix gene expression projects. Running in production for five years, MAF has established itself as a central information hub in the laboratory. Investment in data entry is rewarded by providing researchers with fast answers to common queries (e.g. "what is the expression profile of this new marker in our previously tested tumors?") and by helping extract more biological meaning from collected data.

Availability and requirements

• **Project name:** MicroArray Facility (MAF)

• **Project home page:** http://tagc.univ-mrs.fr/bioinformatics/maf/

• **Operating system(s):** Platform independent

• **Programming language:** Perl, SQL

• **Other requirements:** Web Server (e.g. Apache), RDBMS (Oracle or PostgreSQL), MAGEstk

• **Licensing:** Royalty free, non-exclusive, non-transferable INSERM license for academics (excluding SOPs)

• **Any restrictions to use by non-academics:** Discovery Software license from IPSOGEN SAS

Authors' contributions

PHi wrote the first draft of the manuscript. PHi, SDeb & SG conceived of the study, and participated in its design and coordination. PHo developed the MAF module. RT & SDer developed the CC and GF modules. EB and JR participated in MIAME/MAGE-ML implementation. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank Karine Achache, Cédric Bouttes, Régis Brachet-Cota, Vincent Casteras, Frédéric Marciano, Nicolas Mouriaux, David Ruau, Marc Simoni, Virginie Stéphan and David Zinn for their contributions to MAF development. We also thank Helen Parkinson for her help in testing MAGE-ML upload into ArrayExpress. This development was supported by EU grant QLRT-2001-00015 (Temblor/DESPRAD) and the Ipsogen Bioinformatics Department.

References

1. Chaudhuri JD: **Genes arrayed out for you: the amazing world of microarrays.** *Med Sci Monit* 2005, **11:**RA52-62.
2. Auger CJ, Jessen HM, Auger AP: **Microarray profiling of gene expression patterns in adult male rat brain following acute progesterone treatment.** *Brain Res* 2006, **1067:**58-66.
3. Cardoso RS, Junta CM, Macedo C, Magalhaes DAR, Silveira ELV, Paula MO, Marques MMC, Mello SS, Zarate-Blades CR, Nguyen C, Houlgatte R, Donadi EA, Sakamoto-Hojo ET, Passos GAS: **Hybridization signatures of gamma-irradiated murine fetal thymus organ culture (FTOC) reveal modulation of genes associated with T-cell receptor V(D)J recombination and DNA repair.** *Mol Immunol* 2006, **43:**464-472.
4. Yamamoto H, Imai K, Takamatsu Y, Kamegaya E, Kishida M, Hagino Y, Hara Y, Shimada K, Yamamoto T, Sora I, Koga H, Ikeda K: **Methamphetamine modulation of gene expression in the brain: analysis using customized cDNA microarray system with the mouse homologues of KIAA genes.** *Brain Res Mol Brain Res* 2005, **137:**40-46.
5. Thieblemont C, Nasser V, Felman P, Leroy K, Gazzo S, Callet-Bauchu E, Loriod B, Granjeaud S, Gaulard P, Haioun C, Traverse-Glehen A, Baseggio L, Bertucci F, Birnbaum D, Magrangeas F, Minvielle S, Avet-Loiseau H, Salles G, Coiffier B, Berger F, Houlgatte R: **Small lymphocytic lymphoma, marginal zone B-cell lymphoma, and mantle cell lymphoma exhibit distinct gene-expression profiles allowing molecular diagnosis.** *Blood* 2004, **103:**2727-2737.
6. Charafe-Jauffret E, Bertucci F, Ramuz O, Devilard E, Gaulard P, Brousset P, Houlgatte R, Hassoun J, Birnbaum D, Xerri L: **Characterization of Hodgkin's lymphoma-like undifferentiated carcinoma of the nasopharyngeal type as a particular UCNT subtype mimicking Hodgkin's lymphoma.** *Int J Oncol* 2003, **23:**97-103.
7. Devilard E, Bertucci F, Trempat P, Bouabdallah R, Loriod B, Giaconia A, Brousset P, Granjeaud S, Nguyen C, Birnbaum D, Birg F, Houlgatte R, Xerri L: **Gene expression profiling defines molecular subtypes of classical Hodgkin's disease.** *Oncogene* 2002, **21:**3095-3102.
8. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero J, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R: **Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.** *Oncogene* 2004, **23:**1377-1391.
9. Magrangeas F, Nasser V, Avet-Loiseau H, Loriod B, Decaux O, Granjeaud S, Bertucci F, Birnbaum D, Nguyen C, Harousseau J, Bataille R, Houlgatte R, Minvielle S: **Gene expression profiling of multiple myeloma reveals molecular portraits in relation to the pathogenesis of the disease.** *Blood* 2003, **101:**4998-5006.
10. Vey N, Mozziconacci M, Groulet-Martinec A, Debono S, Finetti P, Carbuccia N, Beillard E, Devilard E, Arnoulet C, Coso D, Sainty D, Xerri L, Stoppa A, Lafage-Pochitaloff M, Nguyen C, Houlgatte R, Blaise D, Maraninchi D, Birg F, Birnbaum D, Bertucci F: **Identification of new classes among acute myelogenous leukaemias with normal karyotype using gene expression profiling.** *Oncogene* 2004, **23:**9381-9391.
11. Maroc N, Morel A, Beillard E, De La Chapelle AL, Fund X, Mozziconacci M, Dupont M, Cayuela J, Gabert J, Koki A, Fert V, Hermitte F: **A diagnostic biochip for the comprehensive analysis of MLL translocations in acute leukemia.** *Leukemia* 2004, **18:**1522-1530.
12. Biberthaler P, Neth P, Bach B, Mayer V, Mussack T, Mutschler W, Jochum M: **Initial RNA expression in human monocytes after multiple injury: a screening pilot study on potentially trauma-sensitive factors by using the microarray-technique.** *Eur J Med Res* 2003, **8:**473-484.
13. Bertucci F, Borie N, Ginestier C, Groulet A, Charafe-Jauffret E, Adelaide J, Geneix J, Bachelart L, Finetti P, Koki A, Hermitte F, Hassoun J, Debono S, Viens P, Fert V, Jacquemier J, Birnbaum D: **Identification and validation of an ERBB2 gene expression signature in breast cancers.** *Oncogene* 2004, **23:**2564-2575.
14. Seidel SD, Sparrow BR, Kan HL, Stott WT, Schisler MR, Linscombe VA, Gollapudi BB: **Profiles of gene expression changes in**

L5178Y mouse lymphoma cells treated with methyl methanesulfonate and sodium chloride. *Mutagenesis* 2004, **19:**195-201.

15. Simpson P, Jones C, Mackay A, Lakhani SR: **Gene expression analysis using filter cDNA microarrays.** *Methods Mol Med* 2006, **120:**415-424.
16. **10 Reasons For Using Nylon Based Microarrays** [http://www.daf.jhmi.edu/microarray/faq.htm]
17. Tanaka TS, Ko MSH: **A global view of gene expression in the preimplantation mouse embryo: morula versus blastocyst.** *Eur J Obstet Gynecol Reprod Biol* 2004, **115(Suppl 1):**S85-91.
18. Herrera L, Ottolenghi C, Garcia-Ortiz JE, Pellegrini M, Manini F, Ko MSH, Nagaraja R, Forabosco A, Schlessinger D: **Mouse ovary developmental RNA and protein markers from gene expression profiling.** *Dev Biol* 2005, **279:**271-290.
19. Wu Y, Zhang X, Bardag-Gorce F, Robel RCV, Aguilo J, Chen L, Zeng Y, Hwang K, French SW, Lu SC, Wan YY: **Retinoid X receptor alpha regulates glutathione homeostasis and xenobiotic detoxification processes in mouse liver.** *Mol Pharmacol* 2004, **65:**550-557.
20. Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR: **Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples.** *Hum Mol Genet* 1999, **8:**1715-1722.
21. Salin H, Vujasinovic T, Mazurie A, Maitrejean S, Menini C, Mallet J, Dumas S: **A novel sensitive microarray approach for differential screening using probes labelled with two different radioelements.** *Nucleic Acids Res* 2002, **30:**e17.
22. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau W, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles – database and tools.** *Nucleic Acids Res* 2005, **33:**D562-6.
23. Ball CA, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Parkinson H, Quackenbush J, Ringwald M, Sansone S, Sherlock G, Spellman P, Stoeckert C, Tateno Y, Taylor R, White J, Winegarden N: **Submission of microarray data to public repositories.** *PLoS Biol* 2004, **2:**E317.
24. Lee JK, Laudeman T, Kanter J, James T, Siadaty MS, Knaus WA, Prorok A, Bao Y, Freeman B, Puiu D, Wen LM, Buck GA, Schlauch K, Weller J, Fox JW: **GeneX Va: VBC open source microarray database and analysis software.** *Biotechniques* 2004, **36:**634-8. 640, 642
25. Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C: **BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data.** *Genome Biol* 2002, **3:**SOFTWARE0003.
26. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis.** *Biotechniques* 2003, **34:**374-378.
27. Killion PJ, Sherlock G, Iyer VR: **The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford Microarray Database (SMD).** *BMC Bioinformatics* 2003, **4:**32.
28. Kokocinski F, Wrobel G, Hahn M, Lichter P: **QuickLIMS: facilitating the data management for DNA-microarray fabrication.** *Bioinformatics* 2003, **19:**283-284.
29. McIndoe RA, Lanzen A, Hurtz K: **MADGE: scalable distributed data management software for cDNA microarrays.** *Bioinformatics* 2003, **19:**87-89.
30. Burgarella S, Cattaneo D, Pinciroli F, Masseroli M: **MicroGen: a MIAME compliant web system for microarray experiment information and workflow management.** *BMC Bioinformatics* 2005, **6(Suppl 4):**S6.
31. Hancock D, Wilson M, Velarde G, Morrison N, Hayes A, Hulme H, Wood AJ, Nashar K, Kell DB, Brass A: **maxdLoad2 and maxdBrowse: standards-compliant tools for microarray experimental annotation, data management and dissemination.** *BMC Bioinformatics* 2005, **6:**264.
32. Maurer M, Molidor R, Sturn A, Hartler J, Hackl H, Stocker G, Prokesch A, Scheideler M, Trajanoski Z: **MARS: microarray analysis, retrieval, and storage system.** *BMC Bioinformatics* 2005, **6:**101.
33. Navarange M, Game L, Fowler D, Wadekar V, Banks H, Cooley N, Rahman F, Hinshelwood J, Broderick P, Causton HC: **MiMiR: a comprehensive solution for storage, annotation and exchange of microarray data.** *BMC Bioinformatics* 2005, **6:**268.
34. Webb SC, Attwood A, Brooks T, Freeman T, Gardner P, Pritchard C, Williams D, Underhill P, Strivens MA, Greenfield A, Pilicheva E: **LIMaS: the JAVA-based application and database for microarray experiment tracking.** *Mamm Genome* 2004, **15:**740-747.
35. Anderle P, Duval M, Draghici S, Kuklin A, Littlejohn TG, Medrano JF, Vilanova D, Roberts MA: **Gene expression databases and data mining.** *Biotechniques* 2003:36-44.
36. Imbert MC, Nguyen VK, Granjeaud S, Nguyen C, Jordan BR: **'LABNOTE', a laboratory notebook system designed for academic genomics groups.** *Nucleic Acids Res* 1999, **27:**601-607.
37. Donovan DM, Becker KG: **Double round hybridization of membrane based cDNA arrays: improved background reduction and data replication.** *J Neurosci Methods* 2002, **118:**59-62.
38. Badiee A, Eiken HG, Steen VM, Lovlie R: **Evaluation of five different cDNA labeling methods for microarrays using spike controls.** *BMC Biotechnol* 2003, **3:**23.
39. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** *Nat Genet* 2001, **29:**365-371.
40. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2006, **34:**D173-80.
41. Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M, Swiatek M, Marks WL, Goncalves J, Markel S, Iordan D, Shojatalab M, Pizarro A, White J, Hubley R, Deutsch E, Senger M, Aronow BJ, Robinson A, Bassett D, Stoeckert CJJ, Brazma A: **Design and implementation of microarray gene expression markup language (MAGE-ML).** *Genome Biol* 2002, **3:**RESEARCH0046.
42. Gardiner-Garden M, Littlejohn TG: **A comparison of microarray databases.** *Brief Bioinform* 2001, **2:**143-158.
43. Brazma A, Robinson A, Cameron G, Ashburner M: **One-stop shop for microarray data.** *Nature* 2000, **403:**699-700.
44. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S, Brazma A: **ArrayExpress – a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2005, **33:**D553-5.
45. Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y: **CIBEX: center for information biology gene expression database.** *C R Biol* 2003, **326:**1079-1082.
46. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95:**14863-14868.
47. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5:**R80.
48. **Ipsogen Profilesoftware™** [http://www.ipsogen.com/services/index_data_analysis.html]
49. Lopez F, Rougemont J, Loriod B, Bourgeois A, Loi L, Bertucci F, Hingamp P, Houlgatte R, Granjeaud S: **Feature extraction and signal processing for nylon DNA microarrays.** *BMC Genomics* 2004, **5:**38.
50. **Fujifilm Arraygauge™ Software** [http://home.fujifilm.com/products/science/si_software/arraygauge.html]
51. **Molecular Devices Genepix™ Software** [http://www.moleculardevices.com/pages/software/gn_genepix_pro.html]
52. **Biodiscovery Imagene™ Software** [http://www.biodiscovery.com/index/imagene]

53. **Microarray Gene Expression Data Society** [http://www.mged.org/]
54. **U. S. Food And Drug Administration Guidance For Industry Pharmacogenomic Data Submissions** [http://www.fda.gov/cber/gdlns/pharmdtasub.htm]