

Software

TAMEE: data management and analysis for tissue microarrays

Gerhard G Thallinger¹, Kerstin Baumgartner¹, Martin Pirklbauer¹,
Martina Uray³, Elke Pauritsch³, Gabor Mehes⁴, Charles R Buck^{4,5},
Kurt Zatloukal⁶ and Zlatko Trajanoski*^{1,2}

Address: ¹Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria, ²Christian Doppler Laboratory for Genomics and Bioinformatics, Petersgasse 14, 8010 Graz, Austria, ³Department of Mathematics C, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria, ⁴ORIDIS Biomed GmbH, Stiftingtalstrasse 3-5, 8010 Graz, Austria, ⁵Bindley Bioscience Center, Purdue University, 475 Stadium Mall Drive, West Lafayette, IN 47907, USA and ⁶Institute of Pathology, Medical University of Graz, Auenbruggerplatz 25, 8036 Graz, Austria

Email: Gerhard G Thallinger - Gerhard.Thallinger@tugraz.at; Kerstin Baumgartner - Kerstin.Baumgartner@gmx.at;
Martin Pirklbauer - Martin.Pirklbauer@gmx.at; Martina Uray - uray@icg.tu-graz.ac.at; Elke Pauritsch - Elke.Pauritsch@gmx.at;
Gabor Mehes - Mehes@oridis-biomed.com; Charles R Buck - CBuck@purdue.edu; Kurt Zatloukal - Kurt.Zatloukal@meduni-graz.at;
Zlatko Trajanoski* - Zlatko.Trajanoski@tugraz.at

* Corresponding author

Published: 7 March 2007

Received: 17 January 2007

BMC Bioinformatics 2007, **8**:81 doi:10.1186/1471-2105-8-81

Accepted: 7 March 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/81>

© 2007 Thallinger et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: With the introduction of tissue microarrays (TMAs) researchers can investigate gene and protein expression in tissues on a high-throughput scale. TMAs generate a wealth of data calling for extended, high level data management. Enhanced data analysis and systematic data management are required for traceability and reproducibility of experiments and provision of results in a timely and reliable fashion. Robust and scalable applications have to be utilized, which allow secure data access, manipulation and evaluation for researchers from different laboratories.

Results: TAMEE (Tissue Array Management and Evaluation Environment) is a web-based database application for the management and analysis of data resulting from the production and application of TMAs. It facilitates storage of production and experimental parameters, of images generated throughout the TMA workflow, and of results from core evaluation. Database content consistency is achieved using structured classifications of parameters. This allows the extraction of high quality results for subsequent biologically-relevant data analyses. Tissue cores in the images of stained tissue sections are automatically located and extracted and can be evaluated using a set of predefined analysis algorithms. Additional evaluation algorithms can be easily integrated into the application via a plug-in interface. Downstream analysis of results is facilitated via a flexible query generator.

Conclusion: We have developed an integrated system tailored to the specific needs of research projects using high density TMAs. It covers the complete workflow of TMA production, experimental use and subsequent analysis. The system is freely available for academic and non-profit institutions from <http://genome.tugraz.at/Software/TAMEE>.

Background

Tissue microarray (TMA) technology is a high-throughput tool for simultaneous analysis of up to 1000 different tissue samples at the DNA, RNA or protein level in a single experiment. TMA technology evolved from different precursors [1-5] to a form described by Kononen *et al.* [6] and includes numerous varieties [7-10]. TMAs as proposed by Kononen consist of a paraffin block (array or recipient block) into which cylindrical tissue cores extracted from different formalin-fixed paraffin-embedded (FFPE) tissue samples (donor blocks) are inserted in an array. From a single array block up to 200 sections can be made and analyzed by immunohistochemistry (IHC), *in situ* hybridization (ISH), or immunofluorescence (IF).

Compared to conventional pathology analyses where single tissue specimens are investigated individually, TMAs provide results from large numbers of different tissue samples in a single experiment. The major advantage is that the experimental conditions for all samples are equal and the amounts of consumables required are markedly reduced. TMAs also preserve precious raw material (archived tissue samples) since a single donor block can provide cores for multiple arrays [11,12].

The TMA workflow comprises the following steps: (i) design of the TMA and selection of tissue samples, (ii) production of the array block, (iii) sectioning of the block and quality control of the sections, (iv) molecular analysis of TMA sections, (v) image acquisition, (vi) evaluation of core images, and (vii) statistical analysis of the results. These steps produce a significant amount of data that must be stored in a centralized repository to give all researchers involved in a TMA-based study access to the relevant information. Special emphasis has to be directed to the standardization of the entered data through the use of existing ontologies and classifications to enable meaningful down-stream analysis. Sample scoring has now become the major bottleneck since instead of a small number of stained whole biopsy sections, hundreds of cores must be evaluated in a reproducible and timely manner.

The power of the TMA technology has led to a rapid increase in the number of labs adopting the technology and an explosion of generated data. Managing and extracting valuable information from such data requires new and efficient data management platforms and computational approaches. Numerous efforts have been undertaken to develop software tools to handle different aspects of the TMA workflow. However, to the best of our knowledge, there is currently no academic data management platform supporting the complete TMA pipeline. Among existing applications spreadsheet based systems [13-16] have been suggested as well as applications con-

centrating on web based manual core evaluation [17-20] and applications with a wider range of workflow components [21-27]. The applications of Chen *et al.* [28,29] and Rabinovich *et al.* [30] focus mainly on the image acquisition step. Only Vrolijk *et al.* [15] and Demichelis *et al.* [25] provide systems that contain a component for automatic core localization and extraction of core images (see Table S1 in Additional file 1 for comparison of features of currently published systems).

One of the most important steps in the TMA workflow – the evaluation of core images – has been addressed in a single database application [25]. The results of the evaluation are the basis for the downstream analysis, where biological markers are correlated with outcome and prognosis or used to determine therapeutic measures. Manual evaluation has been shown to be prone to significant inter- and intra observer variability, especially for borderline cases and with less experienced clinicians [31]. Automated image analysis promises to reduce evaluation variability [31] and to reveal novel findings, not detected by traditional pathologist-based scoring [32]. Different (semi)-automated evaluation algorithms have been described for IHC [15,29,30,33-35], IF [30,32,36,37] and for ISH experiments [38]. Only the algorithm from Dell'Anna *et al.* [35] is integrated into a web-based TMA data management application [25].

Given the main shortcomings in the area of quality control, image acquisition and automated core analysis sparked us to develop TAMEE, a web-based database application using state of the art software technology, ensuring extensibility, scalability, maintainability and portability. It supports the complete TMA workflow, comprising the following features:

- Management of data related to tissues and donor blocks
- Storage of TMA design and array block data
- Tracking of array block sectioning and section quality control
- Automated gridding of the TMA slide images and extraction of core images
- Automated evaluation of core images and flexible storage of evaluation results
- Plug-in framework for the integration of additional evaluation algorithms
- Support for standardized and consistent database content with controlled vocabularies

- Compatibility with data interchange standards
- Operating system and browser independence

Implementation

System architecture

The TMA data management application was implemented using Java, an object-oriented and platform independent programming language [39]. The application is three tiered, consisting of the database tier, the application server tier and the client tier. The open source database MySQL [40] is used as the data-base backend and the business logic is implemented using Enterprise Java Beans (EJBs) [41]. The runtime environment for the EJBs is provided by JBoss, a J2EE compatible application server [42]. The presentation layer is implemented using the Model-View-Controller (MVC) framework Struts [43]. Features requiring complex, immediate response on user input are implemented as Java Applets [44].

Integration of algorithms evaluating core images is facilitated by a plug-in framework in a flexible and extendable way. The framework consists of the following components: (i) an abstract algorithm bean, which defines the methods to be implemented and provides convenience functions which can be utilized by the derived algorithms, (ii) a flexible database model to manage the available algorithms and to store the image evaluation results (see below), and (iii) a web interface to hot-deploy newly created algorithms and to manage existing ones. A description of the framework and Java source code for a sample algorithm is provided as Additional file 2. It can be used as the basis for the development of additional algorithms. Commercial algorithms can be integrated by implementing a "wrapper" algorithm, provided a command line version of the commercial algorithm exists.

The complete application has been designed using UML [45]. The UML representation improves maintainability as the application architecture is immediately visible and the components can be easily identified. The UML model is the basis for code generation with the AndroMDA framework [46], which relieves the developer from tedious and repetitive coding task and allows him to concentrate on the business logic and presentation part of the implementation. The UML model is available as Additional file 3.

Database schema

The database schema supports the complete TMA workflow (see Additional file 4). TMA design and construction is covered by the "Tissues", "Donor" and "Arrayblock" tables. The tissue tables are accompanied by the tables storing the ICD-O diagnosis coding and the TNM staging definitions in the data dictionary. Additionally de-identi-

fied sample identification is stored, which provides a link to both sensitive and insensitive patient data in the clinical database. The array layout including the array attributes like core diameter and spacing is associated with a constructed array block. The "Section" table stores data related to a cut section like related array block, thickness, section index within array block, and type of microtome used. One or more section images (stained or unstained) can be related to a section via the "SectionImage" table. For stained sections, the staining protocol, the antibody used and the related antigen are recorded. Each core stored in the "Punch" table is associated to a specific section, the corresponding array block and the related donor. A list of available evaluation algorithms is located in the "Algorithm" table; the evaluation results of core images in the "PunchImage" table are stored in the "Result" table.

A hybrid database schema has been designed consisting of a conventional relational part for entities with a static attribute set and a part implemented as an entity-attribute-value (EAV) schema [47,48]. In the former, each parameter is represented by a separate column in a database table. The latter is used to support the highly dynamic attribute set resulting from the manual and automatic evaluation of core images. The entity is a certain core image, which has been evaluated, the attribute is the algorithm used for the evaluation of the core image and the value is the result of the evaluation. A simplified EAV approach is used, where all values are stored as strings regardless of their actual data type. Using the EAV approach to model the evaluation data, results from new, currently not foreseen scoring algorithms can be accommodated in a generic way, without the necessity to update the database schema.

Controlled vocabularies are stored on the one hand in the ICD-O related tables ("Topography", "Morphology", "Behaviour" and "Grade") and on the other hand in the "Datadictionary" table.

All images are stored as binary large objects (BLOBs) in the database using two formats, namely resolution reduced, compressed JPEG for quick viewing and high resolution, lossless compressed TIFF for scoring and display of image details.

Image acquisition and extraction of core images

Images of the HE stained donor block sections, which are used to determine the positions where the cores are extracted and the TMA sections used for quality control are scanned with an Epson Perfection 1660 Photo flatbed scanner (Epson, Meerbusch, Germany) with 16 µm per pixel. IF-stained sections were digitized with a GMS-418

laser scanner (Affymetrix, Santa Clara, CA, USA) at 10 μm per pixel.

Extraction of core images is a three step process: (i) marking of corner cores position (ii) localization of cores in the image, (iii) determination of core extents, and (iv) saving of core images in the database. Prior the determination of the core positions, color images are converted to gray-scale, taking the different luminance contributions of the colors into account. Corner core positions are marked manually to be able to determine the overall array orientation. After cropping, mirroring, and rotation of the image based on the marked corner positions, the centers of the cores are determined using the radon transform [49]. Based on the array layout (the number of rows and columns) the local minima in the radon transformed data are used as the starting points for horizontal and vertical lines. The crossings of the lines are regarded as approximate centers of the cores. A binary image is created based on the gray-scale histogram and noise is removed by morphological image operations. The previously determined core centers are used as the seed for the subsequent border tracing algorithm [50] to determine the exact extent of the core in the image. Rectangular sub-images which cover the complete core are extracted from the section image and all pixels outside the cores extend are set to 0 before the image is stored in the database.

Section quality control

After a section has been cut and transferred to the carrier slide, it is scanned with a flatbed scanner to keep track of the section quality. These quality control images are gridded and the images of the cores extracted. For each core, the nominal core area (calculated based on the core diameter from the array layout) is compared to the actual area. If the actual area is below (indicating loss of tissue) or above (indicating overlapping or aggregation of cores) a certain percentage of the nominal one, the core is flagged as bad. Additionally, cores whose shape significantly deviates from circularity are marked.

Analysis of multiple IF stained cores

Double stained sections using antibodies labeled with different fluorophores (e.g. Cy3 and Cy5) are scanned with a laser scanner, which creates two grayscale images, one for each fluorophore [37]. A merged image is gridded and the core images are extracted from both original images as described above. The individual core images are evaluated with respect to the following parameters: median core intensity, integral of core intensity and percentage of core area stained. Additionally, the ratio of the median intensities of the two fluorophores, the percentage to which the stained areas of the two fluorophores overlap, and the median intensity of one fluorophore in the area which is stained by the other are calculated. The evaluation results

are stored in the database, where they are available for further analysis.

Results and discussion

TAMEE is a system which covers all steps in the TMA workflow, starting with the design of a TMA and spanning to the automated analysis of core images. It accommodates the requirements of different tissue types and of current and future image evaluation algorithms. The web-based application uses J2EE application server technology providing industry grade stability, scalability, and extensibility.

Tissue microarray design

Prior to the production of a TMA, tissues appropriate for a given TMA design are selected by querying the clinical database which contains information regarding cases and associated donor blocks. Basic tissue information (like diagnosis and unique sample identification) is exported from the clinical database in tabular form. This table is amended with the bar-code identifier of the donor block and specific comments about core selection and location. The sample data related to an array is uploaded into the database and joined after array production with the layout information. Alternatively, tissue and donor block information can be entered manually if needed. Images of donor block sections stained with Hematoxylin and Eosin (HE), used to identify a representative core position within the specimen, are also stored in the database for quality control and tracking purposes (Figure 1).

Production and quality control

Array blocks are produced with a custom-built robotic TMA Arrayer (Oridis-Biomed GmbH, Graz, Austria), based on the selected position of a core on the donor block and the target location for the core in the array block. Each array block is also identified by a unique bar-code. During array production, the robot control software produces an array description file in tabular form, including the donor block barcode, core location on the donor block, and row and column of the core in the TMA. This file is uploaded into the database and associated to the corresponding donor blocks and to one of several existing array layouts possible with the arrayer. Currently four layouts encompassing 240 to 487 cores are available (Figure 2). Additional array layouts can be defined via the web interface.

Sections created from the TMA block are recorded in the database together with related information like section thickness, type of microtome used, and corresponding depth in the TMA block. Generally, sections are created in batches; this is accounted for in the section entry screen, where section parameters of a batch including the number of sections are defined (Figure 3). To track the quality of

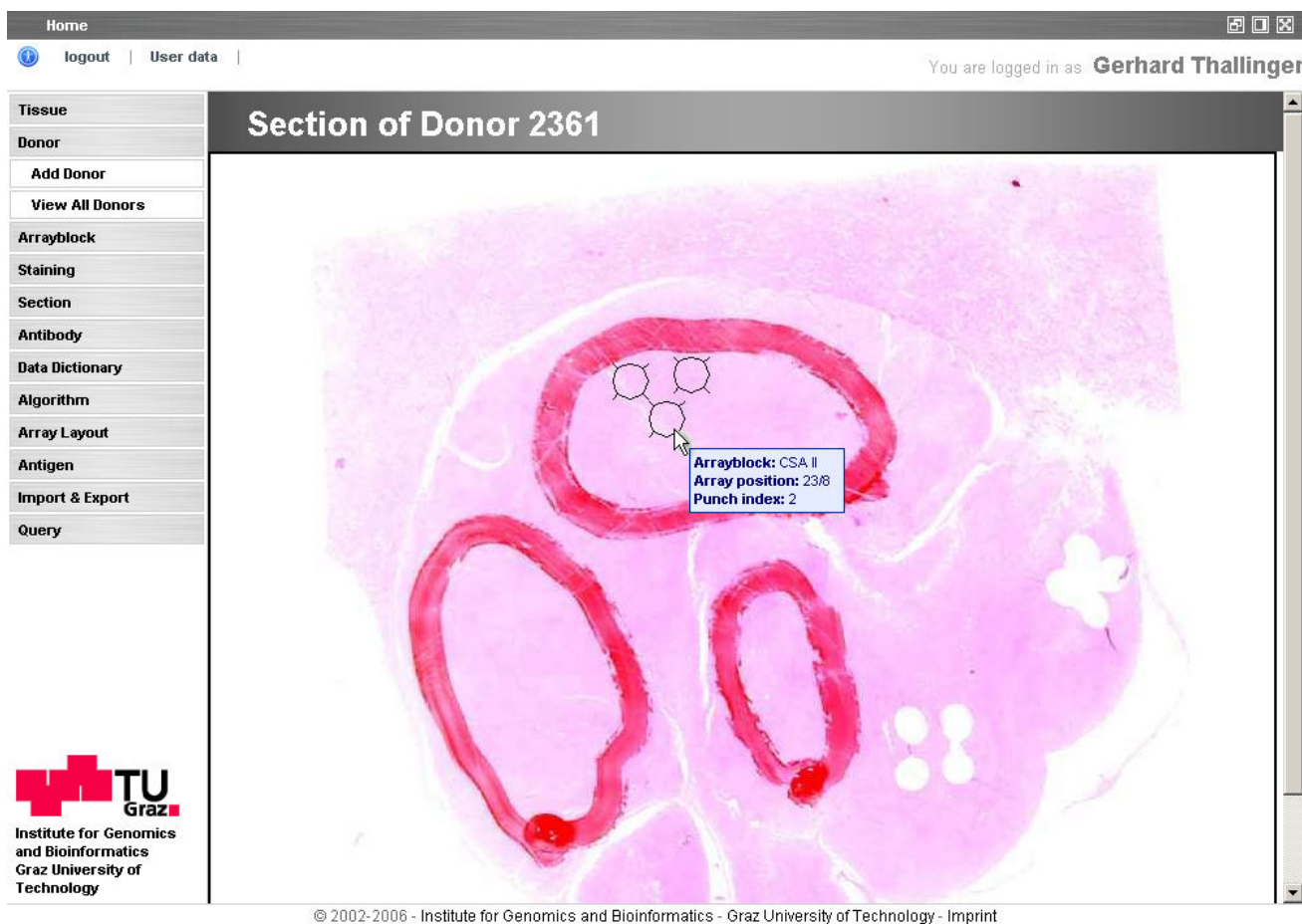


Figure 1
HE stained section of a donor block. The section is used to identify areas of interest in the donor block for extraction of cores to place in the array block (identified by a pathologist with red or black ink). The positions where cores have been extracted are marked with black circles. Moving the mouse over a core position displays detailed information about the related array block, the position of the core on the array block and the index of the core on the donor block.

the cut sections, a flatbed scanner is used to generate a low resolution image of the unstained slide. These images are stored in the database and quality scores that include, for example, core completeness and core shape are assigned via an evaluation algorithm.

Molecular section analysis

TMA sections are useful for a range of molecular analyses that can also be performed on regular tissue sections, such as IHC, ISH, or IF. For every analysis the subject of the analysis (antigen, DNA or RNA sequence) and the experimental conditions are recorded. Type and concentration of antibodies utilized in an IHC or IF analysis are added as an experimental parameter. To ensure reproducibility, the experimental protocol is associated in TAMEE as a PDF file for reference during data evaluation. Digital images of the stained section, obtained either via photom-

icroscopy or from a laser scanner, are related to the analysis parameters and stored for later core image extraction and analysis. In general one grayscale or color image is available from an experiment, in the case of IF experiments with multiple antigens investigated, two grayscale images are created for joint evaluation [37].

Extraction and evaluation of core images

Images of a complete TMA section are uploaded into the database. During core image extraction the individual cores are automatically located (Figure 4) and a square image area with the core in its center is stored for later analysis. Extraction from both HE or IHC stained images (with a white background) and IF images (with a black background) is supported. Pixels not belonging to the core (either background or artifact) are set to zero to provide a standardized basis for core analysis. Gridding and

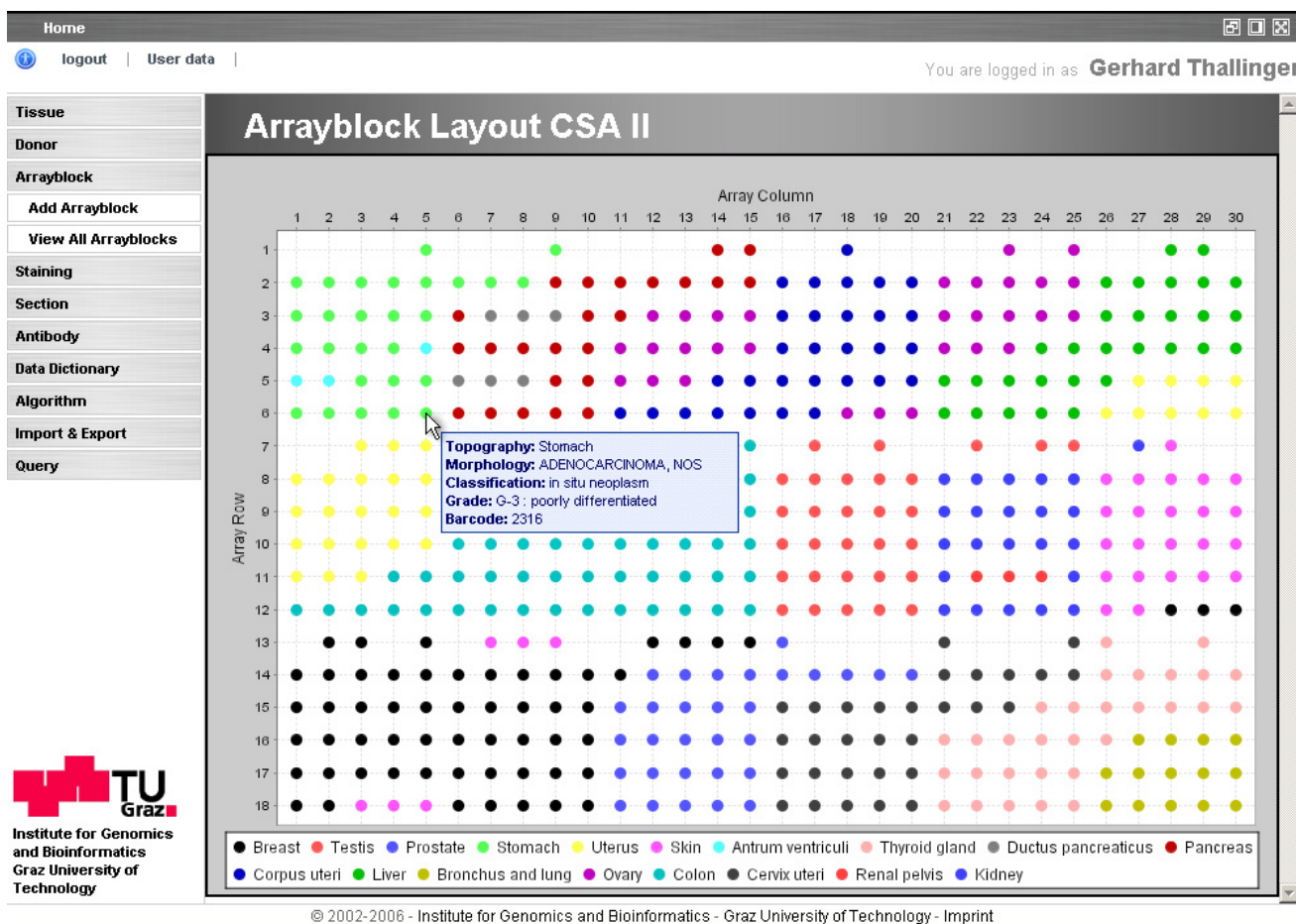


Figure 2
Layout of a TMA with 487 cores. The individual cores of the array are colored according to the topographic origin of the tissue sample. Moving the mouse over a core displays detailed information about the core: topography, morphology, classification, grade (according to ICD-O) and the barcode of the donor block.

extraction of cores from an image of a TMA with 487 cores is accomplished within 35 seconds on an Intel Pentium M system at 1.6 GHz and 1 GB of RAM.

Cores originating from the quality control section images are evaluated for area and shape (roundness). Cores that fall within certain area and roundness ranges are considered acceptable for evaluation in later experimental analysis. If a section shows too many cores outside the acceptable range, the section is abandoned.

IF experiments allow the concurrent investigation of multiple antigens using different fluorophores attached to the antibodies to determine, for example, colocalization of a tumor marker with a cell type specific marker. For each

fluorophore, a grayscale image is created, from which the area stained by the antibody is determined. Additionally, the area of overlap of staining with the different antibodies is calculated. This allows normalization of the results and semi-quantitative analysis of the cores [37].

For each run of an analysis algorithm, evaluation results are written to the database together with the evaluation results, date and time of the run and a reference to the user who invoked the analysis. There is no restriction on the type of evaluation results, categorical data as well as continuous data is supported. The analysis algorithms are embedded in a plug-in framework, which allows the integration and use of additional algorithms without requiring adaptations in other parts of the application (see

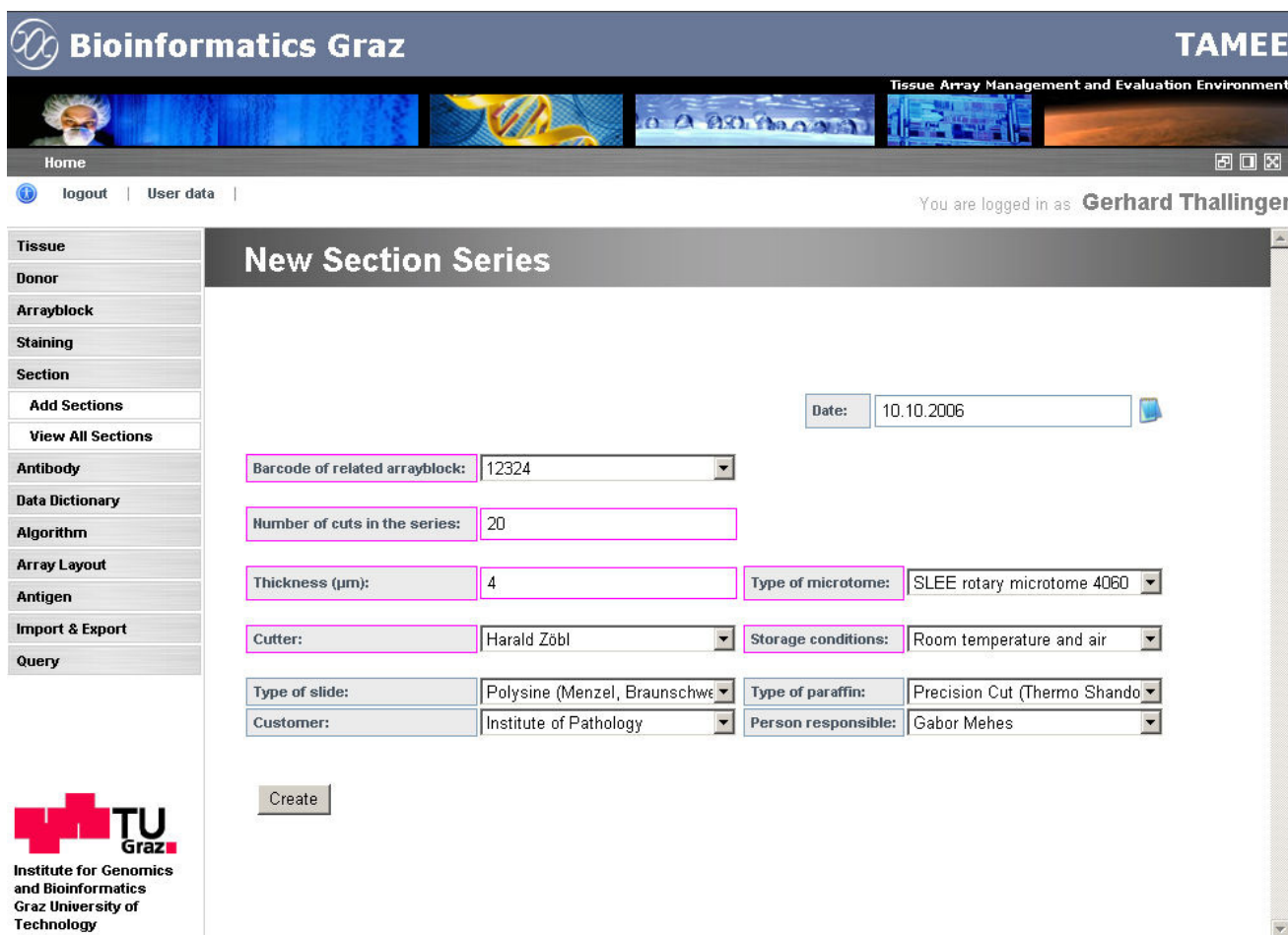


Figure 3
TAMEE sample input screen. Fields where input is required are marked magenta, optional fields are colored blue. Drop-down list boxes for the selection of appropriate attribute values are used extensively to ensure standardized and consistent database content.

Additional file 2 for a description of the framework and a sample algorithm). These additional analysis algorithms can either be newly developed academic ones or existing commercial ones like AQUA [51] or SpotBrowser [52].

Automated core analysis provides several advantages over manual scoring which is affected by intra- and inter-observer variability and allows assignment of categorical scores only. Automated image analysis improves the reproducibility of the evaluation results [31,53] and provides a solid basis for subsequent analysis of the prognostic value of a certain marker [54]. Furthermore, scoring is carried out on a continuous scale, which allows the analysis to be performed without stratification. This increases the statistical power of the analysis and avoids the introduction of a bias [55,56].

Controlled vocabularies

User input for the same entity can be very heterogeneous, depending on educational background, personal preferences and a host of other variables. Non-standardized data is almost impossible to query and therefore diminishes the value of the database. Hence TAMEE uses – wherever possible and reasonable – existing public classifications and self-defined data dictionaries to ensure standardized and consistent data entry. Disease diagnoses are encoded using ICD-O (International Classification of Diseases for Oncology) [57] and the TNM classification scheme [58] is used for the tumor status. Custom data dictionaries have been created for TMA parameters including tissue source, type of paraffin utilized, microtome model used during sectioning, and section storage conditions. The data dictionaries are user extendable, which allows

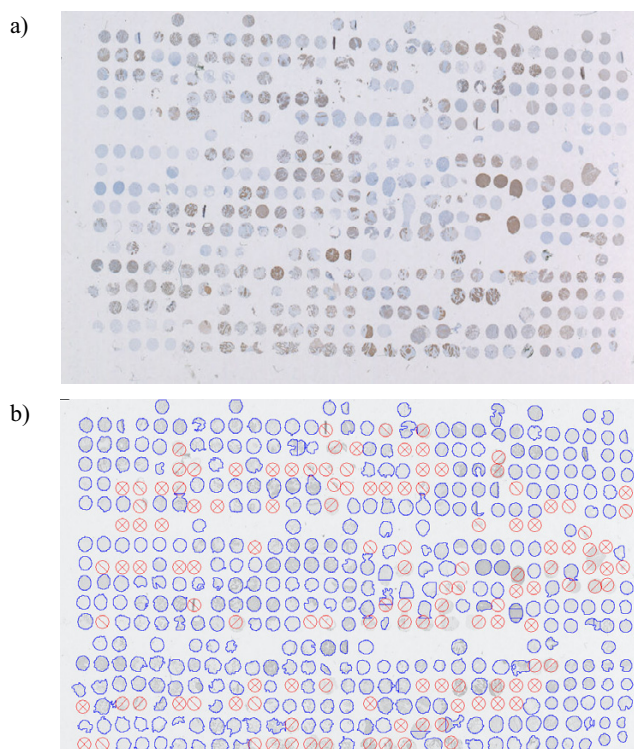


Figure 4

Stained and gridded TMA section. (a) TMA section with about 480 tissue cores from different cancer tissue samples. Cytokeratin has been detected by IHC (brown areas) using a cytokeratin antibody (mouse anti human; DAKO; M3515) and the EnVision detection Kit (DAKO, K4065). Tissue cores are made visible with a Hematoxylin counterstain (light blue). (b) The same section after automatic core localization and shape determination. The outline of a core is displayed in blue. Positions with missing cores (\otimes), cores which are too large (more than 160% of the expected area, \oplus) and cores which are too small (less than 40% of the expected area, \ominus) are shown in red. Array positions without a symbol or outline have been deliberately left empty during TMA production. Gridding ambiguities in heavily distorted areas of the array are corrected manually.

immediate modification of parameter lists without adaptation of the application itself.

Data entry and query

A consistent and clear application interface is provided for users with primarily a medical or biological background. Screens can be divided into two major groups: (i) entry screens for primary data entry, and (ii) list views for the compact display of available entities. Entry screens consist of mandatory fields – colored magenta- and optional

fields in blue (Figure 3). Drop down selection lists are used extensively to ensure standardized data entry. Free text can be entered in fields not used in subsequent analysis queries. Data entered is checked for validity and in the case of an erroneous entry the user is informed appropriately. List view screens present the entities in tabular form and allow paging, querying and the selection of which entity attributes are displayed (Figure 5). Queries can be performed on any combination of the available attributes using a standard set of query operators. After the query is submitted, the displayed list is narrowed to the matching entries of the specified query.

Data access control

Although TAMEE manages de-identified patient data only, access to and modification of the information has to be tightly controlled. Therefore, access to the web application is secured by an authentication and authorization system as described previously [59]. It manages user account information, assignment of users to certain application roles and application resources together with their associated access control lists. By default, TAMEE utilizes three application roles, namely "Administrator", "User", and "Guest". Guest users are granted read only access to all entities. The more privileged User role allows performance of all operations related to the routine TMA workflow. Administrators are enabled to perform management tasks like altering and extending the data dictionary and the definition of array block layouts. Additionally, the right to delete data in the database is granted to administrative users only. The access rules for these groups can be changed according to different requirements and additional groups with special entity access patterns can be defined using the authentication and authorization system.

Support for data exchange specifications

To allow sharing of experimental data in a reproducible and complete form, data interchange standards are becoming increasingly important [60,61]. To this end, the Association for Pathology Informatics proposed the "Tissue Microarray Data Exchange Specification" (TMA-DES) [62-64]. TAMEE can export TMA analysis data in the format proposed in TMA-DES and allows the dissemination of supplemental experimental data related to a publication and the submission of parameters and results to public data repositories in a standardized format. Recently, the "MISFISHIE" standard [65] covering IHC and ISH experiments has been described and TMA-OM [27], an object model for TMA data accompanied by a corresponding XML representation has been suggested. As these data representations gain acceptance, support for them can be added to TAMEE as well.

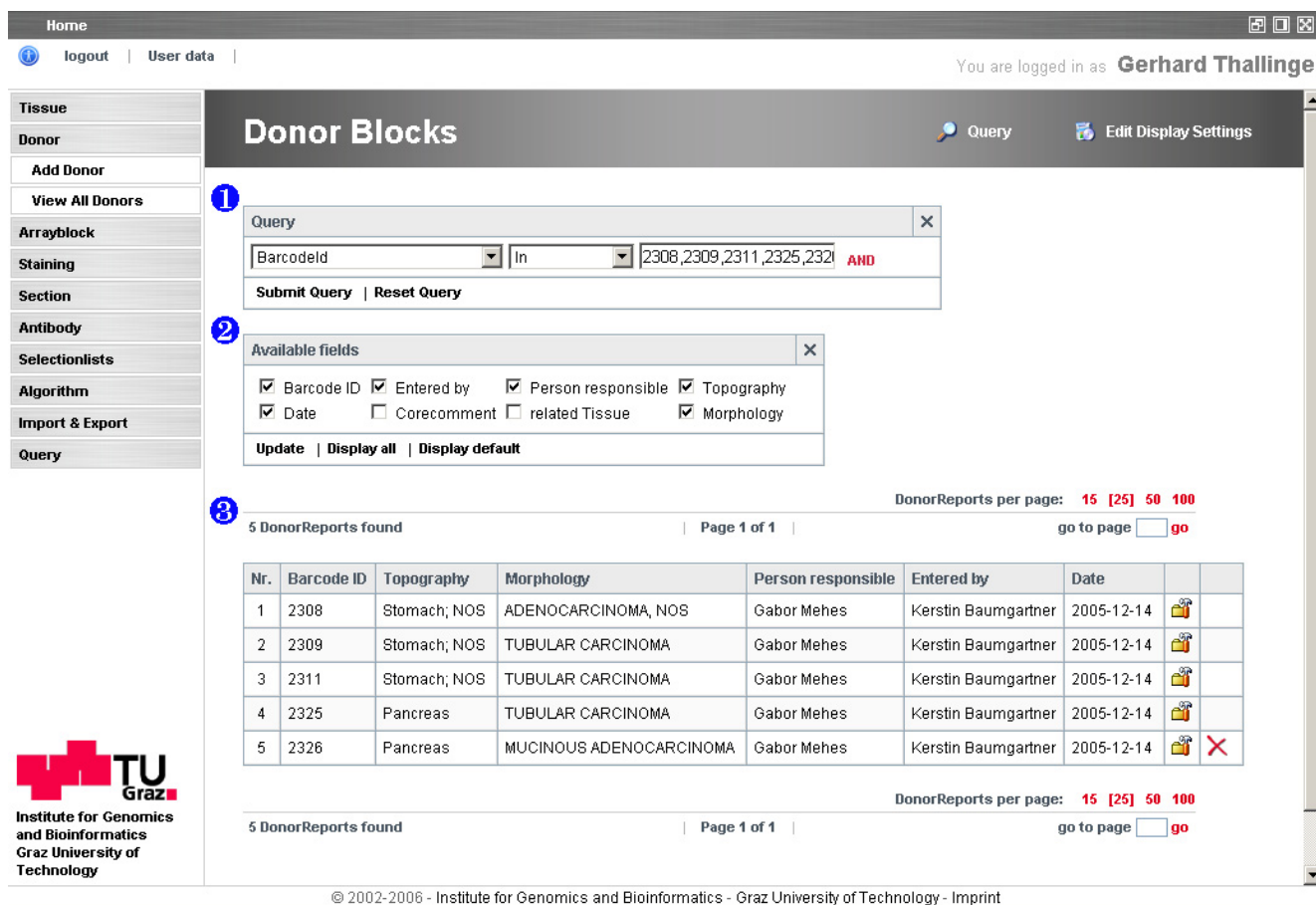


Figure 5
Sample TAMEE list view. A selection of donor blocks based on their barcodes is shown. The list view consists of 1) the query parameters, 2) the customizable display settings and 3) the list itself. Areas 1) and 2) are hidden by default and are made visible by clicking at the respective icons in the header.

Conclusion

We have developed TAMEE, a portable, web-based database application using state of the art application server and database technology tailored to the specific needs of research projects using high density tissue microarrays. The system covers the complete TMA workflow and manages significant amounts of data accumulating during the production and application of TMAs. It allows simultaneous access and update of data as required in multicenter studies. TAMEE facilitates the traceability of all steps during the production and the molecular analyses of TMA sections by collecting the relevant laboratory and quality parameters. Additionally, TMA data can be exported in the standardized TMA-DES format for sharing and dissemination of TMA design, experimental conditions, and evaluation results. Key features are quality control in TMA production and molecular analysis, standardized data entry using controlled vocabularies, and the plug-in

framework for the integration of automated core image evaluation algorithms.

Availability and requirements

- Project name: TAMEE
- Project home page: <http://genome.tugraz.at/Software/TAMEE>
- Operating system: Solaris, Linux, Windows, Mac OS X
- Programming language: Java, HTML
- Other requirements: Java JDK 1.5.x, MySQL 4.0.x with InnoDB engine, a server with at least 1 GBytes of main memory available to the application
- License: None required

- Any restrictions to use by non-academics: None

Installation of the application is straight-forward and should be manageable within a few hours provided the necessary MySQL access rights are granted. Step-by-step instructions are provided at the projects Web site together with the files and scripts necessary for installation. The reference installation of TAMEE is running on a Sun V20z Opteron server (Sun Microsystems Ges.m.b.H, Vienna, Austria) with 8 GB of memory under CentOS Linux [66] using MySQL [40] as database management system. Attached is a Storage Area Network (EVA 5000, Hewlett-Packard Ges.m.b.H., Vienna, Austria) with 5.5 TBytes.

The production instance of TAMEE is hosted on a HP ProLiant ML350G4P (Hewlett-Packard Ges.m.b.H., Vienna, Austria) under Windows 2003 Server with 2 GB of main memory and an 800 GB RAID-5 storage.

List of abbreviations

BLOB Binary Large Object

EAV Entity Attribute Value

FFPE Formalin fixed, paraffin embedded

ICD-O International Classification of Diseases for Oncology

HE Hematoxylin and Eosin

IF ImmunoFluorescence

IHC ImmunoHistoChemistry

ISH *In Situ* Hybridization

J2EE Java 2 Enterprise Edition

JPEG Joint Photography Expert Group

TMA-DES Tissue MicroArray Data Exchange Specification

TIFF Tagged Image File Format

TNM Tumor, Nodes, Metastasis (tumor classification system of the Union against Cancer)

Authors' contributions

GGT designed the application, implemented the section quality control and drafted the manuscript, MP and KB implemented the business – as well as the presentation logic. MU developed the automated gridding of the section images including the extraction of the core images and EP implemented the two color analysis algorithm.

GM designed the cancer survey array and defined quality control procedures. CB and KZ participated in the compilation of the user requirements and contributed to the conception and design of the system. ZT was responsible for the overall project coordination. All authors contributed in reading and revising the manuscript and gave final approval of the version to be published.

Additional material

Additional File 1

Table with features of published TMA applications. PDF file containing a table showing the features of TMA applications described in literature.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-81-S1.pdf>]

Additional File 2

Sample core evaluation algorithm. Zip file containing Java source code of a sample core image evaluation algorithm and a description of the plugin framework as the basis for the development of additional algorithms.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-81-S2.zip>]

Additional File 3

TAMEE UML model. Zip file containing the TAMEE UML model comprising the entity and the service diagrams.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-81-S3.zip>]

Additional File 4

TAMEE entity-relation diagram. Bitmap file containing the TAMEE entity relation diagram.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-81-S4.bmp>]

Acknowledgements

This work was supported by the GEN-AU projects BIN II (Bioinformatics Integration Network II) and GATiB (Genome-Austria Tissue Bank) of the Austrian Federal Ministry of Education, Science, and Culture. We thank S. Trajanoski, W. Schmidt, M. Asslaber, O. Laback and C. Wrighton for fruitful discussions, H. Zöbl for excellent technical support and A. Krogsdam for critically reading the manuscript.

References

1. Lillie RD: *Histopathologic Technic and Practical Histochemistry* 3rd edition. New York, NJ, USA, McGraw-Hill Book Company; 1965:75-76.
2. Battifora H: **The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing.** *Lab Invest* 1986, **55**:244-248.
3. Wan WH, Fortuna MB, Furmanski P: **A rapid and efficient method for testing immunohistochemical reactivity of monoclonal antibodies against multiple tissue samples simultaneously.** *J Immunol Methods* 1987, **103**:121-129.
4. Kraaz W, Risberg B, Hussein A: **Multiblock: an aid in diagnostic immunohistochemistry.** *J Clin Pathol* 1988, **41**:1337.

5. Battifora H, Mehta P: **The checkerboard tissue block. An improved multitissue control block.** *Lab Invest* 1990, **63**:722-724.
6. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4**:844-847.
7. Schoenberg Fejzo M, Slamon DJ: **Frozen tumor tissue microarray technology for analysis of tumor RNA, DNA, and proteins.** *Am J Pathol* 2001, **159**:1645-1650.
8. Hoos A, Cordon-Cardo C: **Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations.** *Lab Invest* 2001, **81**:1331-1338.
9. Howat WJ, Warford A, Mitchell JN, Clarke KF, Conquer JS, McCafferty J: **Resin tissue microarrays: a universal format for immunohistochemistry.** *J Histochem Cytochem* 2005, **53**:1189-1197.
10. LeBaron MJ, Crismon HR, Utama FE, Neilson LM, Sultan AS, Johnson KJ, Andersson EC, Rui H: **Ultrahigh density microarrays of solid samples.** *Nat Methods* 2005, **2**:511-513.
11. Sauter G, Simon R, Hillan K: **Tissue microarrays in drug discovery.** *Nat Rev Drug Discov* 2003, **2**:962-972.
12. Simon R, Mirlacher M, Sauter G: **Tissue microarrays.** *Biotechniques* 2004, **36**:98-105.
13. Liu CL, Prapong W, Natkunam Y, Alizadeh A, Montgomery K, Gilks CB, van de RM: **Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays.** *Am J Pathol* 2002, **161**:1557-1565.
14. Shakhovich R, Celestine A, Yang L, Cattoretti G: **Novel relational database for tissue microarray analysis.** *Arch Pathol Lab Med* 2003, **127**:492-494.
15. Vrolijk H, Sloos W, Mesker W, Franken P, Fodde R, Morreau H, Tanke H: **Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets.** *J Mol Diagn* 2003, **5**:160-167.
16. Liu CL, Montgomery KD, Natkunam Y, West RB, Nielsen TO, Cheang MC, Turbin DA, Marinelli RJ, van de RM, Higgins JP: **TMA-Combiner, a simple software tool to permit analysis of replicate cores on tissue microarrays.** *Mod Pathol* 2005, **18**:1641-1648.
17. Bova GS, Parmigiani G, Epstein JI, Wheeler T, Mucci NR, Rubin MA: **Web-based tissue microarray image data analysis: initial validation testing through prostate cancer Gleason grading.** *Hum Pathol* 2001, **32**:417-427.
18. Manley S, Mucci NR, De Marzo AM, Rubin MA: **Relational database structure to manage high-density tissue microarray data and images for pathology studies focusing on clinical outcome: the prostate specialized program of research excellence model.** *Am J Pathol* 2001, **159**:837-843.
19. Coombes KR, Zhang L, Bueso-Ramos C, Brisbay S, Logothetis C, Roth J, Keating MJ, McDonnell TJ: **TAD: a web interface and database for tissue microarrays.** *Appl Bioinformatics* 2002, **1**:155-158.
20. Conway CM, O'Shea D, O'Brien S, Lawler DK, Dodrill GD, O'Grady A, Barrett H, Gulmann C, O'Driscoll L, Gallagher WM, Kay EW, O'Shea DG: **The development and validation of the Virtual Tissue Matrix, a software application that facilitates the review of tissue microarrays on line.** *BMC Bioinformatics* 2006, **7**:256.
21. Henshall S: **Tissue microarrays.** *J Mammary Gland Biol Neoplasia* 2003, **8**:347-358.
22. Kim R, Demichelis F, Tang J, Riva A, Shen R, Gibbs DF, Mahavishno V, Chinnaiyan AM, Rubin MA: **Internet-based Profiler system as integrative framework to support translational research.** *BMC Bioinformatics* 2005, **6**:304.
23. De Marzo AM, Morgan JD, Iacobuzio-Donahue C, Razzaque B, Faith DA: **TMAJ: Open Source Software to Manage a Tissue Microarray Database [abstract].** *Arch Pathol Lab Med* 2004, **128**:1094.
24. Sharma-Oates A, Quirke P, Westhead DR: **TmaDB: a repository for tissue microarray data.** *BMC Bioinformatics* 2005, **6**:218.
25. Demichelis F, Sboner A, Barbareschi M, Dell'Anna R: **TMABOOST: an integrated system for comprehensive management of tissue microarray data.** *IEEE Trans Inf Technol Biomed* 2006, **10**:19-27.
26. Cowan JD, Rimm DL, Tuck DP: **Cruella: developing a scalable tissue microarray data management system.** *Arch Pathol Lab Med* 2006, **130**:817-822.
27. Lee HW, Park YR, Sim J, Park RW, Kim WH, Kim JH: **The tissue microarray object model: a data model for storage, analysis, and exchange of tissue microarray experimental data.** *Arch Pathol Lab Med* 2006, **130**:1004-1013.
28. Chen W, Foran DJ, Reiss M: **Unsupervised imaging, registration and archiving of tissue microarrays.** *Proc AMIA Symp* 2002:136-139.
29. Chen W, Reiss M, Foran DJ: **A prototype for unsupervised analysis of tissue microarrays for cancer research and diagnostics.** *IEEE Trans Inf Technol Biomed* 2004, **8**:89-96.
30. Rabinovich A, Krajewski S, Krajewska M, Shabaik A, Hewitt SM, Belongie S, Reed JC, Price JH: **Framework for parsing, visualizing and scoring tissue microarray images.** *IEEE Trans Inf Technol Biomed* 2006, **10**:209-219.
31. Bloom K, Harrington D: **Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy.** *Am J Clin Pathol* 2004, **121**:620-630.
32. Camp RL, Chung GG, Rimm DL: **Automated subcellular localization and quantification of protein expression in tissue microarrays.** *Nat Med* 2002, **8**:1323-1327.
33. Wang S, Saboorian MH, Frenkel EP, Haley BB, Siddiqui MT, Gokaslan S, Wians FH Jr., Hynan L, Ashfaq R: **Assessment of HER-2/neu status in breast cancer. Automated Cellular Imaging System (ACIS)-assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard.** *Am J Clin Pathol* 2001, **116**:495-503.
34. Rabinovich A, Agarwal S, Laris C, Price JH, Bemis L, Belongie SJ: **Unsupervised Color Decomposition Of Histologically Stained Tissue Samples.** In *Advances in Neural Information Processing Systems* Edited by: Thrun S, Saul LK and Schölkopf B. Cambridge, MA, MIT Press; 2003:667-674.
35. Dell'Anna R, Demichelis F, Barbareschi M, Sboner A: **An automated procedure to properly handle digital images in large scale tissue microarray experiments.** *Comput Methods Programs Biomed* 2005, **79**:197-208.
36. Rao J, Seligson D, Hemstreet GP: **Protein expression analysis using quantitative fluorescence image analysis on tissue microarray slides.** *Biotechniques* 2002, **32**:924-932.
37. Haedicke W, Popper HH, Buck CR, Zatloukal K: **Automated evaluation and normalization of immunohistochemistry on tissue microarrays with a DNA microarray scanner.** *Biotechniques* 2003, **35**:164-168.
38. Jubb AM, Landon TH, Burwick J, Pham TQ, Frantz GD, Cairns B, Quirke P, Peale FV, Hillan KJ: **Quantitative analysis of colorectal tissue microarrays by immunofluorescence and in situ hybridization.** *J Pathol* 2003, **200**:577-588.
39. Gosling J, Joy B, Steele G, Bracha G: *The Java(TM) Language Specification* 3rd edition. Boston, Addison-Wesley Professional; 2005.
40. MySQL AB: **MySQL database server.** 2005 [<http://www.mysql.com>].
41. Singh I, Stearns B, Johnson M: *Designing Enterprise Applications with the J2EE Platform* 2nd edition. Boston, Addison-Wesley; 2002.
42. JBoss Group: **JBoss Application Server.** 2005 [<http://www.jboss.org/products/jbossas>].
43. Apache Software Foundation: **Apache Struts.** 2006 [<http://struts.apache.org/>].
44. Sun Microsystems: **Java Applets.** 2006 [<http://java.sun.com/applets>].
45. Booch G, Rumbaugh J, Jacobson I: *The Unified Modeling Language User Guide* 2nd edition. Boston, MA, USA, Addison-Wesley Professional; 2005.
46. AndroMDA Core Team: **AndroMDA.** 2006 [<http://www.andromda.org/>].
47. Johnson SB: **Generic data modeling for clinical repositories.** *J Am Med Inform Assoc* 1996, **3**:328-339.
48. Nadkarni PM, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P: **Organization of heterogeneous scientific data using the EAV/IR representation.** *J Am Med Inform Assoc* 1999, **6**:478-493.
49. Toft P: **The Radon Transform - Theory and Implementation.** *PhD Thesis* 1996. Technical University of Denmark, Dept. Math. Modelling.
50. Sonka M, Hlavac V, Boyle R: *Image Processing, Analysis, and Machine Vision* 2nd edition. Pacific Grove, CA, USA, Brooks/Cole Publishing Company; 1998.
51. Rubin MA, Zerkowski MP, Camp RL, Kuefer R, Hofer MD, Chinnaiyan AM, Rimm DL: **Quantitative determination of expression of**

- the prostate cancer protein alpha-methylacyl-CoA race-
 mase using automated quantitative analysis (AQUA): a
 novel paradigm for automated and continuous biomarker
 measurements. *Am J Pathol* 2004, **164**:831-840.
52. Alphelys: **Spot Browser**. 2006 [http://www.alphelys.com/site/us/pTA_StationAnalyse.htm].
 53. Tawfik O, Kimler BF, Davis M, Donahue JK, Persons DL, Fan F, Hagemester S, Thomas P, Connor C, Jewell W, Fabian CJ: **Comparison of immunohistochemistry by automated cellular imaging system (ACIS) versus fluorescence in-situ hybridization in the evaluation of HER-2/neu expression in primary breast carcinoma**. *Histopathology* 2006, **48**:258-267.
 54. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, Pages F: **Type, density, and location of immune cells within human colorectal tumors predict clinical outcome**. *Science* 2006, **313**:1960-1964.
 55. Altman DG, Royston P: **The cost of dichotomising continuous variables**. *BMJ* 2006, **332**:1080.
 56. Royston P, Altman DG, Sauerbrei W: **Dichotomizing continuous predictors in multiple regression: a bad idea**. *Stat Med* 2006, **25**:127-141.
 57. Fritz AG, Ries L: *International Classification of Diseases for Oncology (ICD-O)* 3rd edition. Geneva, World Health Organisation; 2000.
 58. Sobin LH, Wittekind C: *TNM Classification of Malignant Tumours* 6th edition. John Wiley & Sons, Hoboken, New Jersey, USA; 2002.
 59. Maurer M, Molitor 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.
 60. 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.
 61. Orchard S, Hermjakob H, Julian RK Jr., Runte K, Sherman D, Wojcik J, Zhu W, Apweiler R: **Common interchange standards for proteomics data: Public availability of tools and schema**. *Proteomics* 2004, **4**:490-491.
 62. Berman JJ, Edgerton ME, Friedman BA: **The tissue microarray data exchange specification: a community-based, open source tool for sharing tissue microarray data**. *BMC Med Inform Decis Mak* 2003, **3**:5.
 63. Berman JJ, Datta M, Kajdacsy-Balla A, Melamed J, Orenstein J, Dobbin K, Patel A, Dhir R, Becich MJ: **The tissue microarray data exchange specification: implementation by the Cooperative Prostate Cancer Tissue Resource**. *BMC Bioinformatics* 2004, **5**:19.
 64. Nohle DG, Ayers LW: **The tissue microarray data exchange specification: A document type definition to validate and enhance XML data**. *BMC Med Inform Decis Mak* 2005, **5**:12.
 65. Deutsch EW, Ball CA, Bova GS, Brazma A, Bumgarner RE, Campbell D, Causton HC, Christiansen J, Davidson D, Eichner LJ, Goo YA, Grimmond S, Henrich T, Johnson MH, Korb M, Mills JC, Oudes A, Parkinson HE, Pascal LE, Quackenbush J, Ramialison M, Ringwald M, Sansone SA, Sherlock G, Stoeckert CJ Jr., Swedlow J, Taylor RC, Walashek L, Zhou Y, Liu AY, True LD: **Development of the Minimum Information Specification for In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE)**. *OMICS* 2006, **10**:205-208.
 66. The CentOS Team: **CentOS**. 2006 [<http://www.centos.org/>].

Publish with **BioMed Central** and every
 scientist can read your work free of charge

"BioMed Central will be the most significant development for
 disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

