In Cereo and in Silico: Tissue Microarray (TMA) Techniques and Bioinformatics Are Thriving Forces in Medical Science and Personalized Medicine

E. Lang
h_da Darmstadt University, School of Media, Section Information Science & Engineering, Darmstadt, Germany

Summary

Objectives: Tissue microarray (TMA) techniques are among the most promising developments in biomedicine during the last decade. Bioinformatics techniques are indispensable for storing and processing the masses of data related with tissue archive administration and investigation of raw data. Interrelationship between experimental and computational work will be shown.

Methods: Tissue specimen arrays allow parallel analysis of huge amounts of samples. TMA techniques thus produce enormous masses of raw data, and optimal use of data can only be made using modern bioinformatics techniques based on huge storage systems, scalable multilayer software architecture and high-throughput algorithms for retrieval and statistical processing. Further crucial issues addressed by informatics techniques are specimen identification during the whole processing chain, and anonymization whenever scientific work is performed without regard to a certain patient.

Results: TMA supported by bioinformatics methods has helped in identification of biomarkers, mainly in cancer diagnosis. Moreover, it provides powerful means of quality assurance and training in histopathology.

Conclusions: Further statistical analyses seem to be necessary to detect if certain biomarkers are present in nearly all kinds of specimen of the concerned patient, which would allow effective mass screening based on easily accessible specimen. Some investigations showed low dependence on the specimen localization, whereas others suggest to be extremely careful in material selection for the recipient block.

Keywords

Computational biology, data collection, microarray analysis, pathology, biological markers introduction

Introduction

Ten years after completion of the first successful genome elucidation projects [1], [2], bioinformatics and related techniques have achieved a new level of maturity. In the very beginnings of experimental efforts on the molecular scale and of bioinformatics applications, the focus was on gaining whole genome information of a certain species, carried out by long-lasting sequencing work and subsequent alignment. Availability of genome data of more and more species provided a basis for similarity comparison and phylogenetic analyses. Initial overestimation of therapeutic perspectives was balanced by growing knowledge about effects of dynamic aspects like gene expression and regulation, leading to more and thorough investigations. Whole-species characterization based on analysis of individual genes (normally taken from several individuals) was followed by characterization of populations and comparison between individuals [3], [4], leading to insight into gene constellations that are suspected to cause or to indicate diseases. Remarkable progress has been made on this field of investigation during the last years, and huge amounts of data rising from high-throughput experimental work showed to be the boon and the bane for the field of bioinformatics. Large datasets are the prerequisite for all kinds of statistical analyses and data mining, and they require enormous efforts in data storage and retrieval for being processed within a certain research context.

Development of several types of „omics" was also a remarkable step within the last ten years. The initial focus was clearly on DNA for several reasons: DNA is the lead substance in gene analysis, with DNA fragmentation and alignment being the key techniques of analytical work. Moreover, DNA gained initial attention as it was much easier to handle due to its linear chemical structure and clear positions of cleavage. Later developments, however, showed the need for taking into account the behavior of more substance classes like nucleic acids and larger units like proteins and glycoconjugates, creating e.g. proteomics [5], [6], metabolomics [7] and glycomics [8], [9], [10] as new and fruitful fields of investigation.

A new methodology available often raises applications in more and more application fields, leading to dissemination and interdisciplinary investigations. Proteomics revealed that the DNA itself is just a static kind of template, gaining physiological meaning by gene regulation and expression phenomena that decide about how and to which extent the template is used for protein assembly. Glycomics turned attention to the fact that there are much more kinds of biochemical codes [11] besides of the protein-related mechanisms that had been detected in the very beginnings of exploring the realm.

Bioinformatics as a special field of informatics has developed now to being much more than sequence alignment as a context-based form of pattern matching. Actually it involves a range of methods such as pattern recognition, data mining, data fusion, massively parallel computing, information retrieval, development and use of ontologies, and handling of distributed databases.

High-throughput data processing is based on high-throughput data generation: most of the well-known and widespread techniques related with bioinformatics methodology are also related with outstanding laboratory techniques that have been developed and optimized within the last years. The most powerful and universally applied high-throughput techniques are microarrays of different types, providing a suitable tool for many kinds of experimental data generation and data collection, even using existing sample material collections, and with the ability to perform tremendous amounts of sample investigations within a single processing step.

Tissue Microarray Analysis, a State-of-the-art Representative of Modern Bioinformatics Applications

The aforementioned aspects clearly show that nowadays it would be quite impossible to provide a coherent overview of the field of state-of-the-art bioinformatics, as it actually comprises a wealth of different methods and applications. Which subdomain could be estimated as providing a typical single example for the state of the art? A seducing candidate is tissue microarray (TMA) analysis, as this example stands for a broad range of challenges that have been faced and seized within the last years. Therefore, microarray-related techniques as one of the most

promising, still developing and almost indispensable issues of the field will be discussed as a typical example of the interdisciplinary character of modern bioinformatics applications. Combination of laboratory and computing forces has brought remarkable progress in the elucidation of individual disease risk as well as accessibility to certain drug effects with respect to genetic properties of individual organisms. Several issues will be discussed and presented, going into detail concerning several aspects that show generalization potential and will shed light onto phenomena that can be found throughout the field of modern bioinformatics with its deep liaison to experimental work.

Prerequisites and Characteristics of High-content-screening Methods

The development of high-content-screening bioinformatics techniques is essentially related to the availability of appropriate high-throughput material-processing techniques. In the very beginnings of subcellular and molecular structure analysis, over three decades ago, not only were computers slow and data management systems of weak performance and restricted size, but preparation and processing of samples was also restricted to manual work and low amounts of material. The first methods for DNA sequencing, developed in the early Seventies [12], required huge quantities of sample material and reagents and long time to perform. Nearly twenty years later it was possible and roughly affordable to perform complete genome analyses, and another ten years later, this kind of work has nearly become an everyday's task that can be performed for groups of individuals to compare their samples at reasonable cost and feasible time scale. RNA and glycoconjugates seemed to resist to automatic analytical processing for remarkably long time, and examination

of complex tissue structures seemed to be restricted to optical perception and application of a few reaction-specific analyses. Early automation efforts still were restricted to linear processing, but already worked with small sample tubes and lower amounts of sample and reagent material.

Later development brought microarrays as a kind of multidimensional arrangement of samples in order to achieve fast processing. Acceleration, however, was not the only goal that had been intended. An additional core issue that could be addressed was combination and comparison of samples according to their arrangement within the sample matrix, treating it as a kind of addressable array like a data structure. Several techniques have been developed within the last decade that work using sequential and geometrically separated sample positioning in different kinds of arrangement. Solid-state techniques help in gaining data in a context that shows higher relevance with respect to target binding or selectivity.

The scope of (micro)array techniques cannot be overestimated, taking into account that they can cope with macroscopic biological matter (tissue microarray, TMA) [13], [14], and even with matter that had been extracted and preserved considerable time before investigation. This capability leads forward on the way towards multipurpose reuse of data, as it allows to go beyond interpretation of possibly incomplete, inexact and distorted data records and generate experimental results with new accuracy and a high degree of comparability, in certain cases diminished by material alterations due to age and properties of the preservation agents [27].

Arrays in analytical biochemistry and biomedicine can be classified into several types. Mainly tissue microarrays will be discussed further in this work due to their high relevance to bioinformatics. Of course they are mostly interrelated with complementary array technologies such as nucleic acid arrays, dependent on the respective biomedical objectives. Most arrays are micro-

77

In Cereo and in Silico: Tissue Microarray (TMA) Techniques and Bioinformatics Are Thriving Forces in Medical Science and Personalized Medicine

arrays, but there are also macroarray applications [15]. Chemical (substance) microarrays [16] based on liquid or solid-state sample material also represent a relevant type of array-based investigation of biomedical phenomena. TMA results are often combined with those gained by substance array analysis, e.g. when detecting biomarkers. Cellular microarrays [17] can be considered as a kind of bridge between them, as they do not represent tissue structures and therefore do not deliver their typical histological images, but already are complete biological systems with functional and kinetic aspects and with comparable analytical relevance in some main aspects.

From the mid-Eighties on, several techniques had been developed that gain and arrange the tissue sample material in different ways and can use tissue material in different preservation states (mostly embedded in paraffin blocks, in some cases frozen), freshly prepared via different biopsy techniques or derived from xenografts (XMA) or cell lines (CMA). Eguíluz et al. [18] provide a comprehensive overview of almost all of the experimental techniques available and of their benefits and shortcomings. The first micro scale sampling attempt was based on rod-shaped samples taken from donor blocks combined to a sausage-like arrangement in the recipient block [19], [20]. Battifora optimized the sausage block type [21], and Kallioniemi and Sauter et al. [22] introduced the recipient block of core-shaped donor block samples from which all later types of TMA were derived. Core-shaped samples showed remarkable advantages in sample gaining and assembly of the recipient block and therefore now are the basic principle for most tissue microarray techniques. The following years brought quick and fruitful diversification, addressing special challenges of sample gaining and preparation of fresh and preserved material based on paraffin-embedded or frozen material, arranging the samples according to several analytical tasks (combining vari-

ous samples of a single individual or histologically similar samples from several individuals), and application of different staining methods in order to achieve optimal accessibility to automatic pattern recognition methods.

Microscaled amounts of material could hide the fact that they produce macroscaled amounts of data and make processing a challenge that can only be seized by means of specialized and scalable high-content-screening informatics [23].

Taking into account the multitude of tissue cores on one TMA slide just provides a rough number of circa hundred to more than thousand basic data entries per slide. This number of data entries, however, is multiplied by each parameter and step of analysis, such as the matrix of image pixels on the images of each sample core, number of photographs (from pre-array donor block images to several TMA photographs), multitude of image channels that are used to highlight different cell parts using appropriate wavelengths of light, plus some scalar parameters (only one value per core and per slide, respectively), eventually multiplied by the number of time snapshots that are performed in case of a kinetic study. A single tissue core can thus produce hundreds of Megabytes of primary data, and primary data normally are processed using statistical methods in order to derive parameters for the overall description of analysis entities, and for the comparison of entities. Additional metadata are necessary in order to describe experimental parameters like source and age of material, reagents, staining agents and experimental conditions, proceeding of embedding and block formation, temperature, analysis methods, and sampling rate as well as compression factors. Taken together, already very short sequences of analyses lead to data amounts in the Gigabyte and Terabyte range. According to legal and scientific requirements, the data must be stored on highly available or long-time-range media and, if necessary, in redundant copies. Storage Area Networks (SAN) or comparable stor-

age architecture structures like RAID systems have to be used in order to ensure data security and availability. This holds especially for data that have to be recorded as parts of pharmaceutical studies related to drug safety and for studies lasting for a long period. Relation with clinical patient data evolves the need of data protection, restricting access to complete data sets (including personal data) to very few concerned persons and high confidentiality. Storage, however, is just one face of the medal. Retrieval is the crucial part of the process that decides about potential and relevance of a study and that determines the quality of scientific work. This issue is not as obvious as data storage, but has evolved remarkably within the last thirty years and has influenced the way of scientific work based on data comparison and derivation of data taken from different sets of primary data. File management systems and early database systems lacked of retrieval power, providing not more than handling complete files composed from data entries or related to one single data entity, but without an external view of the inherent data structure and certain attribute values. Analysis procedures related to data subsets required reading data files completely and sequentially by multipass parsers, extracting the relevant data sets slowly and with high risk of incompleteness of the result set. Growing data amounts raised the necessity of deriving descriptors that could be used for dividing up the retrieval process into quick prescreening of a small amount of structured descriptive parameters for all entries, followed by slow sequential reading of the complete data of the candidate entry subset that had been filtered during the pre-screening process [24]. Automatic derivation and machine-readable representation of suitable descriptors was a main issue that had first been addressed for chemical structure-based databases like those of Chemical Abstracts Service (CAS) [25]. For long years, the data storage systems were file management systems, later on they were implemented as da-

tabase systems [26], offering the possibility for establishing a multilayer architecture with functional separation of data storage, processing and presentation layer. This kind of architecture (sometimes referred to as three-tier architecture) allows changing separately the different layer applications without influencing the remaining part of the system. The transition from file management systems to database management systems, however, did not necessarily involve direct access to data. The first database attempts were based on a data split that kept descriptive and derived data within the database and according to a highly structured database schema, whereas primary data such as image data and comparably large units of experimental result data were stored as separate files, only related with the database content by means of access keys and index files. This meant that primary data were not accessible to retrieval algorithms and could still only be found by a secondary access guided by prescreening processes. It was only during the last few years that computing power became sufficient enough to enable the retrieval process to be directly performed on the complete data. This is a progress in investigation quality and intensity that cannot be overestimated, as it had been common for decades not only to keep primary data separately, thus restricting the retrieval possibilities, but for long time primary data even were compressed with considerable loss of resolution or other quality issues, or completely rejected after derivation of metadata, thus making it difficult or impossible to reuse experimentally obtained data. This problem shows how especially precious collections of preserved tissues are, as they allow performing experiments again that can deliver data related to the past that originally had not been stored in digital form or that had been compressed with inacceptable loss of data quality. On the other hand, remarkably high effort has to be made in order to ensure identity of donor blocks and identity of certain cores on a slide

of hundreds of cores. Especially access faults on slides prepared from specimen taken from several individuals could lead to wrong mapping of histological findings to patients and cause wrong diagnoses. Rimm et al. [27] report they use RFID labels embedded into the archived donor blocks in order to ensure correct identification. The identity issue is a challenge that could hardly be addressed without the means of informatics technologies.

Further remarkable progress was fostered by sharing and providing collections of experimentally obtained data. One of the first models for comprehensive data sharing was founded by the Cambridge Crystallographic Data Center, collecting and curating all molecular data structure sets in a database that had been obtained by means of diffraction experiments and that had been published in scientific journals [28], [29]. Later similar attempts ensured that protein structure data were made available to the public (Brookhaven Protein Databank, http://www.pdb.org, [30]), followed by more and more data collections for specific substance classes like carbohydrates (http://glycome-db.org/ [31]) or with specific biological function (enzymes, http://expasy.org/enzyme/ [32]).

NCBI (http://www.ncbi.nlm.nih.gov [33]) nowadays presents an outstanding portal for life sciences, having developed from running a merely bibliographic reference database (Medline / PubMed) to a platform providing data collections and algorithms for a broad range of scientific work (http://www.ncbi.nlm.nih.gov/guide/all/). Although it is mainly perceived as and indeed remains the largest bibliographic database in the field of biomedicine and related disciplines, covering thousands of journals and offering powerful retrieval tools, its outstanding value arises from the excellent workbench composed from a wealth of experimental data collections and analysis tools, offering alignment and further types of sequence analysis and comparison, providing ontologies and further types of descrip-

tive tools, and data collections that can be directly accessed via excellent and comfortable retrieval surfaces.

It is not surprising that the NCBI portal contains more than thousands of bibliographic entries concerning TMA-related work. A search performed on all databases shows multiple occurrences in all theme-related databases, showing which kinds of research are carried out based on TMA analysis. Up to now there is, however, no database dedicated to tissue microarray data. Although GEO [34], the database for high-throughput functional genome data, seems to possess a data structure that could accommodate not only DNA and similar array data, but also TMA data, it does not contain explicit TMA image data, although its gene expression data are strongly related to TMA findings and frequently are used in the interpretation of TMA results.

Special software tools for TMA management and data collections dedicated to TMA performance and analysis have been developed during the last years. Some of them are commercially available (SlidePath [35], Aperio [36]), some are open source tools (TmaDB [37], eSlide Suite [38], TAMEE [39] and TIMAN [40]). They differ in the tasks they can support and automatically perform, most of them deal with storage and data mining, some can also handle clinical patient data and support experiment planning. TMA databases can store TMA-related data, but normally are more comprehensive, involving ontology formats and tools supporting several stages of the experimental process and for data processing and mining. The Stanford Tissue Microarray Database [41] (TMAD, http://tma.stanford.edu) is a well-known example that is dedicated to sharing tissue collection (image) data and derived tissue expression data, providing access to Stanford scientists and to the public. Marinelli et al. subsequently enhance the TMAD-related tool collection, providing e.g. statistical tools. Viti et al. [42] present a different approach, TMARep, referring to TMAD but

79

In Cereo and in Silico: Tissue Microarray (TMA) Techniques and Bioinformatics Are Thriving Forces in Medical Science and Personalized Medicine

claiming to provide even better investigation performance by closely integrating ontology data (Open Biomedical Ontology, OBO, and OLS service), MeSH, and image data. Wang et al. [43] propose a wiki-based approach, ArrayWiki, and state that wiki-type implementations are more likely to provide appropriate means for the comprehensive curation work that has to be performed by producers and users of data.

The Human Protein Atlas (http://www.proteinatlas.org/) [44], [45] is a comprehensive portal providing a large collection of images. According to its self-description it is „a publicly available database with millions of high-resolution images showing the spatial distribution of proteins in 48 different normal tissues and 20 different cancer types as well as 47 different human cell lines." Beyond experimental data it provides a „dictionary" section with annotation dictionary and tissue dictionary showing histological tissue images according to anatomic tissue localization or cancer type. This service provides more help to learners than other data sources do.

The following section will show some types of approaches and some examples of investigation and new insight into biological processes gained by using some kinds of the above-mentioned tools.

Relevance of TMA Techniques in Modern Diagnostics and Medical Education

TMA techniques have helped to improve research in existing realms and remarkably extended the fields of application to issues related with cancer research, elucidation of development processes, transcriptomics, proteomics, interpretation of expression data, and detection of biomarkers. In addition, TMA results can help in teaching and provide better means for quality assurance with appropriately designed recipient blocks that help novices in compar-

Table 1 Overview of TMA tools and databases

Name	Source	Functionality
Aperio [36]	www.aperio.com	virtual microscopy, automatic Slide scanning
ArrayWik [43]	http://arraywiki.bme.gatech.edu/index.php/Main_Page	sharing public microarray data repositories and meta-analyses
eSlideSuit [38]	http://www.eslide.net/suite/	set of tools for acquisition, preparation and visualization of digital slides
SlidePath [35]	www.slidepath.com	virtual microscopy, automatic Slide scanning
TAMEE [39]	http://genome.tugraz.at/Software/TAMEE	web-based database application for the management and analysis of data resulting from the production and application of TMAs
TIMAN [40]	http://mitel.dimi.uniud.it/timan/	web-based system basing on open-source software and principles, for pathology cases
TMAD [41]	http://tma.stanford.edu	stores raw and processed data from Tissue Microarray experiments along with their corresponding stained tissue images. provides methods for data retrieval, grouping of data, analysis and visualization as well as export to standard formats
TmaDB [37]	http://www.bioinformatics.leeds.ac.uk/tmadb	comprehensive repository for TMA data
TMARep [42]	http://www.itb.cnr.it/tmabiorep/	managing bio-samples and, through the use of ontologies, enable tissue sharing aimed at the design of Tissue MicroArray experiments and results evaluation
Human Protein Atlas [44, 45]	http://www.proteinatlas.org	comprehensive portal providing a large collection of tissue images, dictionary section

ing tissues und laboratories to verify their preparation, staining and processing routines.

Among the large amounts of projects reported within the last years, just a few typical examples related to elucidation of human cancer genes will be outlined, thus neglecting e.g. work related with model organisms that will surely affect research on human cancers in the near future, but apparently are not directly related to clinical findings.

First approaches mainly aimed to detect specific biomarkers related to certain diseases and disease states with a strong focus on cancers. Recent work remarkably goes beyond, addressing questions of specimen specificity with respect to biomarkers or gene expression variation across specimens taken from several regions of tumor material [46].

Rimm et al. [27] give a comprehensive overview on discovery and validation of cancer biomarkers on the occasion of the ten year anniversary of modern TMA in 2008. They provide a bibliometric diagram showing the rising number of TMA-related articles and the rising percentage of immunohistochemical (IHC) biomarker studies re-

ferring to TMA-based work. Their findings reveal that thorough handling and storage of archival material and semi-assembled recipient structures can prevent specimens from loosing antigenicity against some antibodies, but that experimental results can be remarkably biased by neglecting these prerequisites.

Several investigators [27], [46], [47] addressed the issue of tissue heterogeneity and its impact on detecting related biomarkers. The extremely small volumes of tissue material were suspected not to represent the tumor structure appropriately, namely with respect to the concentration of tumor cells and, hence, of the respective biomarker. It could be clearly demonstrated in many cases, however, that TMA histospots were as representative as their respective entire histological section [48], [49]. Butte et al. [47] even state that, according to their evaluation of several studies, there is strong evidence that for many types of diseases the signature of a disease is robust irrespective of the tissue in which it was studied. They carried out a systematic evaluation on GEO data to evaluate the robustness of

the disease signal across tissues and experiments. Their findings suggest a degree of trans-tissue nature of disease concordance that makes it reasonable to leverage public data to search for biomarkers in easily accessible cell material like blood. Park et al. [50], however, found that normal epithelial and stromal tissue adjacent to tumor tissue, taken from patients with breast cancer, did not show significant gene expression changes. Obviously, further studies should be carried out on this question in order to find out more about the degrees of interrelationship of „hot spot" tissue and certain signal states of gene expression or biomarkers.

Discussion

TMA-based work has influenced the understanding of diseases in a way that could hardly be expected only ten years ago. Traditional histology-based diagnoses using single-slide material were very cumbersome, required high amounts of material and high experience in staining, preparing, and interpretation of tumor material. Moreover, specimen regions had to be chosen very carefully, as visible differences in tissue appearance were required for the correct detection of aberrances. Tissue material of certain individuals could hardly be compared with histologically similar material of other individuals, as variations in staining and preservation made comparison nearly impossible. Assembly of large numbers of tissue cores on one recipient block could overcome problems with different intensities of staining and could build ensembles of samples representing either a multitude of sample regions taken from one individual tissue preparation, or a multitude of tissues taken from a multitude of individuals under comparable histologic conditions. These array-type ordered ensembles enabled and improved research, teaching and quality assurance. Performing crucial steps like staining for whole slides instead

of single specimens avoided scattering of experimental conditions and enabled automated high-throughput processing of array slides. A remarkable number of cancer types could be assessed identifying distinctive gene expression patterns and biomarkers, opening new possibilities for the distinction between normal, conspicuous, and early-cancer-state tissue. Recent work is dedicated to the question of interdependence of specimen origin and gene expression or biomarker intensity in several diseases. A number of findings show low tissue-type specificity for certain diseases, thus promising simple assays based on easily available kinds of specimen that could be used for the development of feasible prescreening tests that could help in early and reliable detection of widespread cancer types. Some tumor types, however, seem to show highly region-dependent expression profiles, requiring high specificity in the election of sample material. Growing archives of tissue material and of data collections related to documented TMA work promise to foster further investigations on this crucial issue.

Conclusions

Bioinformatics is one of the fields of applied computer science that is most deeply interrelated with experimental biomedical science. Vice versa, tissue microarray applications can hardly be imagined without the help of informatics from experiment planning to data storage, administration of donor block arrays and multi-sample slides, and statistical applications for derivation of metadata and cross-study findings. TMA, together with related array techniques like nucleic acid arrays, has enhanced diagnostic progress, quality assessment in laboratories and training in histology within the last ten years to an extent that could hardly be imagined before. With regard to bioinformatics methodology, the main finding is that obviously no completely new

methods had to be developed in order to achieve the challenges rising in the field of TMA applications. Progress lies in the enhancement and the combination of core techniques, in the adoption of applications taken from other fields (e.g. identification of items, originally developed in the field of logistics), and in the remarkable success of ensuring scalability via application of appropriate software architecture design. Thus, TMA applications show in detail the degree of maturity that bioinformatics methodology has achieved in the large.

References

1. Venter JC, Adams MD, Martin-Gallardo A, McCombie WR, Fields C. Genome sequence analysis: scientific objectives and practical strategies. Trends Biotechnol 1992 Jan-Feb;10(1-2):8-11.
2. Broder S, Venter JC. Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. Annu Rev Pharmacol Toxicol 2000;40:97-132.
3. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, et al. Complete Khoisan and Bantu genomes from southern Africa. Nature 2010 Feb 18;463(7283):943-7.
4. Mir KU. Sequencing genomes: from individuals to populations. Brief Funct Genomic Proteomic 2009 Sep;8(5):367-78.
5. Riffle M, Eng JK. Proteomics data repositories. Proteomics 2009 Oct;9(20):4653-63.
6. Vermeulen M, Selbach M. Quantitative proteomics: a tool to assess cell differentiation. Curr Opin Cell Biol 2009 Dec;21(6):761-6.
7. Vinayavekhin N, Homan EA, Saghatelian A. Exploring disease through metabolomics. ACS Chem Biol 2010 Jan 15;5(1):91-103.
8. von der Lieth CW, Bohne-Lang A, Lohmann KK, Frank M. Bioinformatics for glycomics: status, methods, requirements and perspectives. Brief Bioinform. 2004 Jun;5(2):164-78.
9. Orlando R. Quantitative glycomics. Methods Mol Biol 2010;600:31-49.
10. Li J, Richards JC. Functional glycomics and glycobiology: an overview. Methods Mol Biol 2010;600:1-8.
11. Bohne-Lang A, Lang E, Förster T, von der Lieth CW. LINUCS: linear notation for unique description of carbohydrate sequences. Carbohydr Res 2001 Nov 1;336(1):1-11.
12. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977 Dec;74(12):5463-7.
13. Jawhar NM. Tissue Microarray: A rapidly evolving diagnostic and research tool. Ann Saudi Med 2009

81

In Cereo and in Silico: Tissue Microarray (TMA) Techniques and Bioinformatics Are Thriving Forces in Medical Science and Personalized Medicine

Mar-Apr;29(2):123-7.

14. Avninder S, Ylaya K, Hewitt SM. Tissue microarray: a simple technology that has revolutionized research in pathology. J Postgrad Med 2008 Apr-Jun;54(2):158-62.

15. Wang L, Deavers MT, Malpica A, Silva EG, Liu J. Tissue macroarray: a simple and cost-effective method for high-throughput studies. Appl Immunohistochem Mol Morphol 2003 Jun;11(2):174-6.

16. Ma H, Horiuchi KY. Chemical microarray: a new tool for drug screening and discovery. Drug Discov Today 2006 Jul;11(13-14):661-8.

17. Fernandes TG, Diogo MM, Clark DS, Dordick JS, Cabral JM. High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. Trends Biotechnol 2009 Jun;27(6):342-9.

18. Eguíluz C, Viguera E, Millán L, Pérez J. Multitissue array review: a chronological description of tissue array techniques, applications and procedures. Pathol Res Pract 2006;202(8):561-8.

19. Battifora H. The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. Lab Invest 1986 Aug;55(2):244-8.

20. 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 Oct 23;103(1):121-9.

21. Battifora H, Mehta P. The checkerboard tissue block. An improved multitissue control block. Lab Invest 1990 Nov;63(5):722-4.

22. Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 1998 Jul;4(7):844-7.

23. Dunlay RT, Czekalski WJ, Collins MA. Overview of informatics for high content screening. Methods Mol Biol 2007;356:269-80.

24. Dittmar PG, Farmer NA, Fisanick W, Haines RC, Mockus J. The CAS ONLINE search system. 1. General system design and selection, generation, and use of search screens. J Chem Inform Comput Sci 1983;23:93-102

25. Morgan HL. The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. J Chem Doc 1965;5,107-13.

26. Lang E, Förster T, Lieth CW. The integration of the CCDF into the relational information network of the German Cancer Research Center. Gasteiger J , editor. Software development in chemistry 4.

Heidelberg: Springer; 1990. p. 43-50.

27. Camp RL, Neumeister V, Rimm DL. A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. J Clin Oncol 2008 Dec 1;26(34):5630-7.

28. Kennard O, Speakman JC, Donnay JDH. Primary crystallographic data. Acta Cryst 1967; 22:445-9.

29. Motherwell WDS. The CSD-450,000 answers … but what are the questions? Cryst Rev 2008;14: 97-116.

30. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977 May 25;112(3):535-42.

31. Ranzinger R, Frank M, von der Lieth CW, Herget S. Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. Glycobiology. 2009 Dec;19(12):1563-7.

32. Bairoch A. The ENZYME database in 2000. Nucleic Acids Res 2000;28:304-5.

33. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2010 Jan;38(Database issue):D5-16

34. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res 2009 Jan;37(Database issue):D885-90.

35. Conway C, Dobson L, O'Grady A, Kay E, Costello S, O'Shea D. Virtual microscopy as an enabler of automated/quantitative assessment of protein expression in TMAs. Histochemistry and Cell Biology 2008;130,447-63.

36. Conway C, O'Shea D, O'Brien S, Lawler DK, Dodrill GD, O'Grady A; et al. 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.

37. Sharma-Oates A, Quirke P, Westhead DR. TmaDB: a repository for tissue microarray data. BMC Bioinformatics 2005 Sep 1;6:218.

38. Della Mea V, Bortolotti N, Beltrami CA. eSlide suite: an open source software system for whole slide imaging. J Clin Pathol 2009 Aug;62(8):749-51.

39. Thallinger GG, Baumgartner K, Pirklbauer M, Uray M, Pauritsch E, Mehes G, et al. TAMEE: data management and analysis for tissue microarrays. BMC Bioinformatics 2007;8:81.

40. Della Mea V, Bin I, Pandolfi M, Di Loreto C. A web-based system for tissue microarray data management. Diagn Pathol 2006 Oct 11;1:36.

41. Marinelli RJ, Montgomery K, Liu CL, Shah NH, Prapong W, Nitzberg M, et al. The Stanford Tissue Microarray Database. Nucleic Acids Res 2008 Jan;36(Database issue):D871-7.

42. Viti F, Merelli I, Caprera A, Lazzari B, Stella A, Milanesi L. Ontology-based, Tissue MicroArray oriented, image centered tissue bank. BMC Bioinformatics 2008 Apr 25;9 Suppl 4:S4.

43. Stokes TH, Torrance JT, Li H, Wang MD. ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. BMC Bioinformatics 2008 May 28;9 Suppl 6:S18.

44. Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, et al. A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Prote-omics. Mol Cell Proteomics 2005;4(12):1920-32.

45. Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA, Persson A, et al. A genecentric Human Protein Atlas for expression profiles based on antibodies. Mol Cell Proteomics 2008 Oct;7(10):2019-27.

46. Schobesberger M, Baltzer A, Oberli A, Kappeler A, Gugger M, Burger H, et al. Gene expression variation between distinct areas of breast cancer measured from paraffin-embedded tissue cores. BMC Cancer 2008 Nov 25;8:343.

47. Dudley JT, Tibshirani R, Deshpande T, Butte AJ. Disease signatures are robust across tissues and experiments. Mol Syst Biol 2009;5:307.

48. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. Lab Invest 2000 Dec;80(12):1943-9.

49. Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Köchli OR, et al. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. Am J Pathol 2001 Dec;159(6):2249-56.

50. Finak G, Sadekova S, Pepin F, Hallett M, Meterissian S, Halwani F, et al. Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. Breast Cancer Res 2006;8(5):R58.

Correspondence to:

Elke Lang
h_da Darmstadt University
School of Media
Section Information Science & Engineering
Haardtring 100
D-64295 Darmstadt
Germany
Tel: +49 6151 169416
E-mail: Elke.Lang@h-da.de