J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Original Article

Tissue microarray design and construction for scientific, industrial and diagnostic use

Daniela Pilla[1], Francesca M. Bosisio[2], Roberto Marotta[3], Stefano Faggi[4], Paolo Forlani[5], Maurizio Falavigna[5], Ida Biunno[6], Emanuele Martella[4], Pasquale De Blasio[5], Simone Borghesi[7], Giorgio Cattoretti[1,2]

[1]Department of Pathology, San Gerardo Hospital, Via Pergolesi 33, 20900, Monza, [2]Department of Surgical Sciences, Milano-Bicocca State University, via Cadore 48, 20900, Monza, [3]NoemaLife S.p.A., Via Piero Gobetti 52, 40129, Bologna, [4]Nikon Instruments S.P.A, Via Meucci 59, 50041 Calenzano, Firenze, [5]Integrated System Engineering srl, Via Fantoli 16/15, 20138 Milano, [6]Institute for Genetic and Biomedical Research, National Research Center, Via Fantoli 16/15, 20138 Milan, [7]Department of Mathematical Sciences (Matematica e Applicazioni), Milano-Bicocca State University, Via Cozzi 53, 20125, Milano, Italy

E-mail: *Giorgio Cattoretti - giorgio.cattoretti@unimib.it
*Corresponding author

Received: 13 July 12 Accepted: 03 October 12 Published: 20 December 12

This article may be cited as:
Pilla D, Bosisio FM, Marotta R, Faggi S, Forlani P, Falavigna M, Biunno I, Martella E, De Blasio P, Borghesi S, Cattoretti G. Tissue microarray design and construction for scientific, industrial and diagnostic use. J Pathol Inform 2012;3:42.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2012/3/1/42/104904

Abstract

Context: In 2013 the high throughput technology known as Tissue Micro Array (TMA) will be fifteen years old. Its elements (design, construction and analysis) are intuitive and the core histopathology technique is unsophisticated, which may be a reason why has eluded a rigorous scientific scrutiny. The source of errors, particularly in specimen identification and how to control for it is unreported. Formal validation of the accuracy of segmenting (also known as de-arraying) hundreds of samples, pairing with the sample data is lacking. **Aims:** We wanted to address these issues in order to bring the technique to recognized standards of quality in TMA use for research, diagnostics and industrial purposes. **Results:** We systematically addressed the sources of error and used barcode-driven data input throughout the whole process including matching the design with a TMA virtual image and segmenting that image back to individual cases, together with the associated data. In addition we demonstrate on mathematical grounds that a TMA design, when superimposed onto the corresponding whole slide image, validates on each and every sample the correspondence between the image and patient's data. **Conclusions:** High throughput use of the TMA technology is a safe and efficient method for research, diagnosis and industrial use if all sources of errors are identified and addressed.

Key words: Diagnostic, pathology, tissue microarray, tissue micro array, virtual slide, whole slide image

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.104904

Quick Response Code:

INTRODUCTION

In 1998 Kononen and colleagues published in Nature Medicine,[1] the results of a high throughput analysis of tumor cases, named Tissue Micro Array (TMA) technology, via an ingenious substantial modification of a previous method,[2] of transferring cores from several donor blocks into a virgin unique acceptor as replicated samples identified by the position in the rows and columns of the array. Three thousand and counting publications later, the TMA technology has fulfilled the promises of high throughput analytical ability.

The inbuilt geometry of the design and the unsophisticated execution has favoured a practical approach: Common sense, histotechnology expertise and unwritten tips rather than written standards.

It is therefore not surprising that the intrinsic analytical error rate has never been addressed and the standards for design and construction, either for diagnosis or for research, have never been defined, except for cooperative large-scale studies,[3] or for the data exchange specification.[4-6] Notably, none of these methodology suggestions addresses how to control for errors in matching individual samples on the slide with its own data or how safely track those data, except for a quote about asymmetry in order to prevent map orientation confusion.[3] Published data about the analytical error rate in TMA are lacking; unpublished estimates suggests that 1 in 5,000 to 10,0000 samples may be misidentified, even using printed blocks and barcode-driven sample IDs (Stephen M Hewitt, personal communication). More frequently, lack of correspondence between cored-out samples and the original block are attributed to tissue heterogeneity, totally ignoring the contribution of sample switching.[7] The technique is evidently assumed to be totally accurate in identifying each and every single spot; the occurrence of tissue heterogeneity and the high-throughput ability masks and overcomes small individual error-driven experimental noise. Therefore formal demonstration that superimposing a TMA design onto the TMA slide image results in validation of the individual components has never been made.

As a consequence, reports of diagnostic use of the TMA, where positive identification of a patient is crucial, are scarce.[8,9]

This report addresses and solves error control issues and provides scientific ground to the use of the TMA technology for diagnostic but for industrial or scientific purposes as well.

SUBJECTS AND METHODS

Setup of Barcode-Driven Error Control

The Laboratory Information System (LIS) in use in our department (Athena, Noemalife SpA, Bologna, Italy) identifies each object with an unique identification (ID) number, including an 8 digit numeric slide ID. Patient and specimen data are entered into the LIS, creating unique IDs for accession number, procedure, specimen, tissue block, tissue slides, etc. The LIS interface was modified to print its unique ID onto the label of each slide as a Code 128C high density linear barcode, via a label printer (Datamax M4206). We chose to use the tissue slide instead of the tissue block for barcoding because the slide is the item on which the pathologist marks areas to be sampled. Matching the slide with the corresponding block is a standard procedure in histology, which entails two separate and qualitatively distinct procedures totalling 6 seconds or less: Matching first the complete IDs on the label and on the block, then matching the shape of the tissue on the slide with the one on the corresponding block. The pathologist would chose which sample and which area of the sample would be arrayed and a set of marked, barcoded slides would be forwarded to the technician in charge of arraying.

We next generated a query on the LIS database (based on Oracle 10.2), which extracts from the database five items (slide ID, block number, case number, patient ID and date in which the sample was received), when prompted by the ID scanned on the tissue slide [Figure 1]. The query is exported as an Excel file, to be acquired by the Tissue Arrayer software in the Donor Block database. Note that none of the records we chose to be inserted in the file allows identification of the patient by a third unauthorized party. The amount of information extracted from the LIS is limited only by the dimension of the field in the database.

Barcode reading was accomplished with a common hand held linear barcode reader.

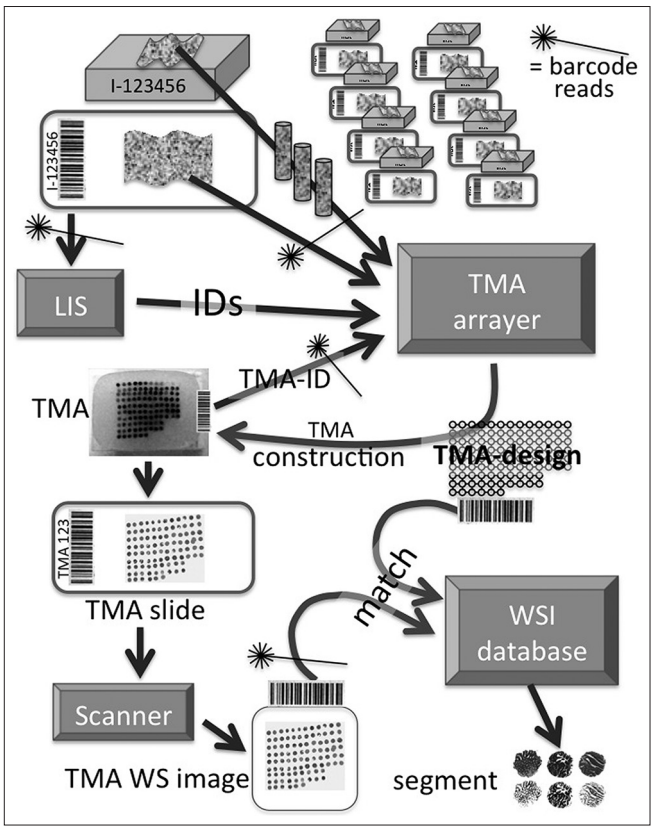

Figure 1: TMA construction through individual case segmentation flow process The figure depicts the flow of samples, images and data through the process. Note that there is no access to the WSI database which does not pass through a barcode reader check (the wand symbol). A final reconciliation of the data contained in the design with the individual cores in the WSI is done by barcode-driven match between the WSI and its own unique design

The file containing the 5 informative items of the donor tissue, is drag into the open window of the Galileo TMA program (Tissue Microarrayer Model TMA Galileo CK4500; Integrated Systems Engineering srl, Milano, Italy), populating automatically all the donor block fields. The unique slide ID is the identifying field for the software.

Next, before the design of the TMA begins, the Galileo software requests that the ID of the TMA acceptor block is inserted; this is accomplished by barcode scanning a 5 digit alphanumeric unique TMA ID, previously generated by a TMA database separated from the LIS, printed and inserted into the back of the virgin acceptor block [Figure 1]. The same barcode ID is then printed together with alphanumeric date, stain, etc on a label glued on each section cut from that TMA [Figure 1].

Asymmetric TMA design

The TMA array is purposely designed with an intrinsic asymmetry, i.e. transformation of the geometry by reflection along the axis generate new, readily distinguishable geometric forms. Care is taken so that the asymmetry produced by inserting or emptying spots is not lost upon accidental loss of single spots or rows on the slide.

During the TMA design construction with the Galileo software, the design is a bi-dimensional matrix of $n \times m$ positions, in which each full spot may be considered "1" and the empty one "0". We define as i and j the indexes identifying the matricial design and a () each element, which is then $a\ (i,j)=0 \mid 1$ where $0<i<m$ e $0<j<n$. By applying to the matrix each of the 7 possible transformations [Table 1], it is possible to calculate the robustness of the asymmetry by calculating the ensuing superimpositions of full and empty spots and ranking the safety of the design. A simple computer program can easily loop through all the possible transformations and count the number of spots that do not overlap after each change. The smallest of these numbers is the overall rank of the array. A rank of 0 discards the design because there is at least one symmetric transformation, thus a chance of error [Figure 2].

The source of asymmetry is consistently placed in one fixed position (e.g. lower right corner), to facilitate the correct orientation of the TMA image before segmenting.

The use of random placement of the samples makes the use of sectors unnecessary; in addition, empty rows or lines within the design run the risk of misidentify spots in case the TMA section is particularly wavy or disrupted.

Exporting Data from the TMA Software into the WSI Scanner

Upon TMA completion, two types of files are generated: An excel file containing the TMA design, the list of all samples and relative data and an xml file, in which the same data are coded in a format readable by the software of the whole slide scanner. The Galileo TMA software natively generates files in the TMA DES (Data Exchange Specification) XML format.[5]

Both files contain an identification field with the TMA ID.

The xml file is imported in Spectrum (Aperio Technologies, Vista, CA, USA) in a designed database, and each xml tag is addressed to a designed field in the database.

Table 1: Listing of the possible transformations of the TMA design consisting of a square array, a design less favourable than a rectangular one

Transformation number	Type	Formula
0	Identity	b(i, j)=a(i, j)
1	Clockwise 90° rotation	b(i, j)=a(j, n+1-i)
2	Clockwise 180° rotation	b(i, j)=a(n+1-i, n+1-j)
3	Clockwise 270° rotation	b(i, j)=a(n+1-j, i)
4	Mirror inversion	b(i, j)=a(n+1-i , j)
5	Mirror inversion+Clockwise 90° rotation	b(i, j)=a(n+1-j, n+1-i)
6	Mirror inversion+Clockwise 180° rotation	b(i, j)=a(i, n+1-j)
7	Mirror inversion+Clockwise 270° rotation	b(i, j)=a(j, i)

Note: a(i, j) are the elements of the original array and b(i, j) are the elements of the transformed array

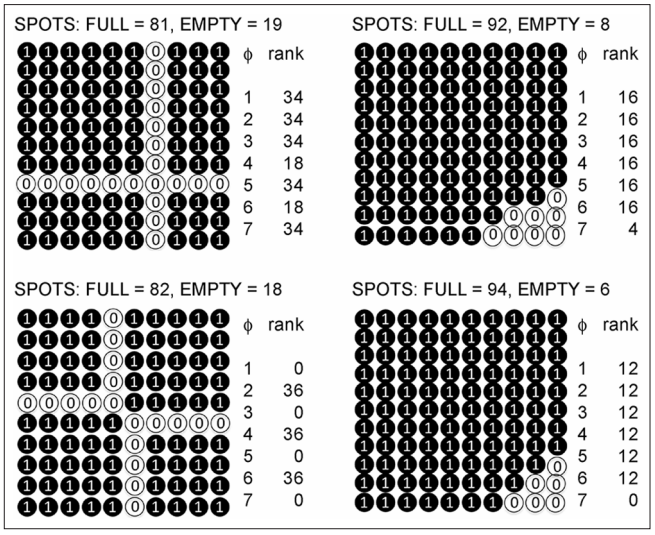

Figure 2: Ranking the asymmetry of the TMA design The figure depicts pair wise comparisons (top, bottom) of TMA design with sectors (left) and solid (right). In all of them, empty spots are marked as "0", filled ones as "1". The number of filled and empty spots is noted above each design. Left of each design are listed the seven transformations (φ) and the calculated rank of that transformation (see text). A rank of "0" in any permutation will result in discarding the design, because of symmetry

The Aperio slide scanner acquires the whole slide image (WSI) of the TMA and the TMA barcode on the label [Figure 1] and stores both in the spectrum database. Additional information can be added to the software by the investigator, before or after the image analysis.

Automated Matching of the Design with the WSI

To ensure exact matching of all patient's ID data with the corresponding sample images, the Spectrum software was modified so that the match of the ID in the design and the ID on the WSI [Figure 1] is highlighted, leaving to the operator the choice to hit the "segment" command. Without a match, no pre-made matching is visualized, alerting the operator of an error.

The segmentation process begins by a superimposition of the design image to the WSI. Once the correct rotation of the WSI is defined by placing the source of asymmetry in the pre-defined corner, the grid of the design is adjusted to accommodate the shearing that may take place when the 4 μm section is placed on the glass slide and to fit all the available spots on the WSI.

RESULTS

Validation of the Matching Between the Data in the TMA Design and Individual Images of the Cores on WSI

We took a mathematical approach to the question on how to validate matching the grid containing the patient's ID and data (the TMA construction design) with the image of the TMA stained section. An array is shown to be "strongly asymmetrical" (intuitively, it has no nontrivial isometries), provided it satisfies two, easy to check, conditions.

We set the following standard notation: $\mathbb{R}$ is the real line, $\mathbb{R}^2$ the plane $\mathbb{R} \times \mathbb{R}$ endowed by the Euclidean distance, $[a, b]$ is the straight segment from the point a to the point b in $\mathbb{R}^2$. We prove the following statement:

Let R be a finite, hence discrete, set of points in $\mathbb{R}^2$, such that the two following two properties hold.

1. Let l_{pq} be the line through the points p, q; consider the set S of all the segments $[p, q]$, $p, q \in$ R with the property that one of the two connected components of $\mathbb{R}^2 - l_{pq}$ has empty intersection with $\mathbb{R}$. Then we require the existence of a segment $[a, b] \in S$ whose length is unique among the lengths of the other segments in S.
2. Let $[a, b]$, $a, b \in R$ be the unique segment of length n in S. Then R should not be symmetrical with respect to the axis perpendicular to $[a, b]$ through its midpoint.

Then if ϕ: $\mathbb{R}^2 \to \mathbb{R}^2$ is an isometry such that $\phi(R) = R$, ϕ must be the identity function. Any finite set of points satisfying this last statement will be called strongly asymmetrical.

We recall than an isometry in $\mathbb{R}^2$ is a function f such that $d_E(x, y) = d_E(f(x), f(y))$ for all $(x, y) \in \mathbb{R}^2$, where d_E is the Euclidean distance in $\mathbb{R}^2$. They include (and are in fact generated by) all reflections, rotations and translations. This statement is false if we require ϕ to be a homeomorphism (i.e. bijective and bicontinuous function) only.

Proof: Being ϕ an isometry, it sends lines in lines, thus $\phi([a, b])$ is a segment connecting the two points $\phi(a)$ and $\phi(b)$ in R. Because of the continuity of ϕ, there exists a connected component of $\mathbb{R}^2 - l_{\phi(a)\phi(b)}$ with empty intersection with R, since so does $\mathbb{R}^2 - l_{ab}$. Therefore $[\phi(a), \phi(b)]$ belongs to the set S, as we are assuming $\phi(R) = R$. The length of such a segment must be the same as the length of $[a, b]$, again because ϕ is an isometry. By the unicity of the segments in S of length n (condition (1) on R), it must be $[\phi(a), \phi(b)] = [a, b]$ and ϕ restricted to $[a, b]$ is an isometry onto $[a, b]$. Either of the following cases must necessarily happen:

A. $\phi(a) = a$ and $\phi(b) = b$. Then ϕ restricted to the line l_{ab} is an isometry onto l_{ab} itself that fixes two points, therefore it must be the identity on the same line. This leaves only one non trivial possibility for ϕ as map on $\mathbb{R}^2$: a reflection along the l_{ab} axis. But in that case, it is impossible that $\phi(R) = R$, since ϕ would swap the connected components of $\mathbb{R}^2 - l_{ab}$ and the points in R not in l_{ab} are all contained in one component. This leaves ϕ = identity on the whole plane $\mathbb{R}^2$ as the only possibility.
B. $\phi(a) = b$ and $\phi(b) = a$. Then ϕ must have at least a fixed point, which must necessarily be the middle point m of $[a, b]$ since ϕ is assumed to be an isometry. Thus it must either be a rotation by π around m, in which case $\phi(R) \neq R$ because ϕ swaps the connected components of $\mathbb{R}^2 - l_{ab}$, or else a reflection through the line perpendicular to $[a, b]$ through m. However, the condition (2) on R implies that $\phi(R) \neq R$ for such an isometry.

This shows that the only possible isometry ϕ of $\mathbb{R}^2$ such that $\phi(R) = R$ is the identity function.

Think of an array of specimens placed in glass slide as the finite set R satisfying the conditions (1) and (2). Assume the shape of the slide to be unknown and that the array is placed randomly on it. This result implies that, choosing any labelling of the specimens and applying an unknown sequence of integrity-preserving transformations to the slide, to recover the initial labelling of the specimens it suffices to move the slide around until the shape of the array matches the original one.

The conditions (1) and (2) are isometry invariant, meaning that, they hold for a set R if and only if they hold for $\psi(R)$, where ψ is any isometry of $\mathbb{R}^2$. As a

consequence, the array of specimens, when glued on the slide, will be strongly asymmetrical, provided the initial (computer generated) one was and all the processing involved in the slide preparation can be assumed to transform the array only just through isometries.

The easiest examples of lattices satisfying the conditions (1) and (2) are rect-angular ones with sides lengths $c \geq d$ with an extra specimen added in a length c boundary row. For that array we can take n to be $c + 1$ and immediately verify that the conditions are fulfilled. Appropriately removing blocks of points from one corner of a rectagular array, with sufficiently many rows and columns, provides an example of strongly asymmetrical array which preserves this property in case of deletion of a few columns, rows and points.

This demonstration is graphically summarized in Figures 3a and 3b.

Implementation and Costs of Diagnostic TMA

The implementation of a barcode-driven error control of the design and execution of a TMA for patient diagnosis requested approximately € 4,000.00 to modify the LIS program, in order to print the slide ID as a linear barcode in addition to the customary alphanumeric identifiers. The generation of the query to extract from the LIS a sample data by barcode scanning costs approx. € 3,000.00 in programmer's time. The modification of the Aperio Spectrum segmentation interface to match the IDs on the slide and in the database was part of this project.

The speed of the TMA construction and the workflow did not change apparently after the implementations were put in place, mostly because of the deep automation already in place; it actually gained speed and smoothness of operation by leaving to the machine the tedious but necessary task of sample checking.

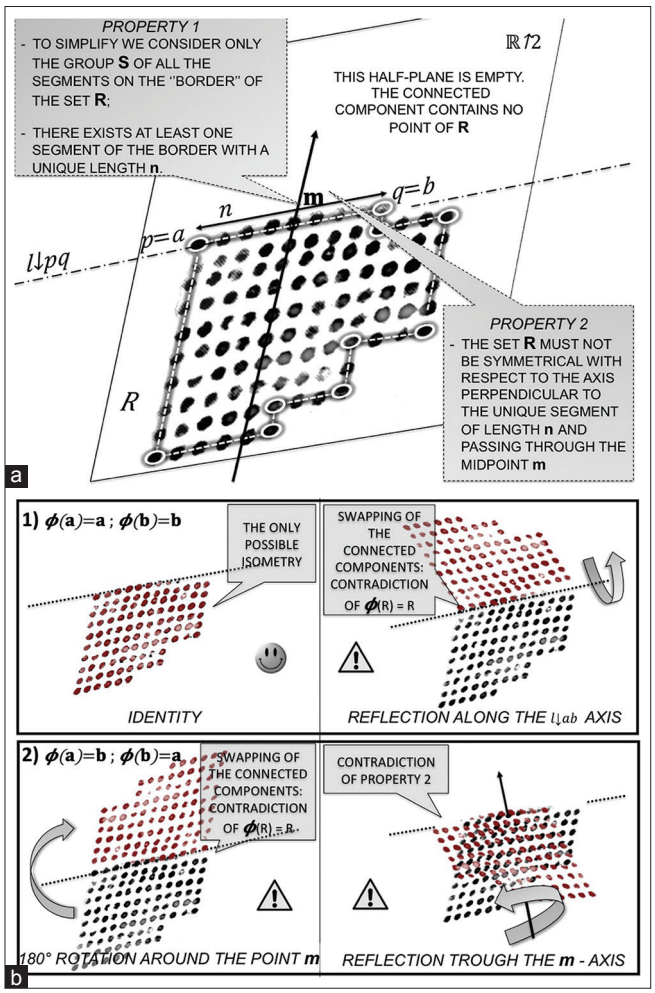

Figure 3: Validation of the TMA design and the WSI match (a) is a graphic rendering of the two properties set forth before analyzing the match between the TMA design and the WSI; (b) represents the four possible isometric alignments of the TMA design and the WSI, only one of which represents the identity of both

DISCUSSION

The focus of our work is on areas which have been taken as granted or obvious and therefore not scientifically investigated such as asymmetry of the design, requirements of the TMA design layout, matching of the design with the image and validation of the correspondence between data and image for individual spots.

We have sought to control and minimize the source of ID errors through the process. This aspect has not been addressed before because of the high data yield achievable with the TMA technique, with reduces the impact of sample heterogeneity (and for that matter, switching individual samples) on the significance of the results within large cohorts of patients.[10,11] Switching individual samples or patients has little bearing if one looks e.g. at the impact of a phenotype on prognosis, particularly if no clinical decisions have to be made on each and every single patient within that cohort.

To obtain error control we have used barcode readings throughout the entire process [Figure 1]. Scanning a linear barcode is not error-free,[12] however it considerably reduces the source of keystroke errors.

Random placement of core replicas within the TMA, a recommended practice,[3] requires a compact design. This makes redundant the use of sectors, introduced both to shape the asymmetry and to facilitate the visual identification of different groups of samples within the TMA. Sectors use up TMA space and do not necessarily guarantee asymmetry (see examples in Figure 2). The randomness of core placement further highlights the obligate handling of the TMA on WSI scanners: Traditional microscopes do not apply anymore.

One additional question we addressed was the validation of the correspondence between the ID and the image of individual cores, obtained upon matching the whole

design geometry with the geometric matrix composed of core sections images. We applied the laws of geometry, which states that once the vectorial transformation of an intrinsically asymmetric matrix (e.g., the WSI) is coincident with another given matrix (e.g., the TMA design) the vector pointing to an individual component has the same value, thus the points coincide and the ID of each and every component is validated. A graphic rendering of the grid, which includes lines connecting all the individual positions in the grid, is of paramount importance to verify the isometries while adapting the grid upon the TMA image.

There are other issues which we did not address; one is the amount of human intervention needed during the segmentation (or "de-arraying") procedure and the efficiency required. This issue has been examined by a handful of papers, one of which,[13] compares a custom made de-arraying system with the one which has been used for the present study. However, equivocal placement of individual cores may still be judged by a human eye.

Another issue we did not solve is how and where a robust asymmetry should be enforced during the process. Our feeling is that asymmetry is too important for safely identify individual samples/patients and that it should be enforced at some point.

One last non-preventable source of analytical error may occur when the combination of loss of cores from the periphery of the section and the deformation of the paraffin sheet upon collecting the section from the water bath results in an equivocal assignment of a core to a row or a line. These cores should be discarded from the analysis. If each sample is represented twice or three times within the TMA,[3] there should be no loss of analysis for that specimen.

Lastly, whether is done automatically or under human supervision, the asymmetry of the design should be visible in the grid superimposed to the WSI before segmenting. Ideally, spots left empty for the asymmetry should be automatically recognized as an additional safeguard against matching the wrong grid with the wrong WSI.

CONCLUSIONS

We have applied for the first time a scientific approach to the ID error control to the construction and data analysis with the TMA technology, we have validated the process of identifying the WSI of single samples/patients in the TMA, setting diagnostic standards for the use of the TMA. While all these aspects seem redundant for the research use of the TMA technology, they are mandatory for any use of it which involves patient's care. However, controlling for analytical errors in experiments in science and industry is a golden standard and we advocate the use of our solutions in all fields using the TMA high throughput technology.

ACKNOWLEDGEMENT

We wish to thank Kate Lillard and Karen Gustashaw (Aperio) for support and encouragement during the project, Maria Cristina Argentieri, Angelita Ferri and the staff of the Pathology Department for help and suggestions. The Tissue Micro Arrayer and the Aperio Scanscope were provided through a grant from the Regione Lombardia (Call for Independent Research, DDG 6716 del 1/7/2009). Mrs Argentieri was funded by a generous donation by Prof. Franco Uggeri (UNIMIB).

REFERENCES

1. 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;4:844-7.
2. Battifora H. The multitumor (sausage) tissue block: Novel method for immunohistochemical antibody testing. Lab invest 1986;55:244-8.
3. Rimm DL, Nielsen TO, Jewell SD, Rohrer DC, Broadwater G, Waldman F, *et al.* Cancer and leukemia group B pathology committee guidelines for tissue microarray construction representing multicenter prospective clinical trial tissues. J Clin Oncol 2011;29:2282-90.
4. Berman JJ, Datta M, Kajdacsy-Balla A, Melamed J, Orenstein J, Dobbin K, *et al.* The tissue microarray data exchange specification: Implementation by the cooperative prostate cancer tissue resource. BMC Bioinformatics 2004;5:19.
5. 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.
6. 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;130:1004-13.
7. Wärnberg F, Amini RM, Goldman M, Jirström K. Quality aspects of the tissue microarray technique in a population-based cohort with ductal carcinoma *in situ* of the breast. Histopathology 2008;53:642-9.
8. Sapino A, Marchiò C, Senetta R, Castellano I, Macrì L, Cassoni P, *et al.* Routine assessment of prognostic factors in breast cancer using a multicore tissue microarray procedure. Virchows Arch 2006;449:288-96.
9. Thomson TA, Zhou C, Ceballos K, Knight B. Tissue microarray for routine clinical breast biomarker analysis. The British Columbia Cancer Agency 2008 experience. Am J Clin Pathol 2010;133:909-14.
10. Nocito A, Bubendorf L, Tinner EM, Süess K, Wagner U, Forster T, *et al.* Microarrays of bladder cancer tissue are highly representative of proliferation index and histological grade. J Pathol 2001;194:349-57.
11. Camp RL, Neumeister V, Rimm DL. A decade of tissue microarrays: Progress in the discovery and validation of cancer biomarkers. J Clin Oncol 2008;26:5630-7.
12. Snyder ML, Carter A, Jenkins K, Fantz CR. Patient misidentifications caused by errors in standard bar code technology. Clin Chem 2010;56:1554-60.
13. Wang Y, Savage K, Grills C, McCavigan A, James JA, Fennell DA, *et al.* A TMA de-arraying method for high throughput biomarker discovery in tissue research. PLoS One 2011;6:e26007.