A spatial statistical framework for the parametric study of fiber networks: application to fibronectin deposition by normal and activated fibroblasts

Short title: Statistical analysis of parametric ECM fiber networks

Anca-Ioana Grapa\textsuperscript{1,2,\#a}, Georgios Efthymiou\textsuperscript{2,\#b}, Ellen Van Obberghen-Schilling\textsuperscript{2}, Laure Blanc-Féraud\textsuperscript{3}, Xavier Descombes\textsuperscript{1}*  

\textsuperscript{1}Université Côte d’Azur, INRIA, CNRS, i3S, Nice, France  
\textsuperscript{2}Université Côte d’Azur, CNRS, INSERM, iBV, Nice, France  
\textsuperscript{3}Université Côte d’Azur, CNRS, INRIA, i3S, Nice, France

\#aCurrent Address: Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK  
\#bCurrent Address: Centre de Recherche en Cancérologie de Marseille, U1068, INSERM, Marseille, France

* Corresponding author

E-mail: xavier.descombes@inria.fr (XD)

Keywords: Fibronectin, Extracellular matrix, Oncofetal isoforms, Statistical parametric maps, Graph networks
Abstract

Changes in the spatial landscape of the extracellular matrix (ECM) in health and disease significantly impact the surrounding tissues. Quantifying the spatial variations in the fibrillar architecture of major ECM proteins could enable a profound understanding of the link between tissue structure and function. We propose a method to capture relevant ECM features using graph networks for fiber representation in normal and tumor-like states of 4 alternatively spliced isoforms of fibronectin (FN) associated with embryonic development and disease. Then, we construct graph-derived statistical parametric maps, to study the differences across variants in normal and tumor-like architectures. This novel statistical analysis approach, inspired from the analysis of functional magnetic resonance imaging (fMRI) images, provides an appropriate framework for measuring and detecting local variations of meaningful matrix parameters. We show that parametric maps representing fiber length and pore orientation isotropy can be studied within the proposed framework to differentiate among various tissue states. Such tools can potentially lead to a better understanding of dynamic matrix networks within the tumor microenvironment and contribute to the development of better imaging modalities for monitoring their remodeling and normalization following therapeutic intervention.

Author Summary

Due to the complex architectural diversity of biological networks, there is an increasing need to complement statistical analyses with a qualitative and local description. The extracellular matrix (ECM) is one such network for which fiber arrangement has a major impact on tissue structure and function. Thus, a flexible numerical representation of fibrillar networks is needed for accurate analysis and meaningful statistical comparison of ECM in healthy and diseased tissue. First, we propose a versatile computational pipeline to study fiber-specific features of the ECM with graph networks. Then, we introduce a novel framework for the statistical analysis of graph-derived parametric maps, inspired from the statistical analysis of fMRI parametric maps. This analysis is useful for the quantitative/qualitative comparison of ECM fiber networks observed in
normal and tumor-like, or fibrotic states. These methods are applied to study networks of fibronectin (FN), a provisional ECM component that dictates the organization of matrix structure. From 2D confocal images we analyzed architectural variations among 4 alternatively spliced isoforms of FN, termed oncofetal FN, that are prevalent in diseased tissue. We show how our approach can be used for the computation and statistical comparison of heterogeneous parametric maps representing FN variant-specific topological/geometrical features. These methods may be further developed and implemented into tumor tissue ECM profiling to decipher the specific roles of ECM landscapes and their remodeling in disease.

Introduction

Organs and tissues are composed of a non-cellular stromal compartment known as the extracellular matrix (ECM), a cell-derived bioregulatory scaffold that provides chemical and mechanical cues to surrounding cells. The composition and architecture of the ECM is tissue- and organ-specific, and depends on the pathophysiological state of the tissue (i.e. normal vs diseased) (1). For example, while healthy interstitial tissues display a loose meshwork-like ECM, their fibrotic and cancerous counterparts are characterized by the presence of dense aligned ECM fibrils. Thus, the physical and structural traits of the tumor matrix have recently drawn much attention as cancer hallmarks and potential therapeutic targets (2)(3). Collagen, among the most prominent matrix components, has been extensively investigated in this context and several studies addressing its structural features and their association to cancer progression, metastasis and treatment have been published (4)(5)(6). Collagen deposition, however, depends on another key component of the ECM, namely fibronectin (FN). FN is a dimeric glycoprotein that forms a provisional matrix framework to which other ECM components integrate to generate a mature ECM (7)(8).

FN exists in two distinct forms: plasma FN (pFN) that is constitutively produced by the liver and circulates at high concentrations in blood plasma, and cellular FN (cFN) which is expressed and deposited locally in tissue by specialized cells in a spatially and temporally restricted manner.
Indeed, cFN is highly expressed during embryogenesis and in certain physiological and pathological conditions including inflammation, wound healing, and cancer. Structurally, the two forms differ by the presence in cFN of alternatively spliced sequences of about 90 amino acids, termed Extra Domain B (EDB) and Extra Domain A (EDA), that enhance FN assembly, impact matrix architecture, and tune cellular responses (9).

Despite the pivotal role of FN and its variants in health and disease (10), comprehensive studies of FN fiber features have not been reported. To that end, we set out to develop a robust pipeline of numerical analyses for the extraction of biologically relevant metrics from confocal microscopy images of FN variant-specific matrices. First, to capture and study the relevant ECM fibrillar features, we developed a graph-based approach that recapitulates the fiber networks. Starting from these graph representations of FN variants, we then designed a framework to statistically compare ECM geometrical features using different parametric maps, such as fiber length and pore orientation isotropy. Using these approaches, we show here that significant parametric differences across variants in health and tumor-like states can be quantified and simultaneously localized.

**Results and Discussion**

**Generation of FN variant-specific cell-derived matrices and induction of a tumor-like state**

Fibroblasts are the major ECM-producing cells of tissues. During tumor progression, quiescent fibroblasts become activated by environmental cues that induce their phenotypic conversion from a “normal” to a “pro-tumoral” state (Fig 1A, top). This process is characterized by actin re-organization and increased contractility of cells that result in their elongation and the deposition of a highly anisotropic ECM (9). In vitro, these changes can be mimicked by treating normal fibroblasts with Transforming Growth Factor-β (TGF-β), a potent cytokine involved in fibroblast activation in the tumor microenvironment (9) (Fig 1A, bottom). For our analyses, normal or tumor-like FN-variant specific matrices were generated by presenting FN-null mouse embryo fibroblasts with recombinant FN containing one, both (or none) of the Extra Domains (prepared
as previously described (9)), in presence or absence of TGF-β, as schematized in Fig 1B. Cell
cultures were decellularized after 7 days, and matrices were visualized by immunofluorescence
staining and confocal microscopy. The organization of variant-specific FN matrices in normal and
activated states (Fig 1C) was then quantitatively analyzed, as described below.
**Fig 1.** (A) (top row) Schematic representation of an epithelium displaying the different architectures of underlying fibroblasts and their ECM in normal and tumor settings (bottom row). Phase contrast images of cultured FN1-null mouse fibroblasts treated as indicated with 5 ng/ml TGF-β. Scale bar, 100 μm. (B) Workflow diagram featuring the linear structure of the purified recombinant FN variants and relative positions of the alternatively spliced Extra Domains (EDB, EDA), the generation of cell-derived matrices, and image acquisition and analysis. (C) Immunofluorescence staining of FN in variant-specific cell-derived matrices. FN-null mouse fibroblasts were presented with the FN variants (15 μg/ml) in the presence or absence of TGF-β.
(5 ng/ml). After removal of the cells, matrices were stained with a rabbit-anti-FN monoclonal antibody and visualized with confocal microscopy. Scale bar: 50 μm.

Fiber detection using Gabor filters and graph extraction

Fiber detection for pattern quantification in FN networks was produced using a pipeline, briefly described in a previous study (1), where it was primarily utilized for the extraction of local topological fiber properties. In the present study, the workflow was extended to enable a statistical analysis of parametric maps, focusing on a global spatial statistical analysis, which is relevant for quantitative and qualitative fiber analysis across multiple scales. The computational pipeline can be tested with a graphical user interface (GUI) on single samples [e.g., immunofluorescence (IF) coupled with confocal microscopy, immunohistochemistry (IHC)], or a batch of images, the output of which can be written into .png image files and fiber specific graph/Gabor-derived features collected in .csv files. The MATLAB source code and sample images can be found on GitHub, at github.com/aigrapa/ECM-fiber-graph. The steps involved in producing a 2D graph-based fiber representation from confocal microscopy images are presented in Fig 2. First, to detect and enhance the fibrillar structures from the 2D confocal images, we selected a bank of Gabor filters (2) tailored to capture fiber elements that appear at multiple frequencies and orientations. The intrinsic parameters of a Gabor filter (S1 Text) are directly linked to corresponding physical properties (i.e. fiber thickness, fiber orientation). As a result of the filter bank convolution, the intensity of each pixel of the enhanced fibers corresponds to the Gabor filter with the highest coefficient (Fig 2). This flexible, yet meaningful detection property is an important asset over alternative multi-resolution techniques commonly used for curvilinear object detection, such as discrete curvelets (3,4), whose complex discrete implementation makes them prone to translation and rotation issues (e.g. energy dissipation in neighboring sectors in the frequency domain). In subsequent steps of the graph-based representation pipeline, a series of image morphological transformations were applied after fiber enhancement, including binarization using hysteresis thresholding and fiber morphological skeleton computation. Converting the binary fiber structure into a morphological skeleton is
useful to generate a simplified representation that preserves the topological properties of the original region and allows a successive skeleton-graph association.

Graphs (i.e. collections of nodes connected by edges) are powerful tools for the structural and pattern analysis of objects, commonly employed for the mathematical study of relations between entities (5). In our work, we propose to use a graph-based framework to depict morphological fiber skeletons, and to provide a geometrical characterization of fiber patterns. Other studies employ graphs to analyze different biological networks, such as osteocyte networks (6) and retinal vasculature (7), illustrating the remarkable potential of spatial graphs for fiber-like object analysis. Within the graph-based skeleton representation, the nodes represent either fiber crosslinks (perceived in the 2D projection of the network onto the image plane) or fiber ends, while the edges capture the fiber length between two given nodes. An additional process of node reconnection to correct for missing fibers due to artefacts from previous morphological operations and noise in the data was deemed necessary. The steps to achieve this correction are detailed in (S1 Text). As previously mentioned, the current pipeline has shown potential for the description of distinct local features among the four FN images of variants, in normal state (1).

An additional step was added to the graph-based workflow to further simplify fiber delineation. Starting from the skeleton graph, we kept all nodes corresponding to fiber extremities, and connected all pairs of nodes with a straight line, if a fiber had previously been identified. For the sake of simplicity, we refer to fiber length as the length of any straight line connecting a pair of graph nodes. This characterization can be encoded in adjacencies matrices (indicating the presence/absence) of a certain edge between the nodes. We note that both representations are useful to extract different local or global fiber properties. The graph-based skeleton fiber delineation (Fig 2) faithfully represents (according to a visual assessment performed by a trained biologist) the geometrical and topological properties of the fibers from the 2D confocal images, while the Gabor-specific (e.g. fiber local orientation, thickness) and graph-derived parameters (e.g. fiber length, number of nodes, etc.) are linked to meaningful physical fiber attributes. This biologically relevant representation enabled us to develop a statistical analysis of the variation of certain fiber parameters for both the "normal" and a "tumor-like" state of the FN network.
Fig 2. Fiber enhancement and graph-based representation starting from confocal 2D images. (A) Confocal microscopy of variant-specific FN matrices. (B) Convolution with a set of Gabor filters. (C) Fiber enhancement. (D) Morphological skeleton of fibers. (E) Skeleton and fiber-associated graph.

Image size is set at 1024x1024 pixels, confocal acquisition is performed at a resolution of 0.27 µm/pixel. (B) Fibers are enhanced with Gabor filters and the maximum response is kept in every pixel, as shown in (C). The morphological fiber skeleton is subsequently extracted (D) and a fiber-
associated graph is assigned \((E)\). Furthermore, a simplified graph representation is derived from the skeleton graph.

**Statistical analysis of parametric maps**

The graph-based representation of FN networks enabled the subsequent design of a novel framework to perform a spatial statistical analysis of ECM patterns, using graph-derived statistical parametric maps. Fiber attributes computed from graphs can be considered for creating these variation maps. To illustrate an application of our method, we considered two different types of parametric maps reflecting meaningful geometrical features, namely the individual fiber lengths (i.e. the length of the connecting line between two graph nodes), and the fiber pore (“gap”) orientation isotropy (i.e. the inverse value of the absolute difference between the median and individual pore orientation). We relied on a local computation of parameters to build these maps: fiber length along graph edges or fiber gap pore orientation, and subsequent extrapolation of these values to build dense maps. Fig 3 illustrates an example for each map category. Details of the computation of parametric maps starting from graph-based representations can be found in (S1 Text).

Inspired from the statistical parameter map (SPM) framework in fMRI analysis (8,9) using Gaussian Random Fields (GRF) (10), the proposed spatial statistical learning framework aims to quantify and simultaneously localize statistically significant differences across the normal and tumor-like FN network maps. These differences are assessed both quantitatively and qualitatively, as detected anomalies with respect to the GRF, corresponding to regions within the maps that cannot be explained by the normal model. Providing both a quantitative and a qualitative assessment of the parametric variations is extremely important for the study of spatial heterogeneity of fibers in pathological conditions. Here we apply our proposed framework to determine if fiber length and pore orientation isotropy can discriminate between normal and tumoral states, and among the distinct cFN variant networks. While topographical differences
between ECM of healthy and tumor tissue have been described (4), no current computational study can, to our knowledge, simultaneously localize and quantify them.

The principle of the current methodology (Fig 4) relies on modelling the normal FN maps as realizations of a GRF within a hypothesis testing framework (12) (11). We hypothesized that the tumor-like maps are realizations of the same GRF, and we determined a set of probabilities that characterize a degree of belonging to the GRF, of certain contiguous regions (clusters) at various intensity thresholds. In other words, the current statistical analysis identifies those foreign regions with respect to the GRF, within both types of maps, under the null hypothesis (i.e., at a given p-value (pval)), to compare their properties (e.g., number and size).
Fig 3. Computation of fiber parametric maps: (A) Starting from the skeleton graph (FN B-A+ tumor-like sample, 1024x1024 pixels), a pore isotropy map is derived (B), as the inverse value of the difference between the median pore angle and each individual one. (C) Starting from the graph-based fiber representation, a parametric map (fiber length, (D)) associates the fiber length, in pixels, to each corresponding connecting line.

In this context, the parametric maps are described by the union of two classes of pixels: those representing a realization of a GRF modelling the normal case, and those that are foreign to the
We expect these foreign elements to occur in regions with very high pixel intensity and/or in larger clusters taken at a specific threshold.

Modelling the parametric maps with GRF is only possible upon gaussianization (i.e. conversion of the empirical distributions into normal distributions) of the GRF marginal distributions. In practice, we only performed the gaussianization of the first-order marginal distribution of the GRF, i.e. the image intensity histogram, considering that the parameter maps under study are smooth enough. Therefore, the image intensity histogram was the only distribution to be gaussianized using an approach based on optimal transport (S1 Text). Thereby, the resulting histogram of intensity follows a normal distribution of zero mean with identical variance as the empirical native histogram.

To estimate the likelihood of a certain cluster to belong to a GRF termed $G_r$, depending on the maximal intensity of this cluster, we used the following formulations taken from the theory of random fields (12). Considering $x_0 = t + H_0$ as the intensity peak of a cluster (at intensity threshold $t$), one can estimate the probability that a cluster having an intensity peak equal to $x_0$, denoted $C_t^{H_0}$ belongs to a realization of GRF. This probability can be seen as the likelihood of a cluster (taken at threshold $t$) of having an intensity peak higher or equal to $t + H_0$:

$$P_H = P(C_t^{H_0} \in G_r) = \frac{x_0}{t} \sigma^{-3} \exp \frac{t^2 - x_0^2}{2\sigma^2}$$

Furthermore, the approximation for the probability of a given cluster having a spatial extent $S$ greater than $S_0$ is given by the following formulation:

$$P_S = P(C_t^{S_0} \in G_r) = \exp -\frac{2\pi^{-3/2}|\Lambda|^{1/2} S_0 \sigma^{-3} \exp -\frac{t^2}{2\sigma^2}}{\phi(-t)}$$

where

- $\Lambda$ is the covariance matrix of the partial derivatives of $G_r$
- $\sigma$ is the standard deviation of $G_r$
\( \Phi(-t) \) is the complementary cumulative distribution function.

Fig 4. Methodology for statistical detection of foreign regions to a GRF, in normal and tumoral parametric maps. (A) Normal and Tumoral fiber length map (FN B-A+). The normal sample is modelled as a realization of a GRF, and we assume that the tumor-like sample is a realization of the same process. Clusters of regions with an intensity higher than a given threshold, \( t = 50 \) (B), \( t = 80 \) (C), and \( t = 100 \) (D).
\( C = 80 \), \( D = 100 \) are declared foreign to a GRF with respect to a \( p\text{val} \), depending on the cluster maximum intensity value or their surface.

The comparison of these probability values against \( p\text{val} \) (e.g. 0.05) was necessary to guide the decision regarding foreign cluster detection (i.e. a cluster is counted as detected if each probability is smaller or equal to \( p\text{val} \)).

This methodology (see Materials and Methods) was applied for a quantitative and qualitative analysis of fiber length and pore orientation isotropy differences, across all four FN variant networks in normal and tumor-like states. We were interested in determining whether the SPM analysis of the selected spatial fiber features could reveal significant variant-specific differences between the FN variant networks in normal (N) and tumor-like (T) states, among the different FN variant networks in a normal state, and finally, among the variant networks in a tumor-like state.

To set up this framework, we divided the available sets of confocal images (1024 x 1024 pixels, 0.27 \( \mu \text{m/pixel} \); 70 images/variant for normal FN (N) and 65 images/variant for tumor-like FN networks) as follows. (1) For comparison of (N) vs (T) FN networks, we considered 50 (N) samples as the learning dataset, 20 (N) as test set for normal, and 65 (T) as test set for tumor-like networks. (2) Upon comparing (N) FN variant networks amongst each other, e.g. \( N_i \) vs \( N_j \), where \( i,j \in \{B-A, B-A+, B+A-, B+A+\} \): 50 samples out of \( N_i \) were taken as the learning dataset, 20 out of \( N_i \) as test set, while 70 out of \( N_j \) comprised the test set. (3) Finally, we compared (T) FN variant meshworks amongst themselves, e.g. \( T_i \) vs \( T_j \), 50 samples out of \( T_i \) were taken as the learning dataset, 15 out of \( T_i \) as test set, while 65 out of \( T_j \) comprised the test set. In all scenarios, a cluster is detected as foreign to GRF if either \( P_S \) or \( P_H \) are less or equal than \( p\text{val} \), \( p\text{val} \leq 0.05 \). Anomalies in fiber length (Fig 5) and pore isotropy maps (Fig 6) were detected at a few intensity thresholds (e.g. 70, 80, 90) and (10, 12, 14), respectively (Fig 5, Fig 6). Thus, using our approach, differences in fiber length and pore orientation isotropy could be localized in regions formed at different intensity thresholds. This property is very useful to obtain a qualitative analysis of parametric maps, where clusters that are statistically different from a normal model can be localized.
A. Fiber length map (N)  

Fiber length map (T)

B. Intensity-based cluster detection

C. Surface-based cluster detection
**Fig 5.** Qualitative analysis - Foreign cluster detection (with respect to the normal statistical model), applied to two samples of fiber length map (FN B-A+ tumor-like), 1024x1024 pixels (A). (B) and (C) depict the detected clusters (pval ≤ 0.05) at various intensity thresholds (70, 80, 90).
Fig 6. Qualitative analysis - Foreign cluster detection (with respect to the normal statistical model), applied to two samples of pore orientation isotropy map (FN B-A+ tumor-like),
1024x1024 pixels (A). (B) and (C) depict the detected clusters ($pval \leq 0.05$) at various intensity thresholds (10,12,14).

For the quantitative analysis, we were interested in determining whether we could detect significant differences between FN variant networks through the average number of identified foreign clusters, as well as the average cluster area per image. It is noteworthy that the group for which clusters foreign to the GRF were found at superior thresholds, had higher parametric values than in the normal model. For example, if significantly different regions occur in the tumor-like matrices compared to the normal ones, with respect to fiber length, then fibers are statistically more elongated within tumor-like networks than normal counterparts. As shown in Fig 7, we found both fiber length and pore isotropy to be significantly different for pairwise comparisons of normal and tumor-like FN variant networks. Essentially, FN architecture in the tumor-like ECM, relative to normal ECM, is represented by statistically longer fibers (Fig 7A, bottom) with a more pronounced pore isotropy (Fig 7B, bottom). Detailed results, including the average area and number of identified foreign clusters to a GRF, are presented in S1 Table1-Table11 (S1 Text).
**Fig 7. Quantitative analysis for detection of differences in fiber length and pore isotropy - Foreign cluster detection (with respect to the normal statistical model), applied for comparison of normal (N) and tumor-like (T) FN (1024x1024 pixels), where each matrix cell corresponds to a pairwise comparison between any two FN-specific variants. Detection is considered significant when at pval ≤ 0.05, the average number of foreign clusters per test database for one of the variants is**

### A. Fiber length

<table>
<thead>
<tr>
<th></th>
<th>B-A-(N)</th>
<th>B+A-(N)</th>
<th>B-A+(N)</th>
<th>B+A+(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A-(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A-(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-A+(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A+(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B. Pore isotropy

<table>
<thead>
<tr>
<th></th>
<th>B-A-(N)</th>
<th>B+A-(N)</th>
<th>B-A+(N)</th>
<th>B+A+(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A-(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A-(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-A+(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A+(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Anomaly detection intensity/surface based at pval ≤ 0.05**

<table>
<thead>
<tr>
<th></th>
<th>B-A-(T)</th>
<th>B+A-(T)</th>
<th>B-A+(T)</th>
<th>B+A+(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A-(T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A-(T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-A+(T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+A+(T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Significant detection**
- **No significant detection**
- **Not assessed**
The presence of FN Extra Domains (EDA and/or EDB) impacts FN assembly and fibrillar network organization by normal fibroblasts (9). This was demonstrated by local analysis of graph-based features (i.e. local node degree ratio, fiber thickness weighed by density, etc). An additional classification architecture was able to discriminate the four variant networks, on smaller images (512x512 pixels) with a relatively high accuracy, 83%. In accordance with these results, our current analysis of discriminating patterns across variant-specific FN fibrils showed that the inclusion of EDA and/or EDB in the cFN molecule, compared to pFN lacking these domains (B-A), affects the length of fibers assembled by normal (non-stimulated) fibroblasts (Fig 7A, top). Following activation, no significant differences in fiber length were observed across FN variant networks (Fig 7A, middle). This finding is not surprising, as activated fibroblasts in the tumor-like state are characterized by their enhanced contractility and elongated phenotype.

Interestingly, pore isotropy of the FN networks was similar for all four variants, in both normal and tumor-like states (Fig 7B, top and middle), suggesting that this parameter is not discriminating for the FN variants or activation state of the cells that assembled them. Intriguingly, our analyses revealed that pores of tumor-like matrices displayed a statistically significant orientation isotropy than matrices in the normal state (Fig 7B, bottom). Mechanisms underpinning this increase in random orientation of pores in the tumor-like FN networks remain to be elucidated.

**Conclusions**

The proposed methodology was designed for the quantification of the differences in the spatial organization of ECM networks, in normal and tumor-like architectures. In this case, we chose to analyze the matrix patterns of four alternatively spliced FN variants, which we have previously been able to discriminate with different learning approaches and relying on graph-based feature analysis (9). The current computational pipeline is now available to be tested with a MATLAB GUI.
Existing ECM analysis methods exploit alternative fiber detection techniques, such as Fast Fourier Transform bandpass filters (13), ridge detection (14), or fast discrete curvelet transform, which is arguably the best suited method to detect curvilinear anisotropic objects, among the previous options. The latter option was used in conjunction with a fiber extraction algorithm (6). Our method relies on a more flexible detection scheme using Gabor filters, thereby avoiding translation/rotation errors, and unlike other methods, associates graph networks to fiber morphological skeletons, enabling diverse fiber analyses. One such analysis, as seen in this work, provides a statistical assessment of the regions which are significantly different in terms of various geometric features.

The statistical approach can be evaluated on parametric maps at different thresholds, both quantitatively and qualitatively, whilst producing results that are more reliable and statistically relevant (by providing pval) than a simple hard thresholding of the parametric maps. The current analysis compares the fiber length and pore orientation isotropy parametric maps, of FN networks in normal and tumor-like states. Parametric maps can be extended to include other fiber or pore-specific parameters (e.g. fiber density, width, length, orientation, waviness, and straightness) useful for differentiating among various biological networks.

Computational analyses of ECM structures can provide essential information about their role in shaping the cellular microenvironment topology in health and during disease development. Prognostic ECM-specific signatures have already been inferred in cancer-related studies, diabetes, wound healing, and nerve repair. There is also a growing interest in the integration of cell and ECM analyses in a spatially resolved manner to further understand the interactions between cells and their matrix microenvironment (15). Our study proposes a versatile pipeline for the analysis of ECM organization in human tissue, in which a future integration of staining for cellular components can be envisioned. For instance, a combined local analysis of parametric maps and metrics describing the organization/morphology of adjacent cells could potentially help elucidate the complex interplay between cellular and non-cellular components.
Materials and methods

Materials and FN preparations

Recombinant human TGF-β1 was from R&D Systems Inc. (Minneapolis, MN, USA). All other chemicals and reagents were purchased from Sigma Aldrich (St Louis, MO, USA) unless otherwise stated. Purified recombinant FN variants were produced as previously described (9).

Cells and culture conditions

Fn1-/- mouse kidney fibroblasts (MFs) were generated and cultured as previously described (9). For experiments, FN was depleted from FCS using gelatin sepharose-4B columns (GE Healthcare, Uppsala, Sweden), and the culture medium was supplemented with Penicillin-Streptomycin 100 U/ml and TGF-β1 (5 ng/ml) where indicated. Absence of Mycoplasma sp. contamination was routinely verified by PCR as described elsewhere (16).

Generation of MF-derived matrices, immunofluorescence staining and microscopy

MF-derived matrices were generated as described previously. For FN immunostaining, primary antibody (rabbit polyclonal anti-FN) was from MerckMillipore (Darmstadt, Germany). Fluorescently labeled (Alexa Fluor 488-conjugated) secondary antibody was from Thermo Fisher Scientific. After staining, the coverslips were mounted in ProLong® Gold antifade reagent (Thermo Fischer Scientific). Confocal imaging was performed on a Zeiss LSM710 confocal system equipped with a 10X/0.45 NA objective. For visual representation, image treatment was performed using Fiji (17).

Statistical analysis based on GRF

Statistical parametric maps (SPMs) are used to evaluate the probability of change in every pixel (18) by using decision tests based on the magnitude of the SPM values (i.e. the peak intensity of a cluster in SPM) and on the spatial extent of these clusters formed at a certain threshold. In our experiments, we focused on two different parametric maps representing the length of fibers and...
the isotropy of pore angles. We embedded the SPM framework, initially developed to analyze single datasets independently, into a machine learning paradigm. To evaluate the differences between two given groups of maps (e.g., FN variants in normal state vs corresponding FN variants tumor-like state), we considered one of the groups as the normal realization of the GRF which we divide into a learning set and a smaller test set. The second group is tested for foreign regions, therefore all the images belonging to this group are considered part of the test set. The proposed method learns the normal GRF model specific parameters, i.e. the average value of \( \Lambda \), from the training set. These learnt parameters are subsequently used to compute the two relevant probabilities, \( P_H \) and \( P_S \).

In the learning phase, we learn the model parameter, as follows.

For all images \( I_j \) (previously gaussianized (S1 Fig)) in the learning set:

- Compute \( \Lambda_j \) (empirical estimator of the covariance of partial derivatives of \( I_j \)). If for an image function \( f \in R^2, \) we consider its gradient vector \( \nabla f = (f_x, f_y) = (\partial f / \partial x, \partial f / \partial y) \) then \( \Lambda_j = \text{cov}(f_x, f_y) = E[(f_x - E[f_x])(f_y - E[f_y])] \).
- Compute \( \sigma_j \), as the \( I_j \)'s sample standard deviation.

The last step involves storing the average \( \Lambda_m, \sigma_m \) of the learning dataset.

The test phase is where the decision whether the clusters, taken at a certain threshold \( t \), are considered foreign to the GRF, is made, according to a given pval.

For all the images \( I_j \) in the test set:

- Gaussianize \( I_j \) so that it can be approximated by a realization of a GRF. The result is a new image \( I_g \), whose histogram is Gaussian with identical variance to that of \( I_j \).
- For a given list of thresholds \( T = (t_1, t_2, ..., t_n) \):
  - Binarize \( I_g \) according to the threshold \( t_i \)
  - Find the list of connected components in the binary image resulted from thresholding and for every (labeled) connected component \( \{l_1, l_2, ..., l_p\} \):
Compute $P_H$ (Eq.1), using the learnt model parameter $\sigma_m$. If its value is lower than a chosen pval, the cluster is considered a foreign element to a GRF.

Compute $P_S$ (Eq.2), using the learnt model parameters $\Lambda_m, \sigma_m$. If its value is lower than a chosen pval, the cluster is considered a foreign element to a GRF.

**Test for anomaly detection in a GRF using simulation**

To better illustrate the principle behind the proposed approach to detect statistical abnormalities in a GRF realization, we considered a simple scenario which includes a synthetic example (normal), representing a simulation of a GRF of zero mean. To this example we added 6 foreign objects (abnormal), representing ellipses with different surfaces, and intensity levels. The aim of this example is to showcase the suitability of our tool when detecting anomalies at various thresholds, based on the maximum intensity of the different regions detected at each threshold, or their spatial extent. Specifically, in this test, we were interested in verifying whether our approach can detect the foreign ellipses, at various thresholds, for pval $\leq 0.05$ based on either an intensity or a surface criterion.

Our method learns the GRF model parameters from the normal example and then uses these to compute the two probabilities of belonging to GRF for each region at various thresholds in the abnormal example. As shown in Fig 8, the current method, compared to a naïve hard thresholding (threshold equal to 5), is better suited to detect only the abnormal elements, i.e. ellipses, thresholds equal to (5,10,15,20). A possible decision could jointly consider the two criteria, i.e. intensity and surface-based detection, and thus lead to the selection of the 6 ellipses together with a few false positive alarms.
**Fig 8.** (A) Realization of a GRF of zero-mean and addition of 6 different sized ellipses. (B) Hard thresholding ($t = 5$ pixels). (C) Intensity-based detection at $pval \leq 0.05$, at thresholds equal to $(5,10,15,20$, see colorbar). (D) Surface-based detection at thresholds equal to $(5,10,15,25$, see...
colorbar). *(E) Example of joint decision of cluster selection based on intensity and surface criteria*

Acknowledgements

We thank the members of the Adhesion Signaling and Stromal Reprogramming in the Tumor Microenvironment team for critical discussion on the manuscript, iBV PRISM imaging platform for the use of their machines and support, in particular, Sébastien Schaub.

References


11. Lafarge F, Descombes X, Zerubia J, Mathieu S. Détectioon de feux de forêt par analyse statistique d’événements rares à partir d’images infrarouges thermiques Forest fire detection by statistical analysis of rare events from thermal infrared images.


Competing interests

The authors declare no competing interests.
Supporting Information (S1)

S1 text: Supporting Methods

Graph-based representation on top of Gabor filters

We defined a set of Gabor kernels $g_k$, characterized by an exponential term which provides the shape of a bivariate gaussian kernel multiplied by a cosine function which describes its oscillations in space:

$$g_k = \exp\left[-\frac{1}{2}v^t(\Sigma^{-1}_\theta)v\right]\cos\left(\frac{2\pi x\theta_i}{\lambda_j} + \phi\right) \quad (1)$$

where we denote by $v = (x,y)^t$, the 2D coordinate vector, indicating pixels localisation in a bidimendional Cartesian coordinate system.

To capture the various geometrical properties of the FN fibers using Gabor filters, we have constructed a set of Gabor kernels $g_k$, $k \in \{1,2,...,60\}$, defined by the following parameters:

- Fiber orientation is represented by $\theta_i$, $\theta_i = i\pi/20$ where $i = \{0,1,2,...,19\}$.

- Fiber thickness is denoted as $\lambda_j$, $j = \{1,2,3\}$ and corresponds to the wavelength (in pixels) of the cosine term, the values of which are equal to $\lambda_j/2$ and vary between 3 and 5 pixels. Therefore, the thinnest fibers are detected when $\lambda_j = 6$ pixels, medium thickness fibers correspond to $\lambda_j = 8$ pixels, while the thickest are characterized by $\lambda_j = 10$ pixels.

- For accurate localization of fibers, the phase of the cosine function, $\phi$, is set to $\{0\}$.

- The spatial support of the kernels is given by $\sigma_x = 5$ pixels and $\sigma_y = 3$ pixels, indicating an anisotropic filter that is appropriate for fiber detection.

At any given location within the filtered image sample, the Gabor kernel that returns the highest coefficient modulus was retained. Once fibers were enhanced using Gabor filters, we extracted their morphological skeleton, using different morphological operations (e.g.
binarization with hysteresis, morphological thinning, etc.). The graph-based representation of
the fiber skeletons was subsequently derived, using a toolbox that produces a network graph
assigned to a morphological skeleton (1). To reconstruct the missing fibers, we used Dijkstra’s
weighted shortest path algorithm, which finds for each node (i.e. fiber end), the shortest path
to other extremities or nodes of the skeleton graph, which are within a given distance to the
considered node. Among all possible paths, we chose the minimal one, relying on the intensity
from the Gabor filtered image. Additionally, the reconnection of the fibers was made using
the angle of the Gabor maximum response as a guideline, which resulted in the reconnection
of the fibers within a predefined cone sector around this local fiber orientation.

Parametric maps derivation
To build a fiber length map, we relied on the simplified graph-based representation, where
we kept the location of the nodes (fiber ends and crosslinks) and replaced the length of the
fiber with a connecting line. During the next step, we identified the 2D pixels coordinates that
approximate the straight line between the nodes, (using Bresenham’s line algorithm) and
assigned the length value of the connecting line to these pixels. The last step for generating
the fiber length map, included the extrapolation of the fiber length values (2) and smoothing
of this result with a Gaussian kernel.

Concerning the pore isotropy parametric maps, the starting point was the skeleton graph.
Pore orientation was computed by first fitting ellipses to each pore and was subsequently
obtained by measuring the angle $\theta_i$ between the horizontal axis and the major ellipse axis.
Pore median value is then subtracted from each $\theta_i$. Finally, the inverse absolute value of the
resulting individual score per pore (indicative of the orientation isotropy) will be assigned to
all pixels filling its corresponding surface, and subsequently smoothed out with a Gaussian
kernel.

1D Optimal Transport
The approach that was considered for map gaussianization was based on the optimal
transport framework (3). More specifically, the problem of "converting" the empirical
intensity distribution of the fiber length map into a normal distribution with the same
variance, can be approached by performing the 1D optimal transport between the two distributions. Briefly, for a map I with an empirical mean and variance of intensity, we consider a second image J, whose intensity pixels follow a normal distribution (S1 Fig). The 1D optimal transport problem will determine how to optimally permute the pixels indices in J, to "recreate" the image I. Since the intensity of the pixels in the permuted version of J does not change, the result will resemble I, but will have the histogram of J (4).

67 Gaussian Random Fields

Given a complete probability space \((\Omega, F, \mathbb{P})\) and T a topological space, then a random field of real values is a measurable application \(X : \Omega \rightarrow R^T\). The finite-dimensional \((F_d)\) distributions of X are defined as the set of functions \(F_{t_1, ..., t_n}\), where \(F_{t_1, ..., t_n}(B) = P\left( (X(t_1), ..., X(t_n)) \in B \right)\), \(\forall n > 0, \forall B \in \mathcal{B}^n, F(B)\) being the Borel sigma algebra on R.

A particular class of random fields is represented by the Gaussian random fields, whose marginal \((F_d)\) distributions are Gaussian vectors \(X = (X(1), ..., X(n))\), characterized by the probability density function (5):

\[
f_X(x) = (2\pi)^{-n/2} |V|^{-1/2} \exp \left[ -1/2 (x - \mu) V^{-1} (x - \mu)^T \right] (2)
\]
**S1 Fig. Gaussianization through optimal transport:** First row illustrates an example of a fiber length parametric map (Map I), a generated image J with normally distributed intensity around 0, and equal variance to that of Map I, and, finally, the converted map I with a corresponding normally distributed intensity. Second row shows the corresponding intensity histograms.

Where \( \mu = \left( E(X(i)) \right)_{i \in [1,n]} \) is the expectation and \( V = \left( E[(X(i) - \mu,)(X(j) - \mu)] \right)_{i,j \in [1,n]^2} \) is the covariance matrix.

The expectation (or mean value) of the number of clusters that appear at a threshold \( t \), of an image modelled by a zero-mean, homogeneous Gaussian field of dimension 2, is the following (5):

\[
E[m_t] = S(2\pi)^{-3/2} |\Lambda|^{1/2} t \sigma^{-3} \exp - \frac{t^2}{2\sigma^2} (3)
\]

where:

- \( m_t \) represents the number of clusters at a certain threshold \( t \)
- \( S \) is the number of pixels of the image
Λ is the covariance matrix of partial derivatives of the GRF

\( \sigma \) is the standard deviation of the GRF

Similarly, the mean value of the number of clusters at a threshold \( t + H_0 \) can be written as such:

\[
E[m_{t+H_0}] = S(2\pi)^{-3/2} |\Lambda|^{1/2} (t + H_0) \sigma^{-3} \exp - (t + H_0)^2 / 2\sigma^2 \tag{4}
\]

Considering \( x_0 = t + H_0 \) as the intensity peak of a cluster (at threshold \( t \)), one can estimate the probability that a cluster (at a threshold \( t \), having an intensity peak equal to \( x_0 \), denoted \( C_t^{x_0} \)) belongs to a realization of this GRF. This probability can be seen as the likelihood of a cluster (taken at threshold \( t \)) of having an intensity peak higher or equal to \( t + H_0 \):

\[
P(C_t^{x_0} \in G_t) = \frac{E[m_{t+H_0}]}{E[m_t]} = \frac{x_0}{t} \sigma^{-3} \exp \frac{t^2 - x_0^2}{2\sigma^2} \tag{5}
\]

Next, we are interested in the estimation of the probability that a cluster (at a threshold \( t \)) belongs to a realization of GRF, depending on its surface (spatial extent - number of pixels). To estimate the number of pixels \( (n_t) \) of a cluster at a threshold \( t \), we use the following equation (6):

\[
E[n_t] = \frac{E[N_t]}{E[m_t]} \tag{6}
\]

where \( N_t \) is the number of pixels of higher intensity than \( t \), and \( m_t \) is the number of clusters at threshold \( t \). Since the intensity values follow a normal (zero mean value) distribution, the expectation of \( N_t \) is the following:

\[
E[N_t] = S \int_t^\infty (2\pi\sigma^2)^{-1/2} \exp - \frac{x^2}{2\sigma^2} \, dx = S\phi(-t) \tag{7}
\]

where \( \phi(-t) \) is the complementary cumulative distribution function. It follows then (Eq. 3, 6, 7) that one can approximate the mean value of \( n_t \), accordingly:

\[
E[n_t] = \frac{\phi(-t)}{(2\pi)^{-3/2} |\Lambda|^{1/2} t \sigma^{-3} \exp - t^2 / 2\sigma^2} \tag{8}
\]
Furthermore, it has been shown in (7) that $n_t$ follows an exponential distribution law, which is commonly defined by a parameter $\lambda_t$, (the inverse of the mean expected value of the random variable).

Consequently

$$P(n_t = x) = \lambda_t \exp (-\lambda_t x) \tag{9}$$

where $\lambda_t = \frac{(2\pi)^{-3/2}|A|^{1/2}t\sigma^{-3}\exp -t^2/2\sigma^2}{\phi(-t)} \tag{10}$

It follows then that the approximation for the probability of a given cluster having a spatial extent $S$ greater than $S_0$ is given by the following formulation:

$$P(C_{\geq S_0} \in G_r) = P(n_t \geq S_0) = \exp (-\lambda_t S_0) = \exp -\frac{(2\pi)^{-3/2}|A|^{1/2}t\sigma^{-3}\exp -t^2/2\sigma^2}{\phi(-t)} \tag{11}$$

### Quantitative results (area and number) of detected clusters foreign to a GRF

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>Intensity</td>
<td>Count</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>79.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td>0.15</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>44.33</td>
<td>4.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**S1 Table 1.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A- and B-A+ FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at $pval \leq 0.05$, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table (B-A vs B-A).
**S1 Table 2.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A- and B+A- FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table (B-A vs B-A).

<table>
<thead>
<tr>
<th>Surface</th>
<th>Count</th>
<th>0.07</th>
<th>0.00</th>
<th>0.00</th>
<th>Surface</th>
<th>Count</th>
<th>0.59</th>
<th>0.34</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td></td>
<td>3.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td>85.20</td>
<td>19.92</td>
<td>9.00</td>
<td></td>
</tr>
</tbody>
</table>

**S1 Table 3.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A- and B+A+ FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table (B-A vs B-A).

<table>
<thead>
<tr>
<th>B-A- vs B-A-</th>
<th>Intensity</th>
<th>Thresholds</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>Intensity</th>
<th>Thresholds</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Count</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Area</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td></td>
<td>0.1</td>
<td>0.00</td>
<td>0.00</td>
<td>Surface</td>
<td>Count</td>
<td>0.77</td>
<td>0.39</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td></td>
<td>17.5</td>
<td>0.00</td>
<td>0.00</td>
<td>Area</td>
<td></td>
<td>63.65</td>
<td>18.44</td>
<td>6.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B-A- vs B-A-</th>
<th>Intensity</th>
<th>Thresholds</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>Intensity</th>
<th>Thresholds</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td>0.1</td>
<td>0.05</td>
<td>0.05</td>
<td>Count</td>
<td></td>
<td>0.11</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td></td>
<td>432.5</td>
<td>68.00</td>
<td>16.00</td>
<td>Area</td>
<td></td>
<td>2066.5</td>
<td>1069.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Surface</td>
<td>Count</td>
<td>0.43</td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Area</td>
<td></td>
<td>772.11</td>
<td>273.00</td>
<td>95.30</td>
</tr>
</tbody>
</table>
**S1 Table 4.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A- (N) and B-A- (T) FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th>B+A- vs B+A-</th>
<th>Thresholds</th>
<th>B+A- vs B+A-</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70  80  90</td>
<td></td>
<td>70  80  90</td>
</tr>
<tr>
<td>Intensity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>0.00  0.00 0.00</td>
<td>Intensity</td>
<td>0.88  0.51 0.26</td>
</tr>
<tr>
<td>Area</td>
<td>0.00  0.00 0.00</td>
<td>Area</td>
<td>3608.82 2194.45 1165.65</td>
</tr>
<tr>
<td>Surface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>0.15  0.10 0.10</td>
<td>Surface</td>
<td>0.92  0.89 0.78</td>
</tr>
<tr>
<td>Area</td>
<td>220.67 44.50 7.00</td>
<td>Area</td>
<td>3127.58 1315.53 506.22</td>
</tr>
</tbody>
</table>

**S1 Table 5.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B+A- (N) and B+A- (T) FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th>B-A+ vs B-A+</th>
<th>Thresholds</th>
<th>B-A+ vs B-A+</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70  80  90</td>
<td></td>
<td>70  80  90</td>
</tr>
<tr>
<td>Intensity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>0.00  0.00 0.00</td>
<td>Intensity</td>
<td>0.72  0.40 0.23</td>
</tr>
<tr>
<td>Area</td>
<td>0.00  0.00 0.00</td>
<td>Area</td>
<td>6038.85 6265.31 6553.13</td>
</tr>
<tr>
<td>Surface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>0.15  0.10 0.10</td>
<td>Surface</td>
<td>0.78  0.54 0.42</td>
</tr>
<tr>
<td>Area</td>
<td>38.00 5.00 1.50</td>
<td>Area</td>
<td>5337.84 4563.03 3591.33</td>
</tr>
</tbody>
</table>

**S1 Table 6.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A+ (N) and B-A+ (T) FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.
### S1 Table 7. Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B+A+ (N) and B+A+ (T) FN (1024x1024 pixels) fiber length maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 70,80,90. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th></th>
<th>B+A+ vs B+A+</th>
<th>Thresholds</th>
<th>Intensity</th>
<th>Count</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### S1 Table 8. Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B-A- (N) and B-A- (T) FN (1024x1024 pixels) pore isotropy maps. Clusters are detected according to intensity and surface criteria, at pval ≤ 0.05, for various thresholds 10,12,14. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th></th>
<th>B-A- vs B-A-</th>
<th>Thresholds</th>
<th>Intensity</th>
<th>Count</th>
<th>10</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>1041.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td>0.25</td>
<td>0.20</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>370.20</td>
<td>37.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B+ vs B+</th>
<th>Thresholds</th>
<th>Intensity</th>
<th>Count</th>
<th>10</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>947.50</td>
<td>292.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td>0.30</td>
<td>0.05</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>400.83</td>
<td>278.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**S1 Table 9.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B+A- (N) and B+A- (T) FN (1024x1024 pixels) pore isotropy maps. Clusters are detected according to intensity and surface criteria, at $p_{val} \leq 0.05$, for various thresholds 10,12,14. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th></th>
<th>B-A+ vs B-A+</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intensity</td>
<td>Count</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>561.67</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>478.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B-A+ vs B-A+</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intensity</td>
<td>Count</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>1169.49</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>651.68</td>
</tr>
</tbody>
</table>

**S1 Table 10.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B+A+ (N) and B+A+ (T) FN (1024x1024 pixels) pore isotropy maps. Clusters are detected according to intensity and surface criteria, at $p_{val} \leq 0.05$, for various thresholds 10,12,14. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.

<table>
<thead>
<tr>
<th></th>
<th>B+A+ vs B+A+</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intensity</td>
<td>Count</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>3933.33</td>
</tr>
<tr>
<td>Surface</td>
<td>Count</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>1565.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B+A+ vs B+A+</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intensity</td>
<td>Count</td>
<td>1034.43</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>633.42</td>
</tr>
</tbody>
</table>

**S1 Table 11.** Foreign cluster detection (with respect to the normal statistical model), applied for the comparison of B+A+ (N) and B+A+ (T) FN (1024x1024 pixels) pore isotropy maps. Clusters are detected according to intensity and surface criteria, at $p_{val} \leq 0.05$, for various thresholds 10,12,14. The average number (count) and area of detected clusters are recorded, where the results corresponding to the normal model are shown in the first half of the table.
References


