IL NUOVO CIMENTO **44 C** (2021) 139
DOI 10.1393/ncc/i2021-21139-9

Communications: SIF Congress 2020

Collagen micro-architecture investigation in tumor sections by means of second-harmonic generation signal multiphasor analysis coupled with non-supervised machine learning techniques

R. Scodellaro(1), M. Bouzin(1), F. Mingozzi(2), F. Granucci(2),
L. D'Alfonso(1), M. Collini(1), G. Chirico(1) and L. Sironi(1)(*)

(1) *Department of Physics, University of Milano-Bicocca - Piazza della Scienza 3, 20126, Milano, Italy*
(2) *Department of Biotechnology and Biosciences, University of Milano-Bicocca Piazza della Scienza 2, 20126, Milano, Italy*

received 23 December 2020

Summary. — Collagen organization changes with the tissue pathological conditions, like cancer, and can be monitored through Second-Harmonic Generation imaging, a label-free method sensitive to the fibrils microstructure. As a consequence, collagen can be exploited as an early-tumor diagnosis marker. Coupling a phasor-based method with a non-supervised machine learning algorithm, our protocol is able to map pixel by pixel crucial features of the collagen fibrils and enlighten different collagen organizations. Basing on these maps, our protocol can automatically discriminate, on fixed tumor sections, tumor area from the surrounding tissue with an accuracy of $\sim 90\%$, opening the possibility to effectively assist histopathologists in cancer diagnosis.

1. – Introduction

Collagen has been recently recognized as a relevant prognostic factor associated with cancer progression and metastasis [1] and as a potential therapeutic target. Thanks to its non-centrosymmetric structure, it can be visualized by means of Second-Harmonic Generation (SHG) microscopy, a label-free nonlinear coherent optical process which is dependent on the polarization of the incident laser (Polarization-dependent SHG, PSHG) [2]. Recently, we developed a 2D phasor analysis, μMAPPS (Microscopic Multiparametric Analysis by Phasor projection of Polarization-dependent SHG signal), which analyzes, at pixel level, the PSHG signal collected on tissue sections images [3], and provides the mean orientation angle of the collagen fibrils θ_F and their susceptibility anisotropy parameter γ. A clustering algorithm [4], working in the phasor spaces, automatically selects pixels sharing similar θ_F and γ values, allowing to segment regions with similar microstructural properties in the tissue. Here, we enrich the μMAPPS approach by defining a clustering-related "fibrils entropy" parameter S, able to provide a local quantitative description of the collagen microstructure disorder state. By applying this new

(*) Corresponding author. E-mail: laura.sironi@unimib.it

Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0)

protocol to mouse-tail tendon samples and to large tissue sections (few mm in size) of a mouse breast cancer model, we demonstrate the effectiveness of the fibrils entropy in automatically segment skin and tumor regions in histopathology sections.

2. – Materials and methods

Optical set-up and analysis method. – The 2-photon microscopy setup is built around a confocal scanning head (FV-300, Olympus) mounted on an upright optical microscope (BX51, Olympus) and coupled to a fs-pulsed Ti:Sa laser (690–1040 nm, 80 MHz repetition rate, Mai Tai HP, Spectra Physics) [5]. The backscattered SHG signal, primed by a 800 nm excitation wavelength, has been collected by a water immersion objective (NA=0.95, WD=2 mm, 20X, XLUMPlan FI, Olympus), filtered by a 400/20 nm bandpass filter (Chroma Inc., HQ400/20) and acquired by a photomultiplier tube (HC125-02, Hamamatsu). Sequential images of mouse tendon samples and entire tumor sections have been acquired by rotating the half-wavelength waveplate from 0° to 180° in steps of 5°, with an excitation laser power of $P_{exc} = 50\,\mathrm{mW}$, measured before the scanning-head. The images have been analyzed by exploiting the μMAPPS method [3], with the PSHG signal [6] defined as

$$I_{PSHG}(\theta_n^L) = k[\sin^2(2\theta) + (\sin^2\theta + \gamma\cos^2\theta)^2] \qquad \text{with } \theta = \theta_n^L - \theta_F, \tag{1}$$

where k is a scale factor for the SHG signal intensity, θ is the angle between the laser polarization and the mean collagen fibrils orientations in the image plane and $\gamma = [\chi^2_{zzz}]/[\chi^2_{zxx}]$ is the ratio of off-diagonal to diagonal elements of the susceptibility tensor χ^2. Pixels sharing similar θ_F and γ values in regions of interest (ROIs) of the tissues have been clustered as described in [4]. The clustering algorithm works pixel by pixel and is applied in the θ_F-γ space. It first recognizes putative centers of the clusters, exploiting a maximum density approach. Each element which is not a putative center is assigned to the cluster characterized by the lowest element-center distance by an iterative procedure. In each cluster, the distance between its center and each element must be lower than arbitrary chosen cut-off values θ_C and γ_C. In each ROI, the "fibril entropy" (further information in [7]) is defined as

$$S = \left[-\sum_{i=1}^{N} p_i \log(p_i)\right] / [-\log(1/E_c)] = \left[-\sum_{i=1}^{N} (x_i/E_c)\log(x_i/E_c)\right] / [-\log(1/E_c)], \tag{2}$$

where N is the total number of clusters and p_i is the ratio between the x_i elements in the i-th cluster and the total number of clustered elements E_c. Only clusters characterized by a number of elements higher than a set threshold (ET), computed as a percentage of the total number of analyzed pixels, have been considered.

Mouse-tail tendons. – Mouse-tail tendons have been harvested and treated as described in ref [3]. 3 FOVs ($225 \times 225\,\mu\mathrm{m}^2$) related to 2 tendons of different mice have been analyzed.

4T1 Breast cancer samples. – BALB/c females mice of 7–12 weeks of age were inoculated in the deep derma in the left flank with the minimal tumorigenic dose of 4T1 cells (5×10^4) at Day 0, as extensively reported in ref. [7]). We analyzed 3 entire tumor sections from 3 different mice (100–300 FOVs of $377 \times 377\,\mu\mathrm{m}^2$ for each section). The

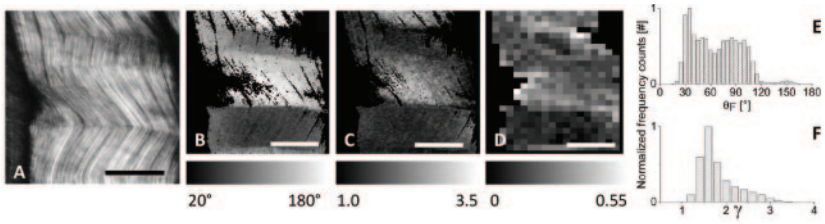

Fig. 1. – Analysis of a mouse tail tendon. (A) Maximum intensity projection image of the analyzed mouse tail tendon. Scale bar: 80 μm. (B), (C) θ_F and γ maps. (D) "fibrils entropy" map. (E), (F) θ_F and γ counts histograms.

mice were kept in a pathogen-free animal facility, in individual and ventilated cages with 12 h light/dark cycles with food and water *ad libitum*. The mice were monitored on a daily basis for signs of discomfort, including hunched posture, ruffled fur and lack of movement within the cage: the body condition score index (a qualitative assessment of an animal overall appearance based on its weight, muscle mass, and bone prominence) was used. Mice did not present signs of distress. Experiments were performed using protocols approved by the Institutional Animal Care and Use Committee of the University of Milano-Bicocca and by the Italian Ministry of Health.

3. – Results

At first, we exploited the μMAPPS method to analyze the mouse-tail tendon PSHG images (an example is reported in fig. 1(A)) and retrieve the θ_F and γ parameters pixel-wise (their maps are reported in figs. 1(B), (C), respectively). The fibrils angular distribution (fig. 1(E)) shows a main peak at $\theta_F = 30°$ and two secondary components at 90° and 150°. The γ parameter spans a wide range, between 1 and 3.5 (fig. 1(F)). In order to quantitatively highlight regions characterized by different local disorder levels, we computed the "fibrils entropy" S (eq. (2)) on 9×9 μm^2 non-overlapping ROIs, starting from the output of the clustering algorithm ($\theta_C = 5°$ and $\gamma_C = 0.2$) applied to the two coupled θ_F and γ phasor spaces (its map is reported in fig. 1(D)). In the crimped regions (fig. 1) we retrieved $S = 0.31 \pm 0.23$, while in regions characterized by regular fibers $S = 0.13 \pm 0.09$ (data not shown). A more diagnostically relevant study is represented by the application of the same protocol to a tumor model surrounded by a skin layer (fig. 2(A)). In order to separate the tumor and the skin regions, we sequentially analyzed $150 \times 150\,\mu$m^2 non-overlapping ROIs encompassing the histology section and extracted the θ_F and γ maps and distributions (figs. 2(D), (E)) related to the skin and tumor ROIs in figs. 2(B), (C), respectively). We computed the S map (fig. 2(F)) by applying the clustering procedure ($\theta_C = 5°$ and $\gamma_C = 0.2$) to each ROI. Higher entropy values (S=[0.4–0.7]) characterize tumor areas, while for skin regions S=[0.1–0.45]. The results obtained for different ROIs dimensions, clustering conditions and selected thresholds are discussed in [7], also for a second tumor model. Finally, we exploited the S parameter to automatically segment tumor and skin regions, characterized by a different collagen structure (fig. 2(G)): we assigned a monochromatic white or gray look-up table (LUT) to pixels whose ROIs show fibrils entropy, respectively, above or below the mean threshold value S_{th} computed on the entire section ($S_{th} = 0.4$ for fig. 2(G)). The skin region is predominantly gray-coloured, with few exceptions close to the section edge, while the tumor

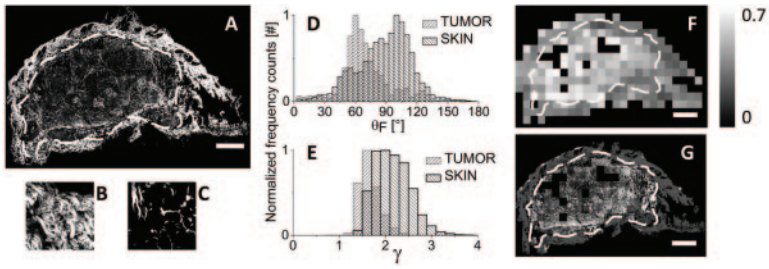

Fig. 2. – Analysis of a tumor section. (A) Maximum intensity projection of the section. Scale bar: 1 mm. The dashed line represents the skin-tumor boundary traced by an expert. (B), (C) $150 \times 150\,\mu\text{m}^2$ ROIs of skin (left) and tumor (right) areas. (D), (E) θ_F and γ histograms of (B), (C). (F) "fibrils entropy" map for $150 \times 150\,\mu\text{m}^2$ ROIs. (G) Result of the segmentation procedure, for $S_{th} = 0.4$ and $150 \times 150\,\mu\text{m}^2$ ROIs. ET=1%. Black color codes for not analyzed ROIs.

area is mainly white-coloured. The accuracy in retrieving the two regions, computed by comparing the number of the pixels classified by the algorithm with those extracted by an expert operator selection, is (88±4)% for the skin region and (91±6)% for the tumor area.

4. – Discussion and conclusion

In this work, we exploited the μMAPPS method [3], coupled to the computation of the fibrils entropy parameter in non-overlapping ROIs encompassing the entire tissue, to automatically recover different collagen microstructures at a mesoscopic level (10–200 μm) from PSHG images of a mouse tendon and in fixed tumor sections, as prepared for histopathology. The fibrils entropy is 0.2–0.3 units higher in the tumor ROIs than in the skin, and the entropy-based segmentation of the two regions agrees with the manually obtained results, with an average accuracy of 90±5%. These results will pave the way to expand the dictionary of features extracted from label-free imaging techniques describing pathological tissues, allowing to assist pathologists for fast, automated and reliable diagnosis at low costs, and overcoming the need of samples fixing and staining, needed for immunofluorescence and immune-histochemistry methods.

REFERENCES

[1] Martins Cavaco A. C. *et al.*, *Cancer Metastasis Rev.*, **39** (2020) 603.
[2] Campagnola P. J. *et al.*, *Nat. Biotechnol.*, **21** (2003) 1356.
[3] Radaelli F. *et al.*, *Sci. Rep.*, **8** (2018) 6314.
[4] Rodriguez A. *et al.*, *Science*, **344** (2014) 1492.
[5] Caccia M. *et al.*, *Eur. Biophys. J.*, **37** (2008) 979.
[6] Rouède D. *et al.*, *Sci. Rep.*, **7** (2017) 12197.
[7] Scodellaro R. *et al.*, *Front. Oncol.*, **9** (2019) 527.