

Analyse statistique de la répartition spatiale d'objets le long de filaments en microscopie à fluorescence

Thibault LAGACHE¹, Brice ROTUREAU², Philippe BASTIN², Jean-Christophe OLIVO-MARIN¹

¹Unité d'Analyse d'Images Biologiques, Institut Pasteur

²Unité de Biologie Cellulaire des Trypanosomes, Institut Pasteur
25 rue du Docteur Roux, 75724 Paris Cedex 15, France

thibault.lagache@pasteur.fr, brice.rotureau@pasteur.fr, philippe.bastin@pasteur.fr,
jcolivo@pasteur.fr

Résumé – L'analyse de la distribution spatiale des molécules le long des filaments cellulaires en microscopie de fluorescence fournit des informations essentielles sur les mécanismes de transport intracellulaire. L'outil standard pour l'analyse de l'organisation spatiale des molécules est la fonction K de Ripley, qui permet de tester statistiquement l'hypothèse de distribution aléatoire des molécules contre leur *clustering* ou leur dispersion. Cependant, deux problèmes techniques majeurs subsistent dans l'analyse automatique et robuste de la distribution spatiale le long de filaments: la détection automatique robuste des filaments sur toute leur longueur et le calcul des quantiles critiques de la fonction K de Ripley qui repose actuellement sur des simulations de Monte-Carlo. Nous proposons ici tout d'abord une approche en deux étapes pour la détection des filaments qui combine le seuillage de l'image dans le domaine des ondelettes orientables puis la reconstruction des parties manquantes des filaments par chemin minimal. Dans un second temps, nous présentons une expression analytique des quantiles de la fonction K de Ripley, conduisant à un test statistique rapide de la distribution aléatoire des molécules. Enfin, nous utilisons notre test statistique pour analyser la distribution spatiale des molécules impliquées dans le transport intraflagellaire le long du flagelle du parasite *Typanosoma brucei*.

Abstract – The analysis of the spatial distribution of molecules along cell filaments in fluorescence microscopy gives essential information on intracellular transport mechanisms. The standard tool for analyzing molecules' organization is the Ripley's K function, which permits to test statistically the hypothesis of molecules' random distribution against clustering or dispersion. However, there are two major technical issues that plague the automatic and robust analysis of molecule spatial distribution along filaments: the accurate detection of filaments in fluorescence microscopy, and the computation of the critical quantiles of Ripley's K function which is currently based on Monte-Carlo simulations. We propose here a two-steps approach that combines image thresholding in steerable wavelets domain and minimal path reconstruction to detect and quantify filaments in their entire length. Then, we present an analytical expression of the quantiles of Ripley's K function, leading to a fast and robust statistical test of molecule random distribution. Thereafter, we used our statistical test to analyze the spatial distribution of molecules involved in intraflagellar transport along the flagellum of the parasite *Typanosoma brucei*.

1 Introduction

L'analyse de la l'organisation spatiale des molécules fluorescentes en biologie moléculaire permet de mieux comprendre les mécanismes de divers processus cellulaires [1]. En particulier, l'étude de la répartition des molécules voyageant le long de filaments cellulaires (actine, microtubules...) permet d'élucider les mécanismes de transport intracellulaire.

En analyse spatiale, les molécules sont le plus souvent représentées comme des objets ponctuels *via* leur centre de masse dans une domaine d'étude délimité (cellule, filaments...). Des statistiques spatiales du second ordre comme la fonction K de Ripley [2] permettent ensuite d'analyser quantitativement les caractéristiques de ces processus ponctuels. En particulier, il est important de savoir si les

points sont distribués aléatoirement suivant un processus de Poisson homogène ou si, au contraire, on peut déceler la présence statistique de motifs particuliers comme du *clustering* ou de la dispersion entre les points (Figure 1). Dans le cas d'un filament, la détection d'un *cluster* permet de démontrer la présence d'un embouteillage ou d'un train de molécules se déplaçant le long du filament. La mise en évidence statistique de motifs spatiaux passe par la comparaison de la fonction K de Ripley avec ses quantiles sous l'hypothèse nulle de distribution aléatoires des points. Ces quantiles dépendent de la longueur du filament et du nombre de points mais ne sont pas connus analytiquement et sont donc calculés à l'aide de simulations Monte-Carlo coûteuses en charge de calcul.

Par ailleurs, la détection robuste des filaments en microscopie de fluorescence reste un enjeu majeur en analyse

d'images quantitative, notamment dans des images bruitées. De plus, des variations d'intensité de fluorescence le long des filaments sont fréquentes (marquage inhomogène, sorties du plan focal...), et des segments de filaments sont ignorés lors de leur détection automatique.

Nous proposons donc dans la section 2 de combiner le seuillage de l'image dans le domaine des ondelettes orientables, afin de détecter les filaments anisotropiques, avec un algorithme de chemin minimal [3] pour reconstruire les segments manquants de filaments. La reconstruction robuste des filaments sur toute leur longueur est une étape préalable indispensable à l'étude de la répartition spatiale des molécules le long de ces filaments. Dans la section 3, nous introduisons alors une nouvelle statistique basée sur la fonction de Ripley 1D et calculons analytiquement ses quantiles sous l'hypothèse nulle de distribution aléatoire en fonction de la longueur du filament et du nombre de molécules (points). Cela nous permet de construire un test statistique rapide et robuste afin de détecter le *clustering* éventuel des molécules le long des filaments cellulaires. Enfin, dans la section 4, nous illustrons notre méthode par l'analyse statistique de la distribution spatiale des molécules le long du flagelle (filament) du parasite *Typanosoma brucei* responsable de la maladie du sommeil.

2 Détection des filaments

La plupart des méthodes de détection telles que le seuillage d'ondelettes isotropes [4] ou les méthodes de déconvolution gaussienne ne parviennent pas à détecter avec précision des structures anisotropes comme les filaments. Pour répondre à ce problème, une méthode basée sur la convolution par des fonctions orientables optimisant un critère de type Canny a été développée par Jacob et Unser [5]. Parallèlement aux méthodes basées convolution, des transformations multi-échelles directionnelles ont été proposées au fil des ans, comme les *curvelets* [6]. Parmi ces méthodes, Unser et ses co-auteurs ont proposé récemment un cadre permettant de concevoir et de paramétrer des *tight frame* d'ondelettes orientables de $L_2(\mathbb{R}^2)$ [7, 8] qui unifie la plupart des transformations directionnelles et généralise la pyramide orientable de Simoncelli [9]. Dans ce cadre, la base d'ondelettes orientables est construite en appliquant la transformée de Riesz à une base primaire d'ondelettes isotropes $\{\psi_{i,\mathbf{k}}\}_{i \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2}$ indexée par une paire i, \mathbf{k} d'échelles de résolution et d'indices de localisation. Ce cadre primaire fournit la décomposition multirésolution du signal, les fonctions $\{\psi_{i,\mathbf{k}}\}_{i \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2}$ étant générées à partir d'une fonction mère ψ :

$$\psi_{i,\mathbf{k}}(\mathbf{x}) = \psi_i(\mathbf{x} - 2^i \mathbf{k}) \text{ with } \psi_i(\mathbf{x}) = 2^{-2i} \psi(\mathbf{x}/2^i). \quad (1)$$

Parmi les différentes fonctions mères proposées par Unser et ses co-auteurs (table 1 dans [7], et [8]), nous avons choisi ici la fonction mère de Simoncelli dont la transformée de

Fourier est donnée par [10] :

$$\hat{\psi}(\omega) = \cos\left(\frac{\pi}{2} \log_2\left(\frac{2|\omega|}{\pi}\right)\right) \mathbf{1}_{\frac{\pi}{4} < |\omega| \leq \pi} \quad (2)$$

La base d'ondelettes orientables est ensuite obtenue en appliquant la transformée de Riesz généralisée $\mathcal{R}_{U_{M,N}}$ de paramètre matriciel $U_{M,N}$ à la base d'ondelettes primaire (1) [7, 8]. $\mathcal{R}_{U_{M,N}}$ est une fonction vectorielle dont la m^{th} ($1 \leq m \leq M$) ondelette orientable $\psi_{i,m}^s(\mathbf{x})$ à l'échelle i est égale à [7, 8]

$$\psi_{i,m}^s(\mathbf{x}) = [\mathcal{R}_{U_{M,N}}(\psi_i(\mathbf{x}))]_m = \sum_{n=-N}^{+N} u_{m,n} \mathcal{R}^n(\psi_i(\mathbf{x})) \quad (3)$$

où $u_{m,n}$ est l'élément ($1 \leq m \leq M, 1 \leq n \leq N$) de la matrice $U_{M,N}$, et \mathcal{R} est la transformation de Riesz standard. Comme proposé dans [8], nous avons restreint la somme de l'équation (3) aux éléments pairs, et nous avons utilisé la pyramide de Simoncelli généralisée $U_{M,N}$ dont la composante $u_{m,2n}$ est donnée par [7, 8]

$$u_{m,2n} = \frac{1}{f} \cos\left((2m - N)\theta_n - \frac{N\pi}{2}\right) \quad (4)$$

où $\theta_n = \left(\frac{\pi n}{M} + \frac{\pi}{2}\right)$ et $f = m!(N - m)!/N! \sqrt{\sum_{m,n} u_{m,n}^2}$ est un facteur de renormalisation. Finalement, nous avons seuillé (seuillage brut) les coefficients d'ondelettes orientables et reconstruit l'image à partir des coefficients seuillés afin de détecter les filaments fluorescents. Nous avons comparé notre méthode avec la méthode basée convolution de Jacob et Unser[5], implémentée dans la plateforme *open-source* d'analyse d'images biologiques Icy [11] (<http://icy.bioimageanalysis.org/>, plugin *Features Detector*, plugin ID : ICY-R1Y2A2), et nous avons observé que notre méthode basée transformation était moins sensible au bruit et aux variations locales d'intensité fluorescente le long des filaments.

Cependant, la reconstruction de l'image à partir du seuillage d'ondelettes n'est pas suffisante pour détecter les filaments sur toute leur longueur à cause des variations locales d'intensité. Afin de combler les parties manquantes de filaments, nous avons donc utilisé ensuite un algorithme de chemin minimal [3] : Soient $(t_i)_{i=1}^{2s}$ les extrémités des s segments reconstruits, notre méthode semi-automatique consiste à déterminer un secteur de recherche maximal D , puis à reconstruire tous les chemins minimaux entre t_i et $t_j, i \neq j$, dès lors que $|t_i - t_j| < D$ [3], avant de sélectionner manuellement les chemins qui conviennent. Nous obtenons ainsi des filaments complets qui peuvent servir de domaines d'étude pour l'analyse statistique de l'organisation spatiale des molécules le long de ces filaments.

3 Construction d'une statistique de test basée sur la fonction K de Ripley

Afin d'analyser statistiquement la distribution spatiale d'objets ponctuels (molécules) le long de filaments afin d'y détecter la présence éventuelle de motifs particuliers comme des *clusters*, nous utilisons une statistique \tilde{K} basée sur la fonction K de Ripley. Pour un paramètre de distance r et n objets le long d'un filament Ω , la fonction K de Ripley est donnée par [2]

$$K(r, n) = \frac{|\Omega|}{n(n-1)} \sum_{\mathbf{x} \neq \mathbf{y}} \mathbf{1}_{\{|\mathbf{x}-\mathbf{y}| \leq r\}} f(\mathbf{x}, \mathbf{y}), \quad (5)$$

où \mathbf{x} et \mathbf{y} sont les positions de deux objets ponctuels le long du filament et $f(\mathbf{x}, \mathbf{y})$ permet de corriger un possible biais dans l'expression de $K(r)$ pour des valeurs importantes du paramètre r à cause de la longueur finie $|\Omega|$ du filament Ω . Nous avons adapté la correction de Ripley [2] pour une géométrie 1D [12] : $f(\mathbf{x}, \mathbf{y}) = 1 + \frac{1}{2} (\mathbf{1}_{\{d(\mathbf{x}, \mathbf{y}) > d(\mathbf{x}, \partial\Omega)\}} + \mathbf{1}_{\{d(\mathbf{x}, \mathbf{y}) > d(\mathbf{y}, \partial\Omega)\}})$, où $\partial\Omega$ sont les extrémités du filament. En utilisant cette correction de bord, nous avons calculé que sous l'hypothèse nulle de distribution aléatoire [12] :

$$\mathbb{E}\{K(r, n)\} = 2r. \quad (6)$$

Afin de construire une statistique de test centrée réduite, nous avons ensuite calculé la variance de K [12] :

$$\text{var}\{K(r, n)\} = \frac{4|\Omega|}{n(n-1)} \left(r - \frac{5r^2}{4|\Omega|} + \frac{(n-2)r^3}{6|\Omega|^2} \right), \quad (7)$$

et proposé la statistique de test suivante [12] :

$$\tilde{K}(r, n) = \frac{K(r, n) - 2r}{\sqrt{\text{var}\{K(r, n)\}}}. \quad (8)$$

Ainsi, $\tilde{K}(r, n) > 0$ indique la présence de *clusters* le long

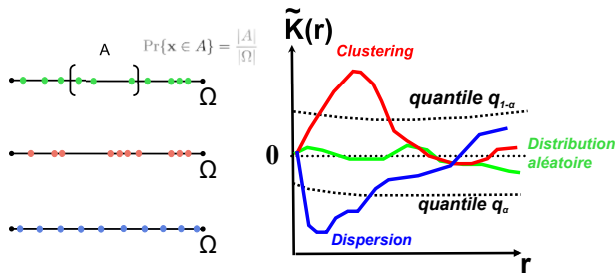


FIGURE 1 – Analyse statistique de la distribution spatiale d'objets le long de filaments L'hypothèse nulle de distribution aléatoire des objets le long du filament Ω est testée grâce à la statistique de test $\tilde{K}(r, n)$. Les déviations statistiquement significatives de $\tilde{K}(r, n)$ peuvent soit indiquer la présence de *clusters* avec un niveau de confiance α quand $\tilde{K}(r) > q_{1-\alpha}(r)$, soit de la dispersion quand $\tilde{K}(r) < q_{\alpha}(r)$, $q_{\alpha}(r)$ et $q_{1-\alpha}(r)$ étant les quantiles de $\tilde{K}(r)$ sous l'hypothèse nulle aux niveaux respectifs $0 \leq \alpha \leq 1$ et $1 - \alpha$. Quand $q_{\alpha}(r) < \tilde{K}(r) < q_{1-\alpha}(r)$, l'hypothèse nulle de distribution aléatoire ne peut pas être rejetée.

du filament alors que $\tilde{K}(r, n) < 0$ indique de la dispersion

entre points (Fig. 1). Afin de caractériser la signification statistique de ces déviations nous avons ensuite calculé analytiquement les quantiles de $\tilde{K}(r, n)$ sous l'hypothèse nulle de distribution aléatoire. Récemment, Lang et ses co-auteurs [13] ont démontré que pour $n \gg 1$, \tilde{K} converge en loi vers la loi normale centrée réduite $\mathcal{N}(0, 1)$ sous l'hypothèse nulle, ce qui revient à approximer les quantiles q_{α} au niveau α de \tilde{K} par les quantiles tabulés z_{α} de $\mathcal{N}(0, 1)$. Cependant, comme nous l'avons montré dans [14, 12] avec des simulations de Monte-Carlo, cette approximation n'est plus valable pour des valeurs intermédiaires de n rencontrées dans de nombreux problèmes en biologie. Nous avons donc utilisé l'approximation de Cornish-Fisher qui généralise le Théorème Central Limite [15] :

$$q_{\alpha} \approx z_{\alpha} + \frac{1}{6} (z_{\alpha}^2 - 1) \gamma \{\tilde{K}(r, n)\} \quad (9)$$

où $\gamma \{\tilde{K}(r, n)\} = \mathbb{E}\{\tilde{K}^3(r, n)\}$ est le coefficient de dissymétrie (*skewness*) de \tilde{K} . Dans [12], nous avons calculé analytiquement $\gamma \{\tilde{K}(r, n)\}$ en fonction de la longueur du filament et du nombre de molécules :

$$\gamma \{\tilde{K}(r, n)\} = \frac{|\Omega|}{n^2(n-1)^2 \text{var}\{K(r, n)\}^{\frac{3}{2}}} \left(8r|\Omega| - 33r^2 + \frac{28r^3}{|\Omega|} + 8(n-2) \left(3r^2 - 5.195 \frac{r^3}{|\Omega|} - \frac{3r^4}{|\Omega|^2} \right) + \frac{1.0292(n-2)(n-3)r^4}{|\Omega|^2} \right). \quad (10)$$

En particulier, nous soulignons que $\lim_{n \rightarrow \infty} \gamma \{\tilde{K}(r, n)\} = 0$, conduisant à l'approximation Normale démontrée par Lang et ses co-auteurs [13].

4 Application Biologique

Nous illustrons ici notre test statistique par l'analyse de la distribution spatiale des molécules se déplaçant le long du flagelle (filament) du parasite *Typanosoma brucei* responsable de la maladie du sommeil. Nous avons imagé une population de trypanosomes en microscopie de fluorescence multi-canaux, des marqueurs étant attachés à des molécules omniprésentes à l'intérieur du flagelle (immunomarquage avec un anticorps mAb25, canal vert), ainsi qu'à des molécules se déplaçant le long des microtubules du flagelle, visualisées dans le canal cyan (Figure 2 (1-2)). Afin d'extraire la position de ces dernières, nous avons utilisé une méthode de détection basée sur le seuillage d'ondelettes isotropes [4] (plugin *Spot Detector* dans Icy, Plugin ID : ICY-R3M2Y2. $n \approx 30$ par flagelle, Figure 2 (3)). Nous avons ensuite extrait le squelette du flagelle en couplant le seuillage de l'image dans le domaine des ondelettes orientables de type Simoncelli, avant de récupérer

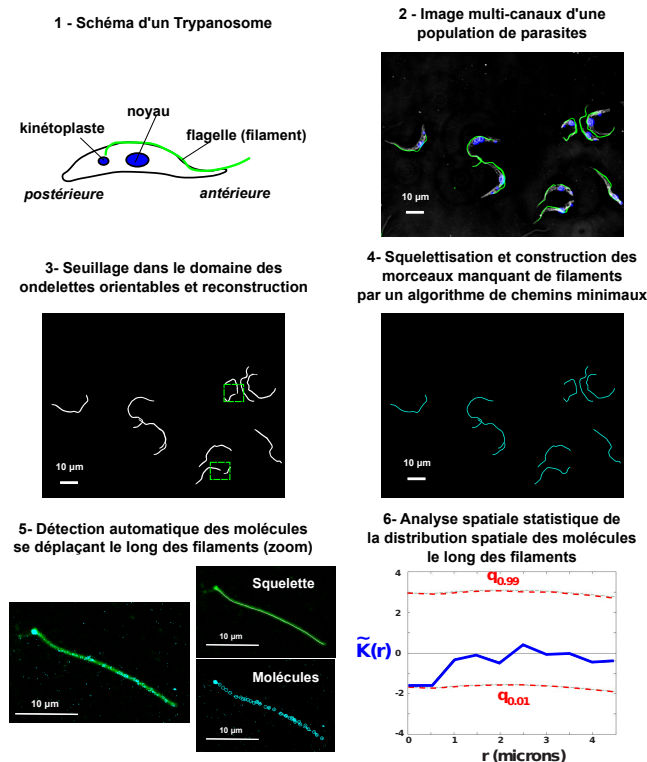


FIGURE 2 – Analyse statistique de la distribution spatiale des molécules le long du flagelle du parasite *Trypanosoma brucei*. 1- Un trypanosome est composé de deux conteneurs d'ADN : un noyau et un kinétoplaste auquel est accroché un unique flagelle. 2- Une image de fluorescence multi-canaux d'une population de parasites permet d'observer les différents flagelles et conteneurs d'ADN, ainsi que les molécules se déplaçant le long du flagelle. 3- Une première étape dans la détection robuste des flagelles consiste à seuiller l'image dans le domaine des ondelettes orientables. Cependant certains segments de flagelles ne sont pas détectés à cause des variations locales d'intensité. 4- Une seconde étape consiste donc à reconstruire ces segments manquants grâce à un algorithme de chemin minimal. 5- Après la segmentation robuste des flagelles sur toute leur longueur, les positions des molécules se déplaçant le long du flagelle sont obtenues automatiquement par seuillage d'ondelettes isotropes [4]. 6- L'analyse statistique de la distribution spatiale des molécules le long du flagelle se fait par comparaison de la statistique de test centrée, réduite $\tilde{K}(r)$ (Eq. 8) avec l'expression analytique de ses quantiles aux niveaux $\alpha = 1\%$ et $1 - \alpha = 99\%$ sous l'hypothèse nulle de distribution aléatoire (Eq. 9).

les segments manquants à cause des variations locales d'intensité grâce à un algorithme de chemin minimal (plugin *Filament Detector* en cours d'implémentation dans Icy) (Figure 2 (4)). Nous avons ainsi pu détecter les filaments de façon robuste sur toute leur longueur ($|\Omega| \approx 20\mu m$), ainsi que les positions précises des molécules à un instant donné (Figure 2 (5)). En calculant la statistique de test $\tilde{K}(r)$ et en la comparant à l'expression analytique de ses quantiles (Eq. 9), nous avons ainsi observé que pour tout $0 < r < 4.5\mu m$, $q_{0,01} < \tilde{K}(r) < q_{0,99}$, indiquant une distribution aléatoire des molécules le long du flagelle (Figure 2 (6)). Cette absence d'embouteillage pourrait être dû à la présence de 9 microtubules au sein du flagelle sur lesquelles peuvent se déplacer les molécules en parallèle, fluidifiant ainsi le trafic.

5 Conclusion

Afin d'analyser la distribution spatiale d'objets ponctuels (molécules fluorescentes) le long de filaments unidimensionnels (flagelle, cytosquelette cellulaire...), nous avons développé ici une méthode robuste de détection des filaments s'appuyant sur le seuillage de l'image dans le domaine des ondelettes orientables, suivi d'une reconstruction des parties manquantes des filaments par un algorithme de chemin minimal. Puis, grâce à la détection automatique des molécules le long du filament, nous avons pu calculer la statistique de Ripley $\tilde{K}(r)$ et la comparer à l'expression analytique de ses quantiles sous l'hypothèse nulle de distribution aléatoire. Nous avons ainsi construit une méthode globale permettant d'analyser l'organisation spatiale des molécules le long des filaments cellulaires et de détecter des motifs spatiaux particuliers comme des *clusters*, apportant des données quantitatives précises sur les mécanismes biophysiques gouvernant le trafic intracellulaire.

Références

- [1] J. Mounier et al., "Shigella effector ipab-induced cholesterol disrupts the golgi complex and recycling network to inhibit host cell secretion," *Cell Host Microbe*, vol. 12, no. 3, pp. 381–9, Sep 2012.
- [2] B.D. Ripley, *Statistical inference for spatial processes*, Cambridge University Press, 1988.
- [3] L.D. Cohen, *Minimal Paths and Fast Marching Methods for Image Analysis*, Springer, 2005.
- [4] J. C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognition*, vol. 35, no. 9, pp. 1989–1996, 2002.
- [5] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, 2004.
- [6] J.-L. Starck, E.J. Candès, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Trans Image Process*, vol. 11, no. 6, pp. 670–84, 2002.
- [7] M. Unser, N. Chenouard, and D. Van de Ville, "Steerable pyramids and tight wavelet frames in $l_2(r(d))$," *IEEE Trans Image Process*, vol. 20, no. 10, pp. 2705–21, Oct 2011.
- [8] M. Unser and N. Chenouard, "A unifying parametric framework for 2d steerable wavelet transforms," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 102–135, 2013.
- [9] E.P. Simoncelli and W. Freeman, "The steerable pyramid : A flexible architecture for multi-scale derivative computation," in *IEEE Second International Conference on Image Processing. Washington DC*, 1995.
- [10] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domains," *IEEE Trans Image Process*, vol. 12, pp. 1338–1351, 2003.
- [11] F. de Chaumont et al., "Icy : an open bioimage informatics platform for extended reproducible research," *Nat Methods*, vol. 9, no. 7, pp. 690–6, Jul 2012.
- [12] M. Lopez-Yunta, T. Lagache, J. Santi-Rocca, P. Bastin, and J.-C. Olivo-Marin, "A statistical analysis of spatial clustering along cell filaments using ripley's k function," in *International Symposium on Biomedical Imaging (ISBI'14)*, 2014.
- [13] G. Lang and E. Marcon, "Testing randomness of spatial point patterns with the ripley statistic," *ESAIM : Probability and Statistics*, vol. 17, pp. 767–788, 2013.
- [14] T. Lagache, G. Lang, N. Sauvonnet, and J.-C. Olivo-Marin, "Analysis of the spatial organization of molecules with robust statistics," *PLoS One*, vol. 8, no. 12, pp. e80914, 2013.
- [15] E.A. Cornish and R.A. Fisher, "Moments and cumulants in the specification of distributions," *Review of the International Statistical Institute*, vol. 5, pp. 307–320, 1937.