



**HAL**  
open science

## Analysis of the spatial organization of molecules with robust statistics

Thibault Lagache, Gabriel Lang, Nathalie Sauvonnet, Jean-Christophe Olivo-Marin

► **To cite this version:**

Thibault Lagache, Gabriel Lang, Nathalie Sauvonnet, Jean-Christophe Olivo-Marin. Analysis of the spatial organization of molecules with robust statistics. PLoS ONE, 2013, 8 (12), 7 p. 10.1371/journal.pone.0080914 . hal-01197614

**HAL Id: hal-01197614**

**<https://hal.science/hal-01197614v1>**

Submitted on 30 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Analysis of the Spatial Organization of Molecules with Robust Statistics

Thibault Lagache<sup>1,2\*</sup>, Gabriel Lang<sup>3</sup>, Nathalie Sauvonnnet<sup>4,2</sup>, Jean-Christophe Olivo-Marin<sup>1,2\*</sup>

**1** Unité d'Analyse d'Images Quantitative, Institut Pasteur, Paris, France, **2** Unité de Recherche Associée 2582, Centre National de la Recherche Scientifique, Paris, France, **3** Unité Mixte de Recherche 518 Mathématiques et Informatique Appliquées, AgroParisTech and INRA, Paris, France, **4** Unité de Biologie des Interactions Cellulaires, Institut Pasteur, Paris, France

## Abstract

One major question in molecular biology is whether the spatial distribution of observed molecules is random or organized in clusters. Indeed, this analysis gives information about molecules' interactions and physical interplay with their environment. The standard tool for analyzing molecules' distribution statistically is the Ripley's K function, which tests spatial randomness through the computation of its critical quantiles. However, quantiles' computation is very cumbersome, hindering its use. Here, we present an analytical expression of these quantiles, leading to a fast and robust statistical test, and we derive the characteristic clusters' size from the maxima of the Ripley's K function. Subsequently, we analyze the spatial organization of endocytic spots at the cell membrane and we report that clathrin spots are randomly distributed while clathrin-independent spots are organized in clusters with a radius of  $2 \mu\text{m}$ , which suggests distinct physical mechanisms and cellular functions for each pathway.

**Citation:** Lagache T, Lang G, Sauvonnnet N, Olivo-Marin J-C (2013) Analysis of the Spatial Organization of Molecules with Robust Statistics. PLoS ONE 8(12): e80914. doi:10.1371/journal.pone.0080914

**Editor:** Joshua Z. Rappoport, University of Birmingham, United Kingdom

**Received:** July 24, 2013; **Accepted:** October 15, 2013; **Published:** December 4, 2013

**Copyright:** © 2013 Lagache et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded in part by grants from the Agence Nationale de la Recherche (ANR-10-INBS-04-06 FranceBioImaging) and the Institut Pasteur (PTR 387). TL is funded by a Bourse Roux from Institut Pasteur. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: thibault.lagache@pasteur.fr (TL); jcolivo@pasteur.fr (JCOM)

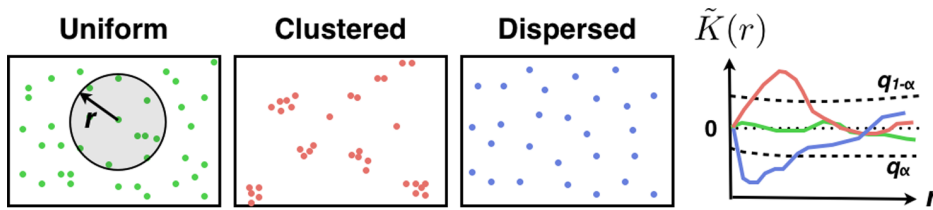
## Introduction

Spatial organization of objects is essential in many scientific areas because it provides information about objects' interactions and their interplay with their environment. Objects' organization can be studied at different scales, ranging from country size in epidemiology [1] to atomic structures in physics [2]. For example, the study of the distribution of leukaemia cases in Britain between 1966 and 1983 in epidemiology revealed some geographical aggregation that may be related to environmental factors [3]. In ecology, the analysis of spatial patterns across ten years in an aspen-white-pine forest [4] showed that tree distribution tended toward greater clumping than that expected from random mortality, which is due to the clonal nature of aspen. At molecular scale, the quantitative analysis of gold particle distribution in electron microscopy helped to analyze the three-dimensional distribution of pyramidal neurons and the related neural circuits [5]. It also gave hints about the distribution of Ras proteins at the plasma membrane [6,7] and the related organization of specialized micro-domains such as lipid rafts. Similarly, the analysis of the spatial distribution of fluorescent markers attached to proteins of interest in confocal microscopy shed light on underlying mechanisms of various cellular processes, such as signaling at immunological synapses [8], and can be used to measure cellular phenotype changes in different conditions, such as during pathogen infection [9].

In all spatial organization studies, objects (disease cases, trees, molecules ...) are represented as points in a delimited field of view (country, forest, cell ...) and quantitative methods are used to

extract features about spatial point distributions. Classical methods are either area-based or distance-based. In the first case, the points' pattern is characterized through its first-order properties such as the spatial variation of its points' density, which is often estimated with patches or kernel methods [10], whereas in the second case, distance-based methods rely on second-order properties of the points' pattern such as inter-point distances, and a major milestone was established by Clark and Evans (1954) who introduced statistics based on the distance of points to their nearest neighbors. An essential piece of information is given by the deviation of points' distribution from complete spatial randomness (CSR) and the concomitant detection of specific patterns such as point clusters (Figure 1). Thus, the two major goals when building a quantitative method are: **1**) assess statistically whether observed specific patterns such as clusters are not due to chance, that is to say points are not randomly distributed in the field of view, and **2**) determine the characteristics of the observed patterns such as the clusters' size. While the first goal is often achieved by the computation of the critical quantiles of the statistics used under CSR, the second one mainly involves fitting to parametric models.

However, these classical methods are plagued with some disadvantages: area-based methods cannot account globally for objects' interactions, and nearest-neighbors methods do not describe objects' interactions at several scales. To answer these problems, a great advance was made by Ripley in 1977 [11] who introduced the distance-based K function which describes the spatial organization of any point process quantitatively at several distance scales by taking into account all neighbors rather than



**Figure 1. Analyzing spatial point patterns with Ripley's K function.** The normalized and centered Ripley's K function  $\tilde{K}(r,n)$  is proportional to the number of pairs of points that are closer than  $r$  in  $\Omega$ . Deviations of  $\tilde{K}(r,n)$  from 0 (CSR) in clustering ( $\tilde{K}(r,n) > 0$ ) or dispersion ( $\tilde{K}(r,n) < 0$ ) conditions have to be compared with objective level of significance that are quantiles  $q_\alpha$  of  $\tilde{K}(r,n)$  at level  $0 < \alpha < 1$ . doi:10.1371/journal.pone.0080914.g001

only the nearest. Yet, Ripley's K function still presents some problems. First, there is no analytical formula that links the critical quantiles of the K function to the number of points and the geometry of the field of view. Consequently, the computation of the critical quantiles is based on intensive Monte-Carlo resampling, which induces an high computational load and requires an initial calibration for each field of view due to specific edge effects. Second, the quantification of pattern characteristics is also problematic because model parameters are currently extracted from fitting procedures involving a functional minimization, such as least square method; few efforts have been made to directly extract key spatial features such as the cluster radius or the minimal distance between dispersed points from the essential properties of the Ripley's curve such as its extrema [12].

Here, we propose two major methodological improvements: **1)** give a closed-form expression of critical quantiles, and **2)** relate standard features such as cluster size to essential properties of the Ripley's K function. Point 1 alleviates the need for Monte-Carlo simulations and point 2 bypasses minimization procedures. Taken together, these two points give rise to a fast, robust and analytical method which is additionally implemented and freely available in Icy [13] (<http://icy.bioimageanalysis.org>).

Thereafter, we used this method to characterize the spatial organization of different endocytic pathways. Endocytosis is indeed a key mechanism for cell homeostasis whereby cells engulf signaling molecules and nutrients from the extra-cellular medium. The most frequent endocytic pathway is mediated by the clathrin protein that forms coats around specific receptors, leading to membrane invagination and molecules entry [14–17]. Many other important pathways do not rely however on clathrin, notably the internalization of interleukin 2 (IL-2) and its receptor (IL-2R) [18] during the cell mediated immunity [19,20]. In both cases, the spatial organization of endocytic spots at the membrane still remains poorly characterized, while it might reflect localized cellular processes such as cell migration and signaling [21]. Here, we compare the spatial organization of clathrin-dependent and -independent endocytosis. We report that both pathways are regularly organized at small distance (for  $r < 1 \mu\text{m}$ ). At larger distance scales, clathrin-independent pathways exhibit clusters with a radius of about  $2 \mu\text{m}$  while clathrin-dependent putative endocytic sites are randomly distributed.

## Results

### Construction of the Test Statistic

We aim at constructing a statistic  $\tilde{K}$  to test whether a points' distribution is random or clustered by comparing its values with critical quantiles under CSR (Figure 1). A standard statistic is the Ripley's K function whose standard expression at distance scale  $r$ , and for  $n$  objects at position  $\mathbf{x}$  in a given field of view  $\Omega$ , is

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

where  $f(\mathbf{x},\mathbf{y})$  is a boundary correction term that prevents a bias in  $K(r,n)$  at larger values of  $r$  due to the finite size of  $\Omega$ . Indeed, some pairs of points closer than  $r$  can fall outside the observation window  $\Omega$ , leading to an underestimation of  $K$ . A widely used boundary correction is the Ripley's correction  $f(\mathbf{x},\mathbf{y}) = \frac{1}{2}(k(\mathbf{x},\mathbf{y}) + k(\mathbf{y},\mathbf{x}))$ , where  $k(\mathbf{x},\mathbf{y})$  is inversely proportional to the proportion of the circle  $\partial b(\mathbf{x},|\mathbf{x}-\mathbf{y}|)$  included in  $\Omega$ :  $k(\mathbf{x},\mathbf{y}) = \frac{|\partial b(\mathbf{x},|\mathbf{x}-\mathbf{y}|)|}{|\partial b(\mathbf{x},|\mathbf{x}-\mathbf{y}|) \cap \Omega|}$ . With this boundary correction and under CSR, ([22], page 39)

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

The problem when using the standard Ripley's K function (Eq.1) is that its mean and variance under CSR vary with distance scale  $r$ , which complicates its quantitative interpretation. A partial answer has been proposed by Besag who introduced the centered  $L$  function [23]  $L(r,n) = \sqrt{\frac{K(r,n)}{\pi}} - r$ . However,  $L$  function is not normalized and we thus propose a new statistic with zero mean and unit variance that uses the analytical expression of  $K$  variance.

The computation of the variance  $\text{var}\{K(r,n)\}$  of  $K(r,n)$  under CSR is made difficult by edge effects, but assuming that  $\Omega$  boundary is locally straight where it intersects  $b(\mathbf{x},|\mathbf{x}-\mathbf{y}|)$ , a closed-form expression of  $\text{var}\{K(r,n)\}$  has been obtained by Ripley [22]:

$$\text{var}\{K(r,n)\} = \frac{2|\Omega|^2 \beta(r)}{n^2} (1 + 0.305\gamma(r) + \beta(r)(-1 + 0.0132n\gamma(r))), \quad (3)$$

where  $\beta(r) = \frac{\pi r^2}{|\Omega|}$  and  $\gamma(r) = \frac{|\partial\Omega|r}{|\Omega|}$ .

Using the closed-form expressions of the variance (Eq. 3), we introduce the normalized and centered statistics

$$\tilde{K}(r,n) = \frac{K(r,n) - \pi r^2}{\sqrt{\text{var}\{K(r,n)\}}}, \quad (4)$$

whose significant deviations from 0 are characteristic of object clustering at length scale  $r$  when  $\tilde{K}(r,n) > 0$  or dispersion for

$\tilde{K}(r,n) < 0$  (Figure 1). To characterize these deviations statistically, we compute hereafter critical quantiles of  $\tilde{K}(r,n)$  under CSR.

**Estimation of  $\tilde{K}(r,n)$  Critical Quantiles Under CSR**

A first attempt of computing the critical quantiles of  $\tilde{K}$  analytically was proposed by Lang and colleagues [24]. They decompose  $\Omega$  in independent sub-domains, and using the central limit theorem, they prove that for  $n \gg 1$ ,  $\tilde{K}$  can be approximated by the standard normal law  $\mathcal{N}(0,1)$  under CSR ( $\tilde{K} \xrightarrow[n \gg 1]{} \mathcal{N}(0,1)$ ).

This is equivalent to approximate  $q_\alpha$ , the quantile at level  $\alpha$  of  $\tilde{K}$ , with  $z_\alpha$ , the quantile of the standard normal law  $\mathcal{N}(0,1)$ :  $q_\alpha \approx z_\alpha$ .

This approximation does not hold for intermediate values of  $n$  or for small distance scales  $r$  (see below), and we propose hereafter a general approximation of  $\tilde{K}$  quantiles that is valid for a large range of  $r$  and  $n$  values. This is based on the standard Cornish-Fisher (CF) expansion which is given by [25]

$$q_\alpha \approx z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)\gamma\{\tilde{K}\} + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)(\kappa\{\tilde{K}\} - 3) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)\gamma^2\{\tilde{K}\} \tag{5}$$

where  $\gamma\{\tilde{K}\} = \mathbb{E}\{\tilde{K}^3\}$  and  $\kappa\{\tilde{K}\} = \mathbb{E}\{\tilde{K}^4\}$  are respectively the skewness and the kurtosis of  $\tilde{K}$ . It can be deduced from Eq.5 that the CF expansion generalizes the central limit theorem which states that  $q_\alpha \approx z_\alpha$ .

At this point, we still face a problem because we do not have expressions for the skewness and the kurtosis of  $\tilde{K}$ , whose expressions are made difficult because of boundary effects. After long and mathematically involved computations which are detailed in File S1, the expressions of the skewness and the kurtosis of  $\tilde{K}$  are given by

$$\gamma\{\tilde{K}(r,n)\} = \frac{4a^3\beta(r)}{n^4\text{var}\{K(r,n)\}^{\frac{3}{2}}} [1 + 0.76\gamma(r) + n\beta(r)(1.173 + 0.414\gamma(r)) + n\beta^2(r)(-2 + 0.012n\gamma(r))], \tag{6}$$

and

$$\kappa\{\tilde{K}(r,n)\} = \frac{a^4\beta(r)}{n^6\text{var}\{K(r,n)\}^2} [8 + 11.52\gamma(r) + n\beta(r)((104.3 + 12n) + (78.7 + 7.32n)\gamma(r) + 1.116n\gamma^2(r)) + n\beta^2(r)((-304.3 - 1.92n) + (-97.9 + 2.69n + 0.317n^2)\gamma(r) + 0.0966n^2\gamma^2(r)) + n^2\beta^3(r)(-36 + 0.0021n^2\gamma^2(r))].$$

with  $\beta(r) = \frac{\pi r^2}{|\Omega|}$ ,  $\gamma(r) = \frac{|\partial\Omega|r}{|\Omega|}$  and  $\text{var}\{K(r,n)\}$  is given by Eq. 3.

At this point, we can make three comments: **1-** Setting apart the assumption that the boundary can be treated locally as a straight line, formulas for skewness and the kurtosis (Eq. 6–7) are exact. **2-** Reintroducing the approximations of variance, skewness and kurtosis (Eq. 3, 6 and 7) in the CF expansion (Eq. 5), we find that  $\tilde{K}(r,n)$  is asymptotically normal:  $\lim_{n \rightarrow \infty} q_\alpha = z_\alpha$  in agreement with [24]. **3-** In many applications,  $\tilde{K}(r,n)$  can be evaluated on  $M > 1$  fields of view and it is then interesting to use the mean statistics

$$\tilde{K}_M(r) = \frac{1}{M} \sum_{j=1}^M \tilde{K}^j(r,n_j) \tag{8}$$

where  $\tilde{K}^j(r,n_j)$  is evaluated on the  $j^{\text{th}}$  field of view. The CF expansion of  $\tilde{K}_M(r)$  quantiles then requires the computation of the skewness and the kurtosis of  $\tilde{K}_M(r)$ , which is detailed in File S1, section V.

**Assessing the Specificity of our Statistical Test on Synthetic Data**

To test the accuracy of the obtained CF approximation of  $\tilde{K}(r,n)$  quantiles (Eq. 5), we tested it against intensive Monte-Carlo resampling in a given field of view. In addition, we also compared the *true* quantiles obtained with simulations with the standard normal approximation.

To ensure the convergence of the Monte-Carlo method, we performed  $M = 10^6$  simulations where we drew uniformly  $n = 50$  points in a  $10 \times 10$  square  $\Omega$ . We then computed the corresponding Ripley's K function  $K_i(r,n)$  (Eq. 1), for  $1 \leq i \leq M$  and  $r$  varying from  $r = 0.3$  to  $r = 3$ . For each  $r$ , we computed the empirical variance  $\widehat{\text{var}}\{K(r,n)\} = \frac{1}{M} \sum_{i=1}^M (K_i(r,n) - \bar{K}(r,n))^2$  where  $\bar{K}(r,n) = \frac{1}{M} \sum_{i=1}^M K_i(r,n)$  is the empirical mean tending to  $\pi r^2$  for  $M \gg 1$ , and we then obtained

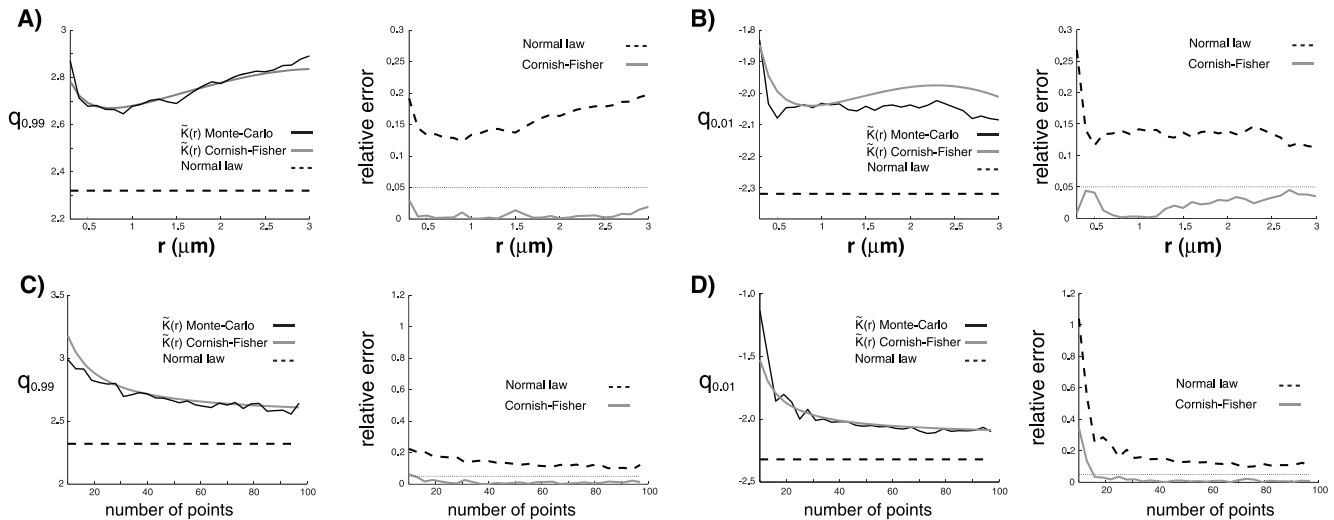
$$\tilde{K}_i(r,n) = \frac{K_i(r,n) - \bar{K}(r,n)}{\sqrt{\widehat{\text{var}}\{K(r,n)\}}} \tag{9}$$

The empirical quantile  $q_\alpha$  of  $\tilde{K}(r,n)$ , for  $\alpha = 0.01$  or  $\alpha = 0.99$  was then computed by sorting the  $\tilde{K}_i(r,n)$  and choosing  $q_\alpha = \tilde{K}_{\alpha M}(r,n)$  with the floor function of  $\alpha M$ .

In Figure 2 A–B, we compare quantiles obtained numerically with Monte-Carlo simulations with the CF expansion (Eq. 5) and the quantiles  $z_\alpha$  of the standard normal law  $\mathcal{N}(0,1)$  ( $z_{0.01} = -2.32$  and  $z_{0.99} = 2.32$ ). Interestingly, we observe that the CF expansion of  $q_\alpha$  (Eq. 5) with the asymptotical variance (Eq. 3), skewness and kurtosis (Eq. 6–7) is very close to Monte-Carlo simulations with a relative error below 5% even for  $r = 0.3$ , while the normal approximation is not satisfactory with a relative error that is around 20% for any  $r$ , and that reaches  $\approx 30\%$  for  $r = 0.3$  and  $\alpha = 0.01$ . The convergence of the CF expansion is linked to the mean number of pairs of points that are closer than  $r$ , which is  $\approx \frac{n^2 \pi r^2}{2|\Omega|}$ . Thus, the number of points  $n(r)$  that is needed for the relative error between the CF expansion and Monte-Carlo simulations to be below 5% should be approximately given by

$n(r) \approx \gamma \frac{|\Omega|^{\frac{1}{2}}}{r}$ . In particular, we found that for  $|\Omega|^{\frac{1}{2}} = 10$  and  $r = 0.3$ ,  $n(r) \approx 45$  indicating that  $\gamma \approx 1.5$ .

We next investigated the accuracy of CF development for an increasing number of points and a fixed  $r = 1$ . Results are given in Figure 2 C–D. We found that the relative error of the CF development to Monte-Carlo simulations is bounded by 5% for  $n \geq 10$  when  $\alpha = 0.99$  and  $n \geq 15$  when  $\alpha = 0.01$ , and fall below 2% for  $n \geq 30$  in both cases. Conversely, the relative error to normal approximation reaches 20% and 30% respectively for  $n = 10, \alpha = 0.99$  and  $n = 15, \alpha = 0.01$ , and is above 10% even for  $n = 100$ . We thus conclude that CF expansion of  $q_\alpha$  is sufficiently accurate to be used in a large range of  $r$  and  $n$  values while the



**Figure 2. Test of CF expansion against Monte-Carlo simulations.** A)-Left: The CF expansion (Eq. (5), grey line) of the quantile  $q_{0.99}$  of  $\tilde{K}(r,n)$  is tested against Monte-Carlo simulations ( $10^6$  simulations, solid black line) in a  $10 \times 10$  square  $\Omega$  for  $0.3 \leq r \leq 3$ . The number of points is set at  $n = 50$ . The quantile  $z_{0.99} = 2.32$  of the standard normal law  $\mathcal{N}(0,1)$  is also represented (black dotted line). A)-Right: Relative errors of CF expansion (grey line) and  $z_{0.99}$  (black dotted line) to Monte-Carlo simulations. The 5% level is represented with a black dotted line. B) Idem to A) for the first percentile  $q_{0.01}$  of  $\tilde{K}(r,n)$  instead of the last one  $q_{0.99}$ . C)-Left: The CF expansion (Eq. (5), grey line) of the quantile  $q_{0.99}$  of  $\tilde{K}(r,n)$  is tested against Monte-Carlo simulations ( $10^6$  simulations, solid black line) in a  $10 \times 10$  square  $\Omega$  for an increasing number of points  $10 \leq n \leq 100$ .  $r$  is set at  $r = 1$ . The quantile  $z_{0.99} = 2.32$  of the standard normal law  $\mathcal{N}(0,1)$  is also represented (black dotted line). C)-Right: Relative errors of CF expansion (grey line) and  $z_{0.99}$  (black dotted line) to Monte-Carlo simulations. The 5% level is represented with a black dotted line. D) Idem to C) for the first percentile  $q_{0.01}$  of  $\tilde{K}(r,n)$  instead of the last one  $q_{0.99}$ .  
doi:10.1371/journal.pone.0080914.g002

normal approximation does not hold even for intermediate values of  $n \approx 50 - 100$ . In addition, we highlight that for  $|\Omega|^{\frac{1}{2}} = 10$  and  $r = 1$ , we found that  $n(r) \approx 15$ , in agreement with  $\gamma \approx 1.5$ .

### Characterizing Objects' Dispersion and Clusters from $\tilde{K}(r,n)$ Statistic

To link the statistical deviations of  $\tilde{K}(r,n)$  from CSR ( $\tilde{K}(r,n) < q_\alpha$  and  $\tilde{K}(r,n) > q_{1-\alpha}$ , see Figure 1) to quantitative properties of point features, we show here how key features such as the minimal distance between dispersed points or the mean cluster size are related to  $\tilde{K}(r,n)$  extrema by using standard models of dispersed and clustered processes. While relating the minimum of  $\tilde{K}$  to the distance separating dispersed objects has not been treated, the relation between the maximum of Ripley's function to the clusters' size has been recently tackled numerically [12]. In their study, Kiskowski et al. modeled clusters with disk-shape domains with radius  $R_c$  that are regularly separated by a distance  $S$ , and they used Monte-Carlo simulations where a part  $\beta n$  of points was randomly distributed in clusters and  $(1 - \beta)n$  points were distributed outside the clusters. They found that the radius of maximal aggregation  $r_{max}$  where the Besag L-function  $L(r,n) = \sqrt{\frac{\tilde{K}(r,n)}{\pi}} - r$  reaches its maximum was between  $R_c$  and  $2R_c$  depending on  $S$ .

To extract the minimal distance between objects in a regular pattern from the minimum of  $\tilde{K}(r,n)$  at small distance scale  $r$ , we model the local objects' organization with a simple inhibition process (chapter 5 [10]), which is a thinned Poisson process (intensity  $\rho$ ) where all pairs of points a distance less than arbitrary  $\delta$  apart would be deleted. Then, the related parametric Ripley's K function reads ([10], page 72)

$$K(r,n) = 1_{\{r > \delta\}} 2\pi \int_{\delta}^r e^{\rho(\pi\delta^2 - U_{\delta}(t))} dt, \quad (10)$$

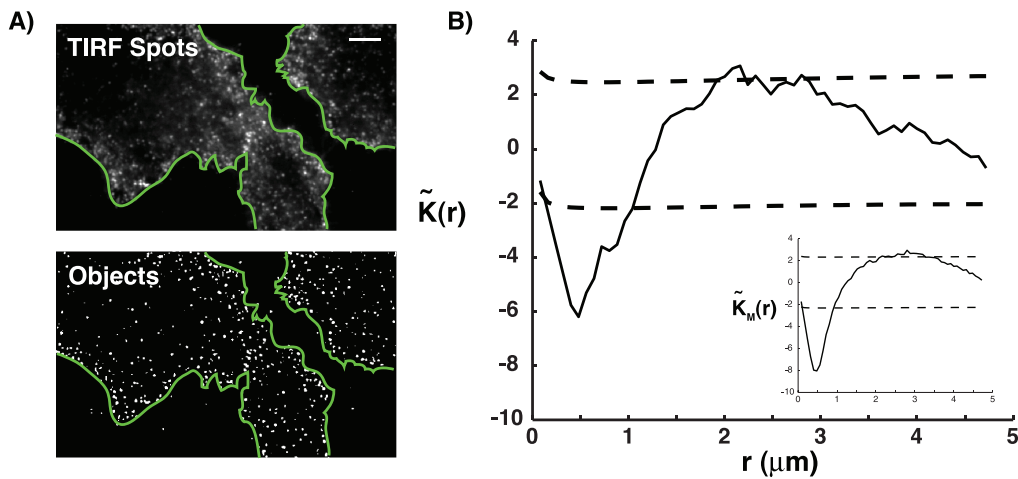
where  $U_{\delta}(t)$  denotes the area of the union of two discs each of radius  $\delta$  and with centers a distance  $t$  apart, that is [26]:  $U_{\delta}(t) = 2\delta^2 \cos^{-1}\left(\frac{t}{2\delta}\right) - \frac{t}{2} \sqrt{4\delta^2 - t^2}$ . Reinjecting the parametric expression (Eq. 10) of the dispersed  $K$  function in  $\tilde{K}(r,n)$  (Eq. 4), we compute the partial derivative  $\frac{\partial \tilde{K}(r,n)}{\partial r}$  of  $\tilde{K}(r,n)$  with respect to  $r$  and obtain that  $\frac{\partial \tilde{K}(r,n)}{\partial r} < 0$  for  $r < \delta$  and  $\frac{\partial \tilde{K}(r,n)}{\partial r} > 0$  for  $r \geq \delta$ , which demonstrates that in an idealized inhibition process, the minimal distance  $\delta$  that separates points from each other is equal to  $r_{min}$  where  $\tilde{K}(r,n)$  reaches its minimum:

$$\delta = r_{min}. \quad (11)$$

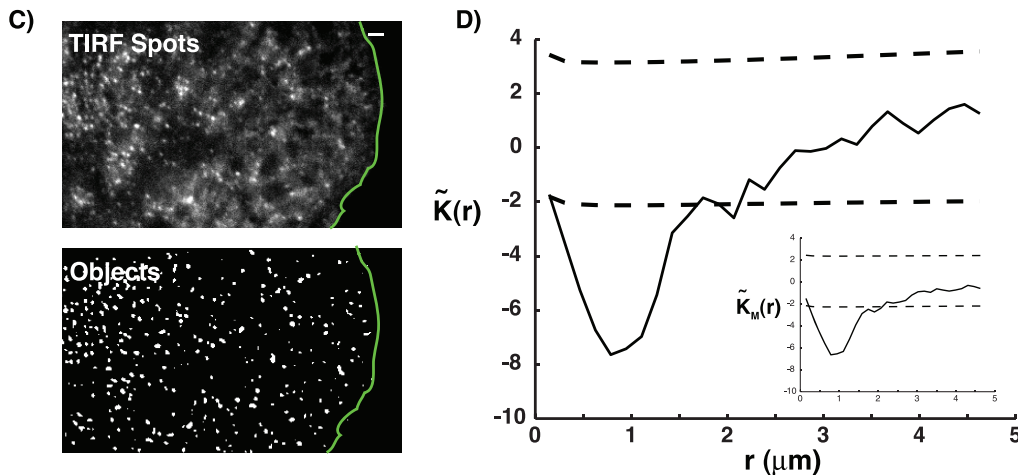
To relate the radius  $r_{max}$  where  $\tilde{K}$  reaches its maximum to the mean clusters' radius  $R_c$ , we assume here that clusters' centers are randomly distributed in  $\Omega$  (density  $\lambda_c$ ), and in that case, the analytical expression of the Ripley's K function is then given by [27], page 376:

$$K(r,n) = \pi r^2 + \frac{2\beta^2}{\pi\lambda_c} \left( \arcsin\left(\frac{r}{2R_c}\right) + \frac{r^2}{R_c^2} \arccos\left(\frac{r}{2R_c}\right) - \frac{r}{4R_c^3} (2R_c^2 + r^2) \sqrt{1 - \frac{r^2}{4R_c^2}} \right). \quad (12)$$

## IL-2R (clathrin-independent)



## Clathrin



**Figure 3. Analysis of endocytosis spatial organization.** (A) Clathrin-independent IL-2R putative endocytic sites. Top: IL-2R is labeled with fluorescent antibodies and imaged using total internal reflexion fluorescence (TIRF) microscopy. Bottom: We delimited individual cells by drawing polygonal (green) Regions of Interest (ROIs) in the software Icy [13] (<http://icy.bioimageanalysis.org>). Positions of putative endocytic sites (objects) inside each cell are then extracted with a multi-scale wavelet analysis [28]. (B) The spatial organization of IL-2R putative endocytic spots is quantified with  $\tilde{K}(r,n)$  (solid black line). CF expansion of  $q_{0.01}$  and  $q_{0.99}$  is represented with black dotted lines. In the bottom-right corner, the mean statistic  $\tilde{K}_M(r)$  (Eq. 8) is plotted against  $r$  ( $M=7$  cells, 4382 objects). (C) Clathrin putative endocytic sites. Top: Clathrin light chain is fused with green fluorescent protein (GFP) and imaged using TIRF microscopy. Bottom: Positions of putative endocytic sites are extracted with a multi-scale wavelet analysis [28]. (D) The spatial organization of clathrin putative endocytic spots is quantified with  $\tilde{K}(r,n)$  (solid black line). CF expansion of  $q_{0.01}$  and  $q_{0.99}$  is represented with black dotted lines. In the bottom-right corner, the mean statistic  $\tilde{K}_M(r)$  (Eq. 8) is plotted against  $r$  ( $M=5$  cells, 2195 objects). Scale bar = 5 microns.

doi:10.1371/journal.pone.0080914.g003

Reinjecting this parametric expression (Eq.12) in  $\tilde{K}(r,n)$ , we link the radius  $r_{max}$  of maximal aggregation with clusters' radius  $R_c$  by solving numerically  $\frac{\partial \tilde{K}(r,n)}{\partial r}(r_{max})=0$ , and find that

$$R_c \approx \frac{r_{max}}{1.3}. \quad (13)$$

### Quantitative Analysis of Endocytosis Spatial Organization

We analyzed two image data sets representative each of clathrin-dependent ( $M=5$  cells,  $n=2195$  points) and clathrin-independent ( $M=7$  cells,  $n=4382$  points) pathways. In each case, after extracting the positions of putative endocytic events thanks to a wavelet-based detection [28] (Figure 3 A–C), we computed the modified Ripley's K function  $\tilde{K}(r,n)$  (Eq. (4)) and CF expansions (Eq. (5)) of quantiles  $q_{0.01}$  and  $q_{0.99}$  in Figure 3 B–D. We first

checked that the characteristic features of  $\tilde{K}$  functions were similar whether they were computed on one cell or averaged on several cells (see inserted box in lower corners Figure 3 B–D).

We found that for both pathways,  $\tilde{K}(r,n)$  is far below the first percentile  $q_{0.01}$  for small  $r < 1 \mu\text{m}$  (IL-2R) and  $r < 2 \mu\text{m}$  (clathrin), indicating that putative endocytic spots are distributed according to a regular pattern, characterized by a minimum distance between points, which corresponds to the value that minimizes  $\tilde{K}(r,n)$ . In the case of clathrin-independent pathway,  $\tilde{K}(r,n)$  reaches its minimum at  $r_{\min}^{\text{IL-2R}} \approx 500 \text{ nm}$  while for clathrin-dependent entry,  $r_{\min}^{\text{clathrin}} \approx 800 \text{ nm}$ . This demonstrates that endocytic sites are non-overlapping and restricted to defined microdomains with respective radii  $\approx 400 \text{ nm}$  for clathrin-dependent endocytosis and  $\approx 250 \text{ nm}$  for clathrin-independent pathway.

At higher distance scale  $r$ , we found that for clathrin-dependent endocytosis,  $\tilde{K}(r,n)$  is comprised between the quantiles  $q_{0.01}$  and  $q_{0.99}$  indicating that clathrin spots are homogeneously distributed on the membrane. Repeating our statistical analysis using labeled transferrin, which is the archetypical cargo for internalization through clathrin-mediated endocytosis [29], we got identical profiles to those obtained with clathrin (Figure S1). By contrast, for clathrin-independent pathway,  $\tilde{K}(r,n)$  is above  $q_{0.99}$  for  $r$  between 2 and 3  $\mu\text{m}$  indicating that clathrin-independent spots are partially organized in clusters. Considering that  $R_c \approx \frac{r_{\max}}{1.3}$  (Eq. 13), we deduced that clathrin-independent spots are partially segregated in clusters with radius  $R_c \approx 2 \mu\text{m}$ .

## Discussion

We have developed a new test statistic based on the Ripley's K function that facilitates the quantitative analysis of the spatial organization of point patterns at multiple scales. This test allows us to statistically assess the presence of specific point patterns such as point clusters or dispersion with no need for resampling by providing an asymptotic closed-form expression of the critical quantiles of the Ripley's K function under spatial randomness. In addition, we related the extrema of our statistics to the geometrical properties of the observed patterns by using standard models of dispersed and clustered point patterns.

We applied our method to study the spatial organization of molecules implicated in different endocytosis pathways, and we found that the spatial organization of endocytosis was different upon the mechanism (dependent or independent of clathrin), which might reflect distinct cellular functions of each pathway. We note that all clathrin and IL-2R spots are not necessarily entering the cell, as some spots might disassemble or detach before being endocytosed [30]. It would thus be interesting to couple our statistical analysis with live cell imaging to compare the spatial organization of real endocytic events and abortive ones.

A major difficulty in Ripley-based statistical tests is their interpretation when the null hypothesis of objects' random distribution is rejected. In particular for IL-2R receptors, the detected aggregation could result either from small clusters, or from a very local increase of the receptors' density near the cell boundary, or from a mixture of both. We have thus repeated our analysis by eroding the cell's contour mask by 300 nm (isotropic ball of radius 3 pixels) and 500 nm (5 pixels) to test boundary effects. Interestingly, we found profiles very similar to those obtained with the whole cell (Figure S2) with a maximum of the Ripley's K function reached for  $r \approx 3$  microns as above. We thus conclude that the local increase of IL-2R receptors at cell boundary does not have much impact on the behavior of Ripley's K function and that receptors are truly organized in clusters with a radius of 2 microns.

In this study, we developed a robust and fast analytical method to test whether an objects' distribution deviates from CSR. A promising extension would be to test whether the spatial organization of points can be described with some specific parametric models, in particular the large classes of Neyman-Scott [27,31] or Strauss [27,32] processes. This would open the door to analytical comparison of points' distributions against each other through embedding and statistical learning.

## Materials and Methods

### Experimental Protocol, TIRF Microscopy

For clathrin-independent endocytosis, Hep2 $\beta$  cells ( $1 \times 10^5$ ) expressing IL-2R were incubated 2 min with anti-IL-2R coupled to Cy3 fluorochrome in a TIRF medium (25 mM Hepes, 135 mM NaCl, 5 mM KCl, 1.8 mM CaCl<sub>2</sub>, 0.4 mM MgCl<sub>2</sub>, 4.5 g/L glucose, pH 7.4 and 0.5% BSA) at 37 C and washed. For clathrin-dependent endocytosis BSC-1 cells, expressing clathrin-light chain fused to GFP were used. Cells were incubated in an environmental control system set to 37 C and movies of 100 s at 1Hz were acquired. Experiments were performed using a TIRF microscope (IX81F-3, Olympus) equipped with a 100x NA 1.45 Plan Apo TIRFM Objective (Olympus) and fully controlled by CellM (Olympus).

### Quantitative Image Analysis

We first delimited cells' contours by drawing polygonal Region of Interest (ROIs) with the Icy software [13] (<http://icy.bioimageanalysis.org>). We then used a wavelet-based detection method [28], implemented as a plugin *Spot detector* in Icy to extract the two dimensional positions of putative endocytic spots at the cellular membrane. In the clathrin-independent pathway, a part of IL-2R spots diffused at the cell membrane and we extracted the signal corresponding to static spots entering the cell by first stacking time sequences in a single image (mean), and by then applying our wavelet-based detection algorithm on the stacked image.

### Supporting Information

**Figure S1 Analysis of the spatial organization of transferrin endocytosis.** (A) Clathrin-dependent transferrin putative endocytic sites. Top: Transferrin is labeled with fluorescent antibodies and imaged using total internal reflexion fluorescence (TIRF) microscopy. Bottom: We delimited manually individual cells by drawing polygonal (green) Regions of Interest (ROIs) in the software Icy [13] (<http://icy.bioimageanalysis.org>). Positions of putative endocytic sites (objects) inside each cell are then extracted with a multi-scale wavelet analysis [28]. (B) The spatial organization of Transferrin putative endocytic spots is quantified with the mean statistic  $\tilde{K}_M(r)$  (Eq. (8) of the main manuscript,  $M = 3$  cells (1086 objects), solid black line). Cornish-Fisher expansion of  $q_{0.01}$  and  $q_{0.99}$  (Eq. (5) of the main manuscript) are represented with black dotted lines. (EPS)

**Figure S2 Analysis of the spatial organization of clathrin-independent endocytosis with erosions of the cell's contour mask.** We have tested the impact of the local accumulation of IL-2R spots at the cell boundary by eroding the cell's contour mask by 300 nm (isotropic ball of radius 3 pixels) and 500 nm (5 pixels) to test boundary effects. The spatial organization of IL-2R putative endocytic spots is quantified with the mean statistic  $\tilde{K}_M(r)$  (Eq. (8) of the main manuscript,  $M = 7$  cells (4382 objects)) for no erosion (black line), 300 nm-erosion (blue line) and 500 nm-erosion (green line). Cornish-Fisher

expansion of  $q_{0.95}$  and  $q_{0.99}$  (Eq. (5) of the main manuscript) are represented with black dotted lines.  
(EPS)

**File S1 Supplementary Methods** Detailed computations of the skewness and the kurtosis of the Ripley's K function.  
(PDF)

## References

- Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21: 256–274.
- White A, Barengi C, Proukakis N (2012) Creation and characterization of vortex clusters in atomic bose-einstein condensates. *Physical Review A* 86: 013635.
- Draper G (1991) *The Geographical Epidemiology of Childhood Leukemia and Non-Hodgkin Lymphomas in Great Britain, 1966–83* (Studies on Medical and Population Subjects, volume 53. HMSO, London.
- Peterson C, Squiers E (1995) An unexpected change in spatial pattern across 10 years in aspen-white pine forest. *Journal of Ecology* 83: 847–855.
- Jafari-Mamaghani M, Andersson M, Krieger P (2010) Spatial point pattern analysis of neurons using ripley's k-function in 3d. *Front Neuroinform* 4: 9.
- Prior IA, Muncke C, Parton RG, Hancock JF (2003) Direct visualization of ras proteins in spatially distinct cell surface microdomains. *J Cell Biol* 160: 165–70.
- Parton RG, Hancock JF (2004) Lipid rafts and plasma membrane micro-organization: insights from ras. *Trends Cell Biol* 14: 141–7.
- Lasserre R, Charrin S, Cuhe C, Danckaert A, Thoulouze MI, et al. (2010) Ezrin tunes t-cell activation by controlling dgl1 and microtubule positioning at the immunological synapse. *EMBO J* 29: 2301–14.
- Mounier J, Boncompain G, Senerovic L, Lagache T, Chrétien F, et al. (2012) Shigella effector ipab-induced cholesterol relocation disrupts the golgi complex and recycling network to inhibit host cell secretion. *Cell Host Microbe* 12: 381–9.
- Diggle P (2003) *Statistical analysis of spatial point patterns*. Hodder Education, second edition.
- Ripley BD (1977) Modelling spatial patterns. *Journal of the Royal Statistical Society B* 39: 172–212.
- Kiskowski M, Hancock J, Kenworthy AK (2009) On the use of ripley's k-function and its derivatives to analyze on the use of ripley's k-function and its derivatives to analyze domain size. *Biophys J* 97: 1095–1103.
- de Chaumont F, Dallongeville S, Chenouard N, Hervé N, Pop S, et al. (2012) Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods* 9: 690–6.
- Ehrlich M, Boll W, Van Oijen A, Hariharan R, Chandran K, et al. (2004) Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* 118: 591–605.
- Diaz E, Sebastian R, Ayala G, Diaz ME, Zoncu R, et al. (2008) Measuring spatiotemporal dependencies in bivariate temporal random sets with applications to cell biology. *IEEE Trans Pattern Anal Mach Intell* 30: 1659–71.
- Mettlen M, Loerke D, Yazar D, Danuser G, Schmid SL (2010) Cargo- and adaptor-specific mechanisms regulate clathrin-mediated endocytosis. *J Cell Biol* 188: 919–33.
- Nunez D, Antonescu C, Mettlen M, Liu A, Schmid SL, et al. (2011) Hotspots organize clathrin-mediated endocytosis by efficient recruitment and retention of nucleating resources. *Traffic* 12: 1868–78.
- Lamaze C, Dujancourt A, Baba T, Lo CG, Benmerah A, et al. (2001) Interleukin 2 receptors and detergent-resistant membrane domains define a clathrin-independent endocytic pathway. *Mol Cell* 7: 661–71.
- Gesbert F, Sauvonnnet N, Dautry-Varsat A (2004) Clathrin-Independent endocytosis and signalling of interleukin 2 receptors il-2r endocytosis and signalling. *Curr Top Microbiol Immunol* 286: 119–48.
- Liao W, Lin JX, Leonard WJ (2011) Il-2 family cytokines: new insights into the complex roles of il-2 as a broad regulator of t helper cell differentiation. *Curr Opin Immunol* 23: 598–604.
- Sigismund S, Confalonieri S, Ciliberto A, Polo S, Scita G, et al. (2012) Endocytosis and signaling: cell logistics shape the eukaryotic cell plan. *Physiol Rev* 92: 273–366.
- Ripley B (1988) *Statistical inference for spatial processes*. Cambridge University Press.
- Besag JE (1977) Comments on ripley's paper. *Journal of the Royal Statistical Society B* 39: 193–195.
- Lang G, Marcon E (2013) Testing randomness of spatial point patterns with the ripley statistic. *ESAIM: Probability and Statistics* (accepted).
- Cornish E, Fisher R (1937) Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute* 5: 307–320.
- mathworldwolfram website (E.W. Weisstein) Available: <http://mathworld.wolfram.com/circle-circleintersection.html>. Accessed 2013 Oct 31.
- Illian J, Penttinen A, Stoyan D, Stoyan H (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley-Blackwell.
- Olivo-Marin JC (2002) Extraction of spots in biological images using multiscale products. *Pattern Recognition* 35: 1989–1996.
- Le Roy C, Wrana JL (2005) Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. *Nat Rev Mol Cell Biol* 6: 112–26.
- Loerke D, Mettlen M, Yazar D, Jaqaman K, Jaqaman H, et al. (2009) Cargo and dynamin regulate clathrin-coated pit maturation. *PLoS Biol* 7: e57.
- Neyman J, Scott E (1952) A theory for the spatial distribution of galaxies. *Astrophys J* 116: 144–163.
- Strauss D (1975) A model for clustering. *Biometrika* 62: 467–475.

## Author Contributions

Conceived and designed the experiments: NS. Performed the experiments: NS. Analyzed the data: TL. Contributed reagents/materials/analysis tools: TL GL. Wrote the paper: TL JCOM.