

RESEARCH

Open Access



# Nfinder: automatic inference of cell neighborhood in 2D and 3D using nuclear markers

Bruno Moretti<sup>1,2,3,5\*†</sup>, Santiago N. Rodriguez Alvarez<sup>1,4†</sup> and Hernán E. Grecco<sup>1,2\*</sup>

<sup>†</sup>Bruno Moretti and Santiago N. Rodriguez Alvarez are contributed equally to this work.

\*Correspondence: morettibruno@gmail.com; hgrecco@df.uba.ar

<sup>1</sup> Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, Buenos Aires, Argentina

<sup>2</sup> CONICET - Universidad de Buenos Aires, Instituto de Física de Buenos Aires (IFIBA), Buenos Aires, Argentina

<sup>3</sup> Present Address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, USA

<sup>4</sup> Present Address: Institute of Physics, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

<sup>5</sup> Present Address: Chan Zuckerberg Biohub—San Francisco, San Francisco, USA

## Abstract

**Background:** In tissues and organisms, the coordination of neighboring cells is essential to maintain their properties and functions. Therefore, knowing which cells are adjacent is crucial to understand biological processes that involve physical interactions among them, e.g. cell migration and proliferation. In addition, some signaling pathways, such as Notch or extrinsic apoptosis, are highly dependent on cell–cell communication. While this is straightforward to obtain from membrane images, nuclei labeling is much more ubiquitous for technical reasons. However, there are no automatic and robust methods to find neighboring cells based only on nuclear markers.

**Results:** In this work, we describe Nfinder, a method to assess the cell's local neighborhood from images with nuclei labeling. To achieve this goal, we approximate the cell–cell interaction graph by the Delaunay triangulation of nuclei centroids. Then, links are filtered by automatic thresholding in cell–cell distance (pairwise interaction) and the maximum angle that a pair of cells subtends with shared neighbors (non-pairwise interaction). We systematically characterized the detection performance by applying Nfinder to publicly available datasets from *Drosophila melanogaster*, *Tribolium castaneum*, *Arabidopsis thaliana* and *C. elegans*. In each case, the result of the algorithm was compared to a cell neighbor graph generated by manually annotating the original dataset. On average, our method detected 95% of true neighbors, with only 6% of false discoveries. Remarkably, our findings indicate that taking into account non-pairwise interactions might increase the Positive Predictive Value up to + 11.5%.

**Conclusion:** Nfinder is the first robust and automatic method for estimating neighboring cells in 2D and 3D based only on nuclear markers and without any free parameters. Using this tool, we found that taking non-pairwise interactions into account improves the detection performance significantly. We believe that using our method might improve the effectiveness of other workflows to study cell–cell interactions from microscopy images. Finally, we also provide a reference implementation in Python and an easy-to-use napari plugin.

**Keywords:** Microscopy, Image analysis, Cell–cell interactions, Tissues, Neighboring cells, Delaunay triangulation



## Background

Understanding the emergence of tissue organization requires quantifying the local and global signaling cues that each cell perceives from the environment. Among these cues, cell–cell interaction is one of the strongest to determine cellular fate during development and maintain homeostasis in adult life. For example, cell–cell contact has an important role during cell migration. A cell in a migrating tissue can change the direction of its movement after direct contact with another cell, in a process that is known as contact inhibition of locomotion (CIL). The number of neighboring cells can alter the outcome of CIL: in a sheet of cells with a free edge, only the cells at the free edge produce lamellipodia, which in turn induces a collective and directed migration of the whole tissue [1]. The number of neighboring cells in a tissue also plays an important role in proliferation. When normal, noncancerous cells are in crowded conditions, they cease proliferation and cell division. This characteristic is lost when cells undergo malignant transformation, resulting in uncontrolled proliferation and tumor formation [2].

At the molecular level, cell–cell interaction is also crucial in a plethora of biological processes. For example, the Notch signaling pathway is highly conserved and plays a key role in cell–cell communication. Since most of this pathway's ligands are transmembrane proteins, signalling is restricted to neighboring cells [3]. In vertebrate embryonic development, the morphological segments—termed somites—that prefigure the bones and muscles of the adult are formed at the posterior end of the elongating body axis. A highly conserved biological clock consisting of an oscillatory gene regulatory network rhythmically differentiates cells into somites [4]. Here, the Notch signaling pathway plays a key role in synchronizing oscillations between neighboring cells [5]. Disruption of the Notch pathway during development causes severe somite malformation in zebrafish, chick and mouse embryos [6–11].

Thus, knowledge of neighbor relations between cells in tissues is crucial to understanding the emergent collective behaviours that arise from individual cell–cell interactions. Assessing neighborhood in tissues using fluorescence microscopy, i.e. precisely identifying which cells share a border, requires labeling cellular membranes. While this is routinely done in a variety of experiments, it is also many times put aside for technical reasons or convenience. For instance, nuclear markers provide a better signal-to-noise ratio and are spatially well-separated, which simplifies cell counting and identification [12–14]. Due to their brightness, they are also commonly used to perform autofocus routines and track cell movements over time [15–18]. Although it would be ideal to use both nuclei and membrane labelling, this is more expensive, takes a longer acquisition time and adds experimental complexity. For these reasons, nuclei labeling is ubiquitously found in literature both for fixed and live-cell imaging, being DAPI and H2B-FP among the most common.

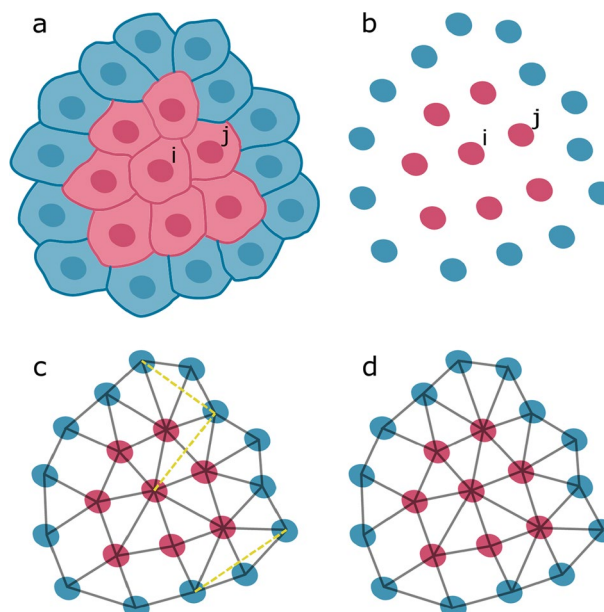
In this work, we developed Nfinder, a fast, automatic and robust method to find the first neighbors of cells in tissues using only a nuclear marker. To validate our method, we applied it to datasets containing both a nuclear and a membrane marker from a variety of organisms such as *Drosophila melanogaster*, *Tribolium castaneum*, *Arabidopsis thaliana* and *C. elegans*. These datasets were manually analyzed to generate ground truth images, which were in turn used to compare our automatically-found results against. We

found that Nfinder is robust and precise for all the test datasets we used, covering 95% of true neighboring relationships with 94% precision, on average.

## Methods

### Delaunay triangulation approximates neighboring cells

The goal of our method is to estimate which cells are in contact based only on the positions of nuclei centroids (Fig. 1a, b). Considering this restriction, we define neighboring cells (Fig. 1, red cells) as those which are (1) close together and (2) not interfered by any other cell. This implies that the building blocks of the cell graph are not edges but triangles, because they are the minimal structures that can provide information about the influence of the graph over a given pair of nodes. Consequently, a triangulation of nuclei centroids could be a good approximation of the neighboring cells network (Fig. 1c). In particular, a Delaunay triangulation is outstanding for this task due to its special properties, which result from being the dual of the Voronoi diagram [19]. The latter divides a set of points  $S$  into touching convex regions, each of which contains only one point in  $S$ . If two regions are in contact in the Voronoi diagram, they are linked in the Delaunay triangulation. As a result of this, the Delaunay triangulation includes the nearest neighbor graph, which is a necessary (but not sufficient) requirement for any approximation of cell–cell physical interactions. In addition to this, given that Voronoi cells are geometrically stable with respect to small changes of the points in  $S$ , so the edges of the Delaunay triangulation are [20]. This property ensures that small displacements of cells do not affect the results, distinguishing this triangulation from others. Another essential feature in planar spaces is that the Delaunay triangulation maximizes the minimum angle in any



**Fig. 1** Overview of the proposed method. **a** System of touching cells, showing membranes and nuclei. Cell  $i$  and its neighbors  $j$  are highlighted (red). **b** Nuclei corresponding to the cells shown in (a). **c** Triangulation of nuclei centroids, including neighboring cells (gray, solid lines) and typical errors (yellow, dashed lines). **d** Final result obtained by filtering the triangulation shown in (c)

triangle, avoiding sharp structures that would appear if cells with an obstacle in between were assumed to be connected [21].

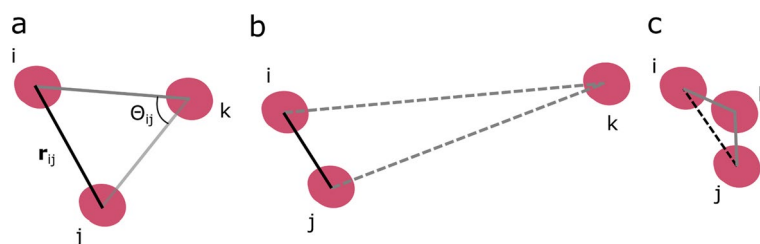
### Automatic communicability efficiency filter

By virtue of these attributes, we designed Nfinder, a method to find the first neighbors of cells within tissues using only nuclear markers, based on an implementation of Delaunay triangulations present in the open-source library SciPy [22]. As mentioned before, the Delaunay triangulation is necessary but not sufficient to determine true neighboring relations and, therefore, a filtering procedure is needed (Fig. 1d). To achieve this goal, we characterized each pair of nuclei ( $i, j$ ) by their distance  $r_{ij}$  and the maximum angle subtended with other nuclei  $\Theta_{ij}$  (Fig. 2a). The latter measures the interplay between ( $i, j$ ) and its neighborhood, as it informs about how other cells might be blocking the path connecting  $i$  and  $j$ .

To understand the role of  $\Theta_{ij}$ , it is instructive to analyze the possible outcomes for a set of three cells ( $i, j, k$ ) in two extreme scenarios. If ( $i, j$ ) are close to each other while  $k$  is very distant,  $\Theta_{ij}$  is very small, indicating that ( $i, j$ ) cannot be interfered by  $k$  (Fig. 2b). On the other hand, when  $k$  is near to the midpoint of the pair, it will be very unlikely that ( $i, j$ ) get in contact, no matter how close they are (Fig. 2c). Here, the relevance of  $\Theta_{ij}$  can be appreciated: its value depends on the local spacing between cells, ranging from  $0^\circ$ , when ( $i, j$ ) are isolated, to  $180^\circ$ , when there is a neighbor in between. Accordingly, ( $i, j$ ) are said to be neighbors if  $r_{ij} \leq r^*$  and  $\Theta_{ij} \leq \Theta^*$ , where  $r^*$  and  $\Theta^*$  are the maximum values of distance and angle determined by the preference of the biological system towards a particular spatial organization. To estimate these parameters, and based on a previous work that introduced the concept of communicability angle between a pair of nodes in a graph [23], we developed a new statistic that measures cell–cell communicability efficiency, defined as

$$E(r, \Theta) = F(r, \Theta) - L(r, \Theta), \quad (1)$$

where  $F$  is the fraction of connected pairs and  $L$  is the communicability efficiency loss given that the thresholds are  $r$  and  $\Theta$ . Intuitively, the communicability efficiency is a metric that weighs the actual number of connected cells against the number of cells that would be connected if the spatial distribution of cells were random. Consistently, the



**Fig. 2** Characterization of the neighboring cells network. **a** Nuclei (red) are connected by solid edges representing cell membrane physical contact. Edge ( $i, j$ ) is characterized by nucleus-nucleus distance  $r_{ij}$  and the maximum angle subtended with nearby nuclei  $\Theta_{ij}$ . **b** Small-angle limit, showing  $i$  and  $j$  close together while  $k$  is far away. As a result of this, ( $i, j$ ) are neighbors and  $k$  is isolated from both  $i$  and  $j$ , which is indicated by dashed edges. **c** Large-angle limit, in which  $k$  is potentially blocking the connection between  $i$  and  $j$

maximum communicability efficiency would be achieved by connecting the maximum number of cells at the minimum cost. The proposed critical values  $r^*$  and  $\Theta^*$  are those that maximize the communicability efficiency of the cell graph

$$(r^*, \Theta^*) = \arg \max E(r, \Theta). \quad (2)$$

When the thresholds are  $r_{min}$  and  $\Theta_{min}$ , there are no connectivity losses ( $L = 0$ ). However, this goes to the detriment of the communicability efficiency, as only one pair would be connected. Conversely, when the thresholds are  $r_{max}$  and  $\Theta_{max} = 180^\circ$ , then  $L = 1$  and  $F = 1$ , leading to null communicability efficiency. For this reason, maximizing the communicability efficiency provides a good thresholding criteria, as it balances out the costs and benefits of increasing the critical values of distance and angle.

It is worth noticing that  $F$  is simply calculated as the Empirical Cumulative Distribution Function (ECDF) of  $r$  and  $\Theta$ , which corresponds to the fraction of pairs with  $r_{ij} \leq r$  and  $\Theta_{ij} \leq \Theta$ . On the other hand,  $L$  is not unequivocally determined by  $r$  and  $\Theta$ . The only constraint on the loss function is that it must be an increasing function of both  $r$  and  $\Theta$ . In this paper, we show that measuring loss as

$$L_{mean}(r, \Theta) = \frac{1}{2} \left( \frac{r - r_{min}}{r_{max} - r_{min}} + \frac{\theta - \theta_{min}}{\theta_{max} - \theta_{min}} \right) \quad (3)$$

yields accurate predictions on the neighbor graph. Mathematically,  $L_{mean}$  is the average probability of randomly choosing  $r$  and  $\Theta$  as the critical values. Since biological systems are not completely random, not all  $r$  and  $\Theta$  are equally probable. As a result of this, there must be some values  $(r^*, \Theta^*)$  that maximize the communicability efficiency of the cell graph and, therefore, are good estimators of the critical values of distance and angle. Source code of a reference implementation in Python for 2D and 3D datasets as well as a user-friendly napari plugin can be found at <https://github.com/santi-rodriguez/nfinder>.

#### Performance metrics used to validate Nfinder

To validate our method, we applied it to the nuclear channel of datasets containing both nuclear and membrane labelling. For each dataset, the double staining allowed us to manually generate the neighboring cells graph of each dataset. First, we segmented and labelled each nucleus via intensity thresholding. Based on this labelling, we manually created a list of connected nuclei defined as those whose membranes were in contact according to what the membrane channel showed. This list of touching cells was used as a ground truth for the neighboring cells graph and compared to our automatic estimation based on nuclear information only. Edges from the estimated graph that matched the ground truth were labelled as true positives (TP). On the contrary, all pairs connected by the algorithm that were not actual neighbors were recorded as false positives (FP). Finally, false negatives (FN) were assigned to touching cells that the algorithm did not detect, while true negatives (TN) corresponded to those edges that were not connected in the ground truth nor in the estimated cell graph.

The metrics used to quantify the performance of the method were the True Positive Rate (TPR)

$$TPR = \frac{TP}{TP + FN}. \quad (4)$$

Positive Predictive Value (PPV)

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

and Jaccard Index (JI)

$$JI = \frac{TP}{TP + FN + FP}. \quad (6)$$

The TPR measures the coverage of positives, while the PPV indicates the fraction of true neighbors among all predictions. Finally, the JI reports the overlap between the list of computed cell–cell edges and the ground truth. Additionally, in the Additional file 1, we show the F1-score

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (7)$$

and the Matthews correlation coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}. \quad (8)$$

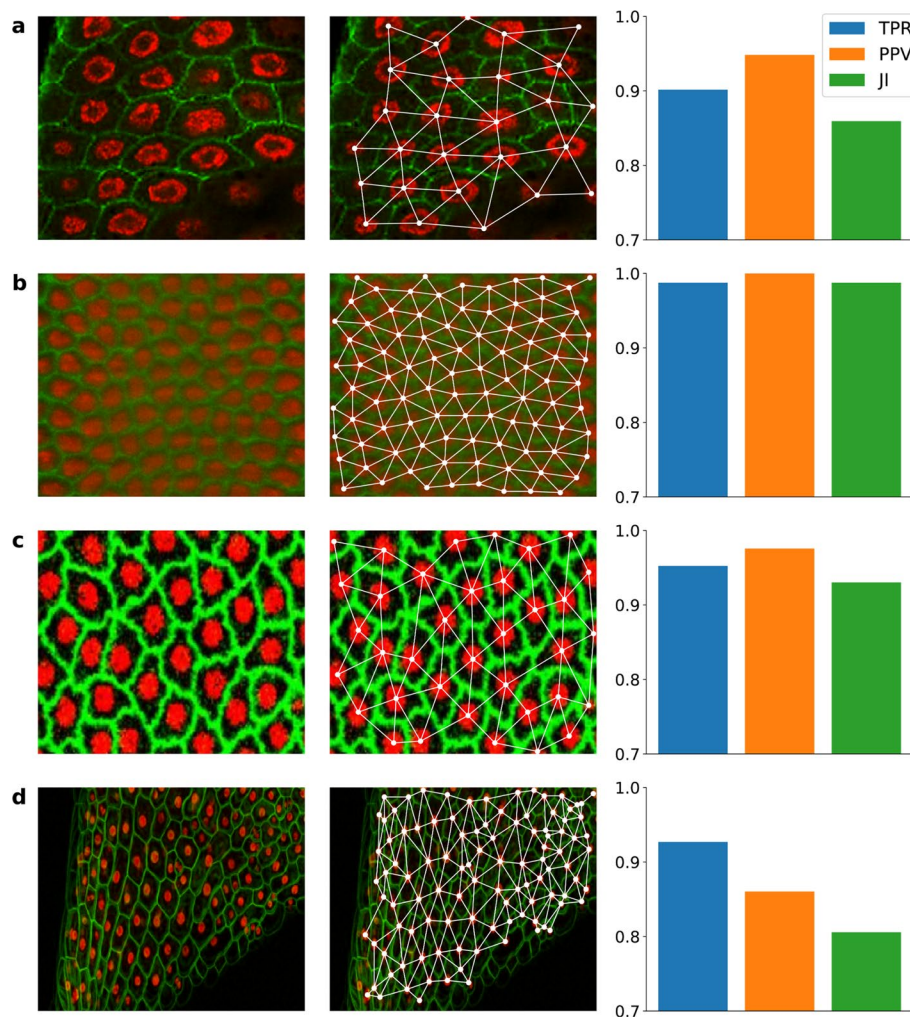
The F1-score represents the harmonic mean of PPV and TPR, while the MCC is the Pearson correlation coefficient between the ground truth classification and the output of the algorithm.

Remarkably, Nfinder is deterministic and scale, reflection and rotational invariant. Thus, it does not depend on knowing the typical distance between neighboring cells, as this is inferred from the image itself. In the next section, we show that neighboring cells can be precisely detected with this method in very diverse tissues.

## Results

### Delaunay triangulation-based method can be applied to diverse tissues

As a proof of concept, we applied Nfinder to different datasets available in the literature. Each dataset consisted of a single image that contained both nuclear and membrane fluorescent markers. The analyzed samples included (i) *Drosophila-1*: a sample of 27 intestinal epithelium cells from the adult midgut of *Drosophila melanogaster* immunostained for the cell periphery (bPS integrin) (Fig. 3a, green) and nucleus (DAPI) (Fig. 3a, red) [24], (ii) *Drosophila-2*: a *Drosophila* embryo containing 93 cells imaged with SiMView microscopy expressing H2B-eGFP to label the nuclei (Fig. 3b, red) and lyn-tdTomato under the control of the  $\beta$ -actin promoter to label cell membranes (Fig. 3b, green) [25], (iii) *Tribolium*: an average intensity projection of 36 cells from a uniform blastoderm of *Tribolium castaneum* embryo with H2B-RFP-labelled nuclei (Fig. 3c, red) and GAP43-YFP-labelled membranes (Fig. 3c, green) [26] and (iv) *Arabidopsis*: a confocal micrograph showing the expression of different fluorescent proteins in 105 cells from a stem of *Arabidopsis thaliana* (Fig. 3d) [27]. For details about each dataset, refer to Additional



**Fig. 3** Neighboring cells estimation based on Nfinder. **a** Sample of intestinal epithelium from the adult midgut of *Drosophila melanogaster* immunostained for the cell periphery (green) and nucleus (red) ( $N = 27$ ,  $TPR = 0.902$ ,  $PPV = 0.948$ ,  $JI = 0.859$ ). Adapted from O'Brien and Bilder 2012 [24]. **b** *Drosophila* embryo imaged with SiMView microscopy. Nuclei are labelled with H2B-eGFP (red) and cell membranes are labelled with lyn-tTomato under the control of the  $\beta$ -actin promoter (green) ( $N = 93$ ,  $TPR = 0.988$ ,  $PPV = 1.0$ ,  $JI = 0.988$ ). Adapted from Stegmaier et al. 2016 [25]. **c** Average intensity projection of a uniform blastoderm of *Tribolium castaneum* embryo with H2B-RFP-labelled nuclei and GAP43-YFP-labelled membranes ( $N = 36$ ,  $TPR = 0.952$ ,  $PPV = 0.976$ ,  $JI = 0.93$ ). Adapted from Benton et al. [26]. **d** Confocal micrograph showing the expression of different fluorescent proteins in the stem of *Arabidopsis thaliana* ( $N = 105$ ,  $TPR = 0.927$ ,  $PPV = 0.86$ ,  $JI = 0.806$ ). Adapted from Federici and Haseloff [27]. First column: Original sample. Second column: filtered Delaunay triangulation of nuclei. Third column: performance metrics calculated by comparing the final result to the manually created ground truth. N, Number of cells; TPR, True Positive Rate; PPV, Positive predictive value. JI, Jaccard Index. Scale is not shown for images as the method is scale invariant

file 1: Fig. S1 (*Drosophila-1*), Additional file 1: Fig. S2 (*Drosophila-2*), Additional file 1: Fig. S3 (*Tribolium*) and Additional file 1: Fig. S4 (*Arabidopsis*).

The performance of Nfinder was tested by comparing our results to the manually generated ground truth for the neighboring cells graph of each dataset, as described in the previous section. For evaluation, we used metrics that do not depend on true negatives (TN). Such metrics were avoided because they are highly biased when negatives are dominant [28], which is the case given that there are many more possible links between

cells (negatives) than links between cells that are actually touching (positives). In particular, we analyzed the True Positive Rate (TPR, also called sensitivity) (Eq. 4), Positive Predictive Value (PPV or precision) (Eq. 5) and Jaccard Index (JI) (Eq. 6).

After applying our automatic communicability efficiency filter, the percentage of links removed from the Delaunay triangulation was 20%, 11%, 12% and 13% for the datasets *Drosophila-1*, *Drosophila-2*, *Tribolium* and *Arabidopsis*, respectively. As it is shown in Fig. 3, TPR was over 90% in all cases, supporting the idea that the Delaunay triangulation of nuclei provides a good estimation of neighboring cells. In addition, PPV was 94.6% on average and a maximum of 100% was reached in a *Drosophila* embryo tissue with highly regular spatial distribution (Fig. 3b). This indicates that the more regular the tissue, the better the method's performance. However, its range of application has proven wider as results were satisfactory in irregular tissues, such as the one from the *Arabidopsis* dataset (Fig. 3d). Even in this case, the poorest performance was 80% in JI.

### Nfinder is easily extensible to 3D datasets

As Delaunay triangulations exist in all  $n$ -dimensional spaces, Nfinder can be easily extended to 3D stacks. In contrast to the planar space, the basic structure of the 3D Delaunay network is not a triangle but a tetrahedron. Therefore, before applying the proposed method, each tetrahedron is decomposed into four triangles. Next steps of the pipeline are the same as those described in the previous sections. To show the usefulness of Nfinder in 3D, we estimated the neighboring cells graph of a 3D stack of a *Caenorhabditis elegans* embryo with 24 cells (Fig. 4a and Additional file 2: Video S1), whose nuclei and membranes were stained with mCherry and GFP, respectively [29]. To facilitate the visualization of overlapping cells, we color-coded nuclei according to their position in  $z$ , and highlighted the nucleus with the maximum  $x$ -coordinate and its neighbors.

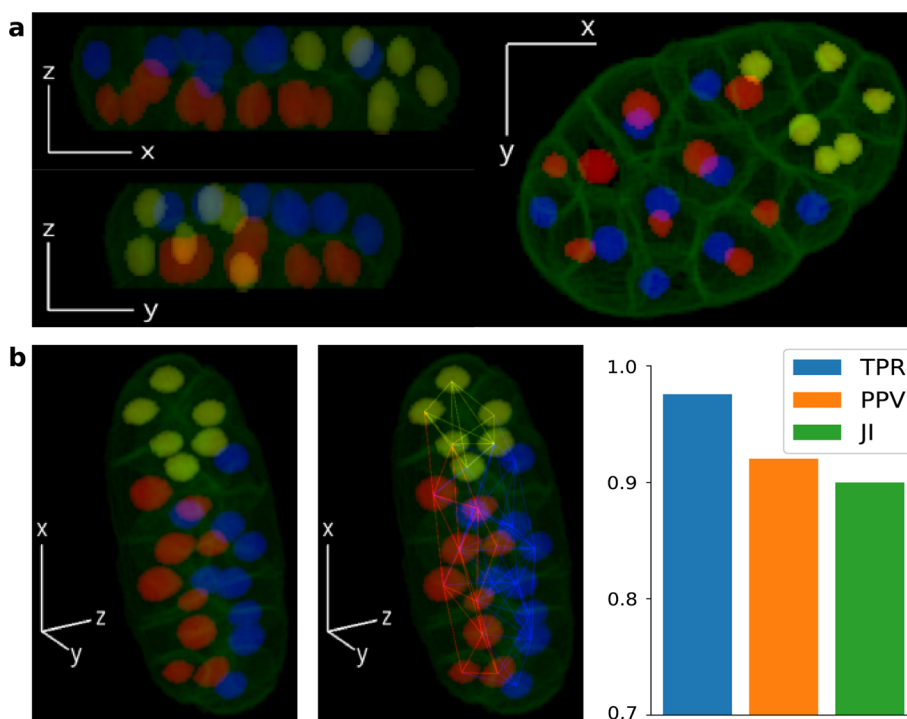
In this case, the percentage of links removed from the Delaunay triangulation by automatic communicability efficiency filtering was 16%. As shown in Fig. 4b, our method was successful in all performance metrics ( $\geq 90\%$ ), with a remarkable TPR of 97.6%.

### Discussion

We developed Nfinder, an automatic and robust method for estimating neighboring relationships between cells using only the information provided by nuclear markers. This is a remarkable feature because membrane staining is not always possible nor convenient in many applications, ranging from cell signalling to tissue development. The robustness of our method is mostly determined by making an educated guess of neighbor relationships between cells via a Delaunay triangulation, which makes results stable against small displacements of cells. In particular, this property is useful in cell tracking, as it is expected that small fluctuations in the positions of cells between timepoints do not affect their connection. Regardless of being designed for nuclear markers, Nfinder can also be applied to cell centroids whenever cytosolic or cytoplasmic fluorescent markers are available.

Although similar strategies have been used before [15, 30, 31], previous works were not focused on the detection of cell neighbors and, therefore, their performance was not thoroughly analyzed. In addition, links between cells were only filtered by a distance threshold, neglecting the importance of non-pairwise interactions [31]. Furthermore,





**Fig. 4** Nfinder results in 3D. **a** Planar views of a 3D stack of a *C. elegans* embryo, showing cell membranes (green) and nuclei ( $N = 24$ ). Colors were chosen so that the embryo was divided by a plane in  $z = 8.83 \mu\text{m}$  (red:  $z < 8.83 \mu\text{m}$ ; blue:  $z \geq 8.83 \mu\text{m}$ ). Detected neighbors of the cell with the maximum x-position are highlighted (yellow). Adapted from Azuma and Onami 2017 [29]. **b** Final results: 3D view of the analyzed specimen (left), estimated cell graph (center) and performance metrics (right; TPR=0.976, PPV=0.92, JI=0.9). N, Number of cells; TPR, True Positive Rate; PPV, Positive Predictive Value; JI, Jaccard Index. Scale is not shown for images as the method is scale invariant

this threshold was an input of the algorithm and, thus, its reusability and generalization were user-dependent. Nfinder is the first method to infer neighboring cells from nuclei labelling that is (i) fully automatic and (ii) includes non-pairwise interactions. The automation of our method was achieved by maximizing a new empirical statistic that assigns a communicability efficiency to each edge of the Delaunay triangulation (Additional file 1: Fig. S5). As for the quantification of non-pairwise interactions, we used the communicability angle of each pair of cells, defined as the maximum angle subtended with shared neighbors.

The proposed automatic filtering of the Delaunay triangulation based on the communicability efficiency of the cell graph improved the performance of the neighboring estimation. To test this, we compared the TPR, PPV and JI obtained for all datasets by selecting edges based on distance thresholding to our method’s results (Additional file 1: Fig. S6). This analysis was performed following the protocol explained in the Methods section. Here, the distance threshold was chosen by maximizing the communicability efficiency modeled as

$$E(r) = F(r) - L(r) \tag{9}$$

where the cumulative distribution of edges  $F$  and the loss function  $L$  only depend on the distance threshold  $r$ . Here,  $L$  was formulated as

$$L(r) = \frac{r - r_{min}}{r_{max} - r_{min}} \quad (10)$$

being  $r_{min}$  and  $r_{max}$  the minimum and maximum edge length in the Delaunay triangulation, respectively. Considering  $r$  as the distance threshold, this loss function represents the fraction of edges that would be filtered out from the Delaunay triangulation if the distance between cells followed a uniform distribution between  $r_{min}$  and  $r_{max}$ . Contrasting the performance metrics of the described distance thresholding with our method's results, we observed that, on average, incorporating the communicability angle increases the JI by +4.2% and PPV by +5.6%, while the TPR did not change significantly (-0.1%). Remarkably, this contribution improved PPV by +11.5% in the dataset *Drosophila-1*. Performance differences seem to be more noticeable when the image contains a low number of cells (*Drosophila-1*) or the spatial distribution of cells is more complex (*C. elegans*). These aspects could be relevant, for instance, when tracking cell–cell interactions during the early stages of embryonic development in 3D.

Importantly, we observed that the performance of Nfinder is affected by how regular the geometry of the cell graph is. This may be due to the correspondence between the Delaunay triangulation and the Voronoi diagram. Since Voronoi cells are convex polygons, neighboring relations between irregular-shaped cells, such as dendritic cells or neurons, might not be well approximated by the Delaunay triangulation of nuclei centroids. A solution to this problem could be replacing these with a more representative set of points, determined by non-convex structures [32]. This analysis is beyond the scope of this work and could be the purpose of future research. However, the validation of our method showed that it is highly accurate and versatile, working with similar performance in diverse tissues both in 2D and 3D.

## Conclusion

In this paper, we developed Nfinder, a precise and automatic method to find neighboring cells from nuclei labelling in 2D and 3D. This automation was carried out by maximizing a novel statistic that measures the communicability efficiency of a spatial graph. Unlike previous works, our method includes cell–cell and non-pairwise interactions, which improves the detection performance significantly. After a systematic evaluation of our method's performance, we found that the importance of considering non-pairwise interactions cannot be neglected if the number of cells is low or the cell spatial distribution is non-uniform. We believe that using our approach might add biologically relevant information to other pipelines, especially those aiming at tracking cell–cell interactions and studying spatiotemporal correlations at the tissue level.

## Abbreviations

CIL	Contact inhibition of locomotion
ECDF	Empirical Cumulative Distribution Function
TP	True positive
FP	False positive
FN	False negative
TN	True negative
TPR	True Positive Rate

PPV	Positive Predictive Value
Jl	Jaccard Index
GFP	Green fluorescent protein

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05284-2>.

**Additional file 1.** Supplementary Figures S1–S6. Additional performance metrics.

**Additional file 2. Video S1:** 3D animation of *C. elegans* embryo showing the estimated neighboring cell graph.

### Acknowledgements

We thank Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and University of Buenos Aires (UBA) for financial support to B.M. and H.E.G.

### Author contributions

The original concept is due to BM and HEG. The idea of using a communicability efficiency filter corresponds to SNRA. All of the authors developed the report, tested the illustrative examples, and read and approved the final manuscript.

### Funding

This work was supported by the following Grants: PICT 2014-3658; PICT 2013-1301 (Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación).

### Availability of data and materials

Image datasets used in this manuscript can be found in references [24–27]. Project name: Nfinder. Project home page: <https://github.com/santi-rodriguez/nfinder>. Operating system(s): Platform independent. Programming language: Python. Other requirements: Python 3.7 or higher, License: MIT License. Any restrictions to use by non-academics: None.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2022 Accepted: 12 April 2023

Published online: 03 June 2023

## References

1. Mayor R, Carmona-Fontaine C. Keeping in touch with contact inhibition of locomotion. *Trends Cell Biol.* 2010;20:319–28.
2. Pavel M, Renna M, Park SJ, Menzies FM, Ricketts T, Füllgrabe J, et al. Contact inhibition controls cell survival and proliferation via YAP/TAZ-autophagy axis. *Nat Commun.* 2018;9:2961.
3. Bray SJ. Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol.* 2006;7:678–89.
4. Krol AJ, Roellig D, Dequeant M-L, Tassy O, Glynn E, Hattem G, et al. Evolutionary plasticity of segmentation clock networks. *Development.* 2011;138:2783–92.
5. Venzin OF, Oates AC. What are you synching about? Emerging complexity of Notch signaling in the segmentation clock. *Dev Biol.* 2020;460:40–54.
6. Oates AC, Mueller C, Ho RK. Cooperative function of deltaC and her7 in anterior segment formation. *Dev Biol.* 2005;280:133–49.
7. van Eeden FJ, Granato M, Schach U, Brand M, Furutani-Seiki M, Haffter P, et al. Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. *Development.* 1996;123:153–64.
8. Jiang YJ, Aerne BL, Smithers L, Haddon C, Ish-Horowitz D, Lewis J. Notch signalling and the synchronization of the somite segmentation clock. *Nature.* 2000;408:475–9.
9. Dale JK, Maroto M, Dequeant M-L, Malapert P, McGrew M, Pourquie O. Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. *Nature.* 2003;421:275–8.
10. Oka C, Nakano T, Wakeham A, de la Pompa JL, Mori C, Sakai T, et al. Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Development.* 1995. <https://doi.org/10.1242/dev.121.10.3291>.
11. Kusumi K, Sun ES, Kerrebrock AW, Bronson RT, Chi DC, Bulotsky MS, et al. The mouse pudgy mutation disrupts Delta homologue Dll3 and initiation of early somite boundaries. *Nat Genet.* 1998;19:274–8.
12. Stanoev A, Mhamane A, Schuermann KC, Grecco HE, Stallaert W, Baumdick M, et al. Interdependence between EGFR and phosphatases spatially established by vesicular dynamics generates a growth factor sensing and responding network. *Cell Syst.* 2018;7:295–309.e11.

13. McQuin C, Goodman A, Chernyshev V, Kamensky L, Cimini BA, Karhohs KW, et al. Cell Profiler 3.0: next-generation image processing for biology. *PLoS Biol.* 2018. <https://doi.org/10.1371/journal.pbio.2005970>.
14. Delaune EA, François P, Shih NP, Amacher SL. Single-Cell-resolution imaging of the impact of notch signaling and mitosis on segmentation clock dynamics. *Dev Cell.* 2012;23:995–1005.
15. Cervino AS, Moretti B, Stuckenholtz C, Grecco HE, Davidson LA, Cecilia CM. Furry is required for cell movements during gastrulation and functionally interacts with NDR1. *Sci Rep.* 2021;11:1–17.
16. Wolff C, Tinevez JY, Pietzsch T, Stamatakis E, Harich B, Guignard L, et al. Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb. *Elife.* 2018;7:1–31.
17. Attardi A, Fulton T, Florescu M, Shah G, Muresan L, Lenz MO, et al. Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. *Development.* 2019;146:175620.
18. Tomer R, Khairy K, Amat F, Keller PJ. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nat Methods.* 2012;9:755–63.
19. Delaunay B. Sur la sphère vide. A la mémoire de Georges Voronoï. *Bulletin de l'Académie des Sciences de l'URSS Classe des sciences mathématiques et na.* 1934;793–800.
20. Musin OR. Properties of the Delaunay triangulation. In: *Proceedings of the thirteenth annual symposium on computational geometry.* New York: Association for Computing Machinery; 1997. pp. 424–426.
21. Reem D. The geometric stability of Voronoi diagrams with respect to small changes of the sites. In: *Proceedings of the twenty-seventh annual symposium on computational geometry.* New York: Association for Computing Machinery; 2011. pp. 254–263.
22. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17:261–72.
23. Estrada E, Hatano N. Communicability Angle and the spatial efficiency of networks. *SIAM Rev.* 2016;58:692–715.
24. O'Brien L, Bilder D. CIL:39686, *Drosophila melanogaster*, intestinal epithelial cell, stem cell. CIL Dataset. 2012. <https://doi.org/10.7295/W9CIL39686>.
25. Stegmaier J, Amat F, Lemon WC, McDole K, Wan Y, Teodoro G, et al. Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Dev Cell.* 2016;36:225–40.
26. Benton MA, Akam M, Pavlopoulos A. Cell and tissue dynamics during *Tribolium* embryogenesis revealed by versatile fluorescence labeling approaches. *Development.* 2013;140:3210–20.
27. Federici F, Haseloff J. CIL:38805, *Arabidopsis thaliana*. CIL. Dataset; 2011. <https://doi.org/10.7295/W9CIL38805>.
28. Luque A, Carrasco A, Martín A, de las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* 2019;91:216–31.
29. Azuma Y, Onami S. Biologically constrained optimization based cell membrane segmentation in *C. elegans* embryos. *BMC Bioinform.* 2017;18:307.
30. Cilla R, Mechery V, Hernandez de Madrid B, Del Signore S, Dotu I, Hatini V. Segmentation and tracking of adherens junctions in 3D for the analysis of epithelial tissue morphogenesis. *PLoS Comput Biol.* 2015;11:e1004124.
31. Javed S, Mahmood A, Fraz MM, Koohbanani NA, Benes K, Tsang Y-W, et al. Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Med Image Anal.* 2020;63:101696.
32. Meyer-Hermann M. Delaunay-object-dynamics: cell mechanics with a 3D kinetic and dynamic weighted Delaunay-triangulation. *Curr Top Dev Biol.* 2008;81:373–99.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

