Collective cell migration is an important process that determines cell reorganization in a number of biological events such as development and regeneration. Random cell reorganization within a confluent monolayer is a popular in vitro model system for understanding the mechanisms that underlie coordination between neighboring cells during collective motion. Here we describe a simple automated C++ algorithm to quantify the width of streams of correlated cells moving within monolayers. Our method is efficient and allows analysis of thousands of cells in under a minute; analysis of large data sets is therefore possible without limitations due to computational time, a common analysis bottleneck. Furthermore, our method allows characterization of the variability in correlated stream widths among a cell monolayer. We quantify stream width in the human retinal epithelial cell line ARPE-19 and the fibroblast cell line BJ, and find that for both cell types, stream widths within the monolayer vary in size significantly with a peak width of 40 μm, corresponding to a width of approximately two cells. Our algorithm provides a novel analytical tool to quantify and analyze correlated cell movement in confluent sheets at a population level and to assess factors that impact coordinated collective cell migration.

Collective cell migration is critical in numerous biological processes such as embryo morphogenesis (1,2), angiogenesis (3), and wound healing (4). It also plays a central role in pathological situations such as metastasis, allowing some tumor cells to invade surrounding tissues (5). However, despite the importance of collective cell migration, how cell coordination occurs during group migration is not well understood (6). A better understanding of the underlying behaviors during collective cell migration will provide insight into embryo development and tissue reorganization during regeneration and disease. Furthermore, the ability to control collective migration will provide novel tools for engineering tissues with reproducible structures.

There are a number of in vivo (7–9) and in vitro model systems (5,10–12) used to study collective cell migration, both directed migration and random migration. Two common in vitro models are confluent sheets of cells (13) with or without a wound intended to induce directed migration. These simple 2-D model systems are powerful because they allow tracking of individual cells within the sheet and careful characterization of the group behavior (14) and cell-cell coordination (15) that arises from cell-cell interactions between neighboring cells. Specifically, cells within the sheet randomly exchange places with their neighbors, resulting in a net displacement from their original location (16) and the formation of randomly oriented cellular “streams” characterized by correlated cell motion (16,17). When a wound is present, the streams at the wound edge become oriented and directed cell movement into the free space occurs to close the wound. Within a cellular stream, during both random and directed migration, motion is coordinated between neighboring cells (i.e., across the width of the stream) and between cells moving behind and ahead of each other (i.e., along the length of the stream). While cell coordination along the stream length could arise passively from the creation of free space behind a cell, coordination across the width of the stream requires more complex cellular interactions. Therefore, understanding the behavior of cells along the width of a stream can potentially provide more insight into the mechanisms of cell-cell coordination during sheet reorganization. Furthermore, quantifying stream width is potentially useful for designing culture surfaces (18) with adhesive or topographic features (unpublished data) that allow manipulation of this streaming behavior, in which feature dimensions must be on the scale of the stream width (19).

A common method for quantifying cell coordination is the velocity correlation function (16,20–22), which evaluates the overall correlation of cell trajectories as a function of distance from a cell using the Equation 1:

$$V'(r) = \frac{1}{N} \sum_{n=0}^{n=N} \sum_{j=0}^{j=N} \frac{1}{n} \ddot{\theta}(\dot{r})$$

[Eq. 1]

where \(N\) refers to the total number of objects being analyzed, \(n\) is the number of neighboring objects within a radius, \(r\) from the current object \(i\), and \(j\) refers to all other objects within radius \(r\). By averaging all dot product values for all cell pairs, the function provides an overall description of neighboring cell correlation as a function of distance from a cell. When the velocity correlation function falls to \(\sim 0\) (or \(\theta_{average} = \pi/2\)), cells are no longer considered correlated. The radius at which this occurs provides a measurement of
were grown in DMEM (ATCC) supplemented with 10% FBS and 1% pen/strep. All cells were maintained in a humidified atmosphere at 37°C and 5% CO₂. For some experiments we overexpressed GFP-N-cadherin in ARPE19 cells using a lentivirus overexpression vector, followed by sorting of the cells by flow cytometry to ensure a homogenous population.

**Materials and methods**

**Cell culture**

We conducted experiments using human retinal epithelial cells (ARPE-19; ATCC, Manassas, VA, USA) and human foreskin fibroblast cells (BJ; ATCC). ARPE-19 cells were grown in Dulbecco’s modified Eagle's medium/nutrient F-12 (DMEM/F-12; Invitrogen, Burlington, ON, Canada) supplemented with 10% fetal bovine serum (FBS; VWR International, Mississauga, ON, Canada) and 1% penicillin/streptomycin (pen/strep; VWR International). BJ cells were grown in DMEM (ATCC) supplemented with 10% FBS and 1% pen/strep. All cells were maintained in a humidified atmosphere at 37°C and 5% CO₂. For some experiments we overexpressed GFP-N-cadherin in ARPE19 cells using a lentivirus overexpression vector, followed by sorting of the cells by flow cytometry to ensure a homogenous population.

**Live cell imaging and tracking**

We seeded 64,000 cells/cm² in a custom-made 96-well polydimethylsiloxane (PDMS) flat bottom plate to generate confluent cell sheets (Figure 1A). For wound healing studies, 90,000 cells/cm² were seeded in a 24-well tissue culture polystyrene (TCPs) plate coated with 314 µg/mL PureCol (Advanced Biomatrix, San Diego, CA, USA) for 1 h prior to seeding and wounds were made using a P10 tip. We tracked cells a day later using a previously established protocol (15,23). Briefly, cells were stained with 500 ng/mL Hoechst 33342 (Invitrogen) in culture medium for 30 min, followed by a phosphate-buffered saline (PBS) wash and addition of culture medium. An ImageXpress Micro high-content screening microscope with a live-cell imaging module (Molecular Devices, Sunnyvale, CA, USA) was used to take images of the cell nuclei every 20 min for 8 h for tracking in confluent sheets (Supplementary Figure S1 and Supplementary Movie S1) and for 23 h for wound healing (Supplementary Figure S2 and Supplementary Movie S2). Throughout the experiment, cells were maintained in a humidified atmosphere at 37°C and 5% CO₂. Positional tracking of cells in each well was performed using the Multi-Dimensional Motion Analysis application module in the MetaXpress software package (Molecular Devices). This tracking algorithm uses nuclear shape and fluorescent intensity measurements to track the position of individual cells (Supplementary Figures S1 and S2) (24,25).

**Stream visualization and quantification**

To assess the width of cellular streams, we imported the text files generated by the MetaXpress tracking algorithm into our analysis program. The text files contain information for individual cells over a series of time frames including the cells’ assigned object numbers, their positions and radii from a designated origin, and the angles that they are moving relative to the previous time frame. These variables are specifically used in the correlation algorithm to evaluate the correlated stream widths. While we used text files generated by MetaXpress, any method of cell tracking that gathers this coordinate data, including manual tracking, could be used with our algorithm with minor modifications to the code. For comparison, we also calculated the velocity correlation function using Equation 1 (16,20–22). The resulting average of the velocity correlation function falls between zero and one (zero means no correlation between cell streams, and one means that all streams are moving in the same direction) (16,18,26).

The MetaXpress tracking algorithm also outputs an image file that allows the visualization of the moving streams (Figure 1B), but the color coding is random. To better visualize the coordination between cells in the stream, we wrote a Matlab (The MathWorks, Natick, MA, USA) program that specifically colors the moving cell streams based on the direction of motion from the text files (Figure 1, C and D). Using these latter images, we estimated the accuracy of our algorithm by comparing reported stream widths for specific cells with manual measurements for those cells from the angle-based colored images.

**Algorithm design**

An overview of the algorithm is shown in Figure 2. To evaluate the widths of the moving cell streams, the algorithm compares the angle of displacement for a given cell (the comparison cell) at a given time point with the angles of displacements of certain surrounding cells at that same time point. Specifically, the algorithm compares the angle of displacement of surrounding cells located within an 140° arc perpendicular (above and below) to the angle of displacement of the comparison cell (Figure 2A). We did this to avoid overestimating stream width due to correlated movement along the stream length (6). For each surrounding cell, the algorithm designates the cell as correlated with the comparison cell if it is moving within 10° relative to the direction of the comparison cell. This threshold angle can be changed if desired. Only objects in the same time frame are compared, because objects moving in the same direction as the comparison
The algorithm assesses the correlation of surrounding cells at progressively increasing distances from the comparison cell and defines stream width as the distance at which correlation is lost. Specifically, the algorithm examines the cells in the defined arcs that are within a ring of 20 µm thickness (r_{in} to r_{out} in Figure 2) from the comparison cell, starting at an inner radius of 0 µm (Figure 2A). This ring thickness was chosen because ARPE-19 and BJ cells have widths that are within a 20 to 30 µm range (25.4 ± 6.4 µm and 25.7 ± 7.4 µm, respectively); this number should be altered if larger or smaller cell types are being assessed. Each successive ring is considered to be correlated if the number of correlated cells is greater or equal to the number of uncorrelated cells in either one or both arcs of the ring. If the algorithm detects the cells in a particular radius are correlated, the algorithm expands the inner radius by a further 20 µm (Figure 2B) and compares the number of uncorrelated and correlated cells in arcs in the new ring region. This comparison operation is continued until a ring is reached in which the number of uncorrelated cells is greater than the number of correlated cells (Figure 2B) in both 140° arcs regions of that ring. The inner radius of this ring is defined as the stream width for that object (comparison cell) for that time frame. Since each 140° arc is defined separately, if one or both of the arcs do not contain any cells in a given time frame, the inner radius is increased until new objects are found. If objects are not found within 100 µm, the arc is declared done, and the inner radius where cells were last previously detected is logged as the stream width. This component of the algorithm is important in situations where cells are more spread out, at lower cell densities, at the boundaries of the image, or if a wound is present. To account for situations where the comparison cell being analyzed is in the center of a stream, the radius is doubled if both arcs finish at the same time. For a given comparison cell, the stream widths are analyzed for all of the time frames, and the average radius is calculated and recorded as the stream width. This is then repeated for all cells in the image. The algorithm therefore outputs a distribution of stream widths (one for each cell in the image).

Our algorithm assesses the average stream width for a given cell at only selected time frames (frames 1, 4, 8, 12, 16, 20, and 24). This decision was made because examining every frame resulted in odd computations due to any random “wiggle” effects from non-straight movement of the cells. By eliminating frames, the cell movement was approximated to be more linear (see increase in straightness of path in skipped frame image compared with nonskipped frame image in Figure 1, C, and D). Also, the 140° arcs are subdivided into two halves (four sections total). This prevents the algorithm from terminating early in cases where rings contain neighboring streams moving in different orientation. This was not an issue at small assessment radii but became a problem after the size of the arcs was increased to find more objects in the same frame.

Results and discussion

As expected, our live cell imaging demonstrated that both cell types we tested reorganize within the confluent cell sheet. Our Matlab recoloring tool, as has been seen previously (15), allows cells moving in the same direction to be more easily visualized (Supplementary Figure S3). Using our algorithm, we characterized the cell streams for one epithelial (ARPE-19) and one fibroblast (BJ) cell line, as shown in histograms in Figure 3, A and B. We selected these cell types because we expected them to show differing migratory behavior due to differences in the extent to which each cell type interacts with neighboring cells within the sheet: epithelial cells form stronger and more junctions between neighboring cells than fibroblasts. Surprisingly, both cell types showed similar behavior: a large number of the cells did not participate in streams, as evidenced by the peak at <20 µm in Figure 3, A and B, while for cells moving within streams, we observed a range of stream widths, with a peak at 40 µm, corresponding to a stream width of approximately two cells. Our analysis highlights the non-normal variation in stream widths for different cells within the sheet, suggesting a population average measurement is not necessarily an accurate representation of the behavior of all cells within the sheet.

We also quantified the size of the streams using the standard velocity correlation function method (shown in Figure 3, C and D). Figure 3, C and D, show data averaged for all cells and suggest that the radii at which cells become noncorrelated (in which the inverse cosine of the function levels off at ~π/2) are ~200 and ~250 µm for ARPE-19 and BJ cells, respectively. To allow easier comparison with our nonaveraged data, we also quantified for each cell the radius at which the velocity correlation function approached 0 (specifically when it fell below 0.273 or φ = π/17°). This was selected to account for a nonperfect drop to exactly 0 in the function when cells become noncorrelated. Figure 3, E and F, show the distribution of “stream widths” defined using this method. The distribution of stream widths calculated using the velocity correlation function method was wider than that found using our algorithm, and peaked at ~100 µm for both cell types as seen in Figure 3, E and F, as opposed to the peak of stream width around 40 µm predicted by our algorithm as seen in Figure 3, A and B. We attribute this difference to the impact of correlation along the stream length and the error this introduces depending on where a cell lies within the stream. The value returned by the velocity correlation function is a combined measure of both the length and width of the stream. Our algorithm, on the other hand, provides a novel method to quantify stream width specifically.

We also used our algorithm to quantify stream width in ARPE-19 cell sheets in the presence of a wound (Supplementary Figure S2 and Supplementary Movie S2) and with cells expressing GFP-N-cadherin. Figure 3, G and H, shows the stream width distribution for each situation. In the wound healing situation, the number of cells not participating in streams decreases, and an increase in the number of cells participating in streams of 40 µm were observed. These observations are consistent with the initiation of directed collective migration on introduction of the wound. Peak stream width does not appear to significantly increase in contrast to what the track images suggest. This illusion is due to the directed motion of the cell in different streams following similar paths, but at different time points. Interestingly, in cells expressing GFP-N-cadherin, where we expected increased cell-cell interactions, the distribution of stream widths shifted to lower values. This observation could be due to reduced cell motion in response to increased adhesiveness, and further studies are currently underway to probe this effect and identify the key cellular parameters that determine stream width within a sheet of cells. Beyond the wound and confluent sheet examples shown here, we believe our algorithm will also be useful in the future for defining the impact on stream width of cells moving within confined spaces (19).
Our algorithm was written efficiently in C++ to allow researchers to run data sets of any size (varying number of time frames) in minutes. The algorithm variables can be easily fine-tuned to quantify collective migration behavior for any cell type. For example, it is easy to adjust parameters such as the number of frames to be analyzed (based on how tortuous the cell migration path is) and the size of radii increments (based on average cell width). The program is fully automated, allowing analysis of the data for all of the wells from a 96-well plate with a single instruction, which is useful for integration with a high-content screening platform. On average, the program can run data containing ~2400 cells (one site of a confluent well imaged at 4× magnification) over 24 time frames in ~20 s. Vast numbers of cells can therefore be analyzed in a short time period, which is important as often computational time is an analysis bottleneck.

We assessed the accuracy of our algorithm in estimating stream width by comparing its output to manual measures of stream width from our directionally colored images for randomly selected cells. Specifically, the stream widths of thirty randomly selected cells from three wells were evaluated manually for each cell type. Accuracy was computed as

\[ \text{Accuracy} = \frac{\text{manual measurement} - \text{algorithm measurement}}{\text{algorithm measurement}} \times 100 \]

The average accuracy is displayed on the bar chart in Figure 4A and was not significantly different for each test case. The algorithm had a calculated accuracy of 89.7% ± 8.9% for ARPE-19 cells, of 85.1% ± 10.1% for BJ cells, 75.5% ± 12.5% for ARPE-19 GFP-N-cadherin expressing cells, and 82.2% ± 10.5% for the ARPE-19 cells in the wound healing assay. The inaccuracy was a result of a combination of factors. First, manually estimating the stream widths by hand was incredibly difficult due to the number of surrounding tracks, making our accuracy benchmark inherently inaccurate. This was a major contribution to the large deviation in accuracy, and further illustrates the need for an algorithm to evaluate stream widths. Secondly, errors may arise from cells that are not exactly in the center or edge of a stream. In the algorithm used for all our experiments, to account for situations where the comparison cell being analyzed is exactly in the center of a stream, the radius is doubled if both arcs finish at the same time; this was computationally easy to implement and was expected to improve the accuracy of the measurement. A similar calculation for cells located between the edge and center of a stream would have unnecessarily increased the complexity of the code and hence was not implemented; therefore, these cells will report narrower stream widths. We do not expect this effect to be significantly detrimental to overall accuracy, however, unless assessing very wide streams. Still, this could be improved in future iterations of the algorithm. Thirdly, the algorithm may be too lenient when classifying neighboring cells as correlated. If a nearby cell moved in the same orientation for less than ~5 µm, our observations determined the cell movement to be uncorrelated, whereas the program marked it as correlated, for that time point. This could be adjusted in the future by only considering data from pairs of cells that both move...
should be selected to correspond approximately to the width of one cell to get the most accurate results. We also assessed the impact of the angle used to define correlated movement (Figure 4C). We selected ±10° for correlation based on visual estimates of the angle between cells in a stream from our live imaging movies, but this could easily be adjusted to be more or less stringent. Decreasing this angle to 5° did not significantly change the shape of the stream width distribution or peak location, but did slightly increase the number of cells considered to be part of very narrow streams. Increasing this angle to 50°, however, significantly impacted the stream width distribution. Selection of an appropriate angle to define correlated movement, therefore, is also an important parameter for obtaining accurate results.

Acknowledgments
The authors thank Maria Jimena Loureiro and Sahar Javaherian for technical contributions. This work was funded by a National Science and Engineering Research Council (NSERC) Discovery grant (to A.M.).

Competing interests
The authors declare no competing interests.

References


Received 6 November 2012; accepted 17 January 2013.

Address correspondence to Alison P McGuigan, Dept of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College St., Toronto, Ontario, M5S 3J9, Canada. e-mail: alison.mcgugan@utoronto.ca

To purchase reprints of this article, contact: biotechniques@fosterprinting.com