1
2
Cell Geometry Distinguishes Migration-Associated Heterogeneity in Two-Dimensional Systems

3

4   **Running Title:** Geometric heterogeneity of cell migration

5

6   Sagar S Varankar[1,2,3], Kishore Hari[1], Sharmila A Bapat[2,*], Mohit Kumar Jolly[1,*]

7   *Corresponding Authors: mkjolly@iisc.ac.in, sabapat@nccs.res.in

8

9   [1]Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore
10   560012, India.

11   [2]National Centre for Cell Science, Savitribai Phule Pune University, Ganeshkhind, Pune
12   411007, India.

13   [3]Present Affiliation: Wellcome – MRC Cambridge Stem Cell Institute, Cambridge, UK.

14

Abstract:

Background: In vitro migration assays are a cornerstone of cell biology and have found extensive utility in research. Over the past decade, several variations of the two-dimensional (2D) migration assay have improved our understanding of this fundamental process. However, the ability of these approaches to capture the functional heterogeneity during migration and their accessibility to inexperienced users has been limited.

Methods: We downloaded published time-lapse 2D cell migration datasets and subjected them to feature extraction with the Fiji software. We used the 'Analyze Particles' tool to extract ten cell geometry features (CGFs), which were grouped into 'shape', 'size' and 'position' descriptors. Next, we defined the migratory status of cells using the 'MTrack2' plugin. All data obtained from Fiji were further subjected to rigorous statistical analysis with R version 4.0.2.

Results: We observed consistent associative trends between size and shape descriptors and validated the robustness of our observations across four independent datasets. We used these descriptors to resolve the functional heterogeneity during migration by identifying and characterizing 'non-migrators (NM)' and 'migrators (M)'. Statistical analysis allowed us to identify considerable heterogeneity in the NM subset, that has not been previously reported. Interestingly, differences in 2D-packing appeared to affect CGF trends and heterogeneity of the migratory subsets for the datasets under investigation.

Conclusion: We developed an analytical pipeline using open source tools, to identify and morphologically characterize functional migratory subsets from label-free, time-lapse migration data. Our quantitative approach identified a previously unappreciated heterogeneity of non-migratory cells and predicted the influence of 2D-packing on migration.

Introduction

Migration is a fundamental cellular function which contributes to development, tissue maintenance and disease progression across biological systems. Initiated during early embryonic morphogenesis, migration involves dynamic changes to the cellular architecture and cell-niche interactions[1–3]. Apart from well-defined molecular signatures, phenotypic parameters are often used to identify migrating cells albeit with limited accuracy during homeostasis. Changes in cell phenotype or, from a quantitative perspective, geometry are most evident during processes like the epithelial-mesenchymal transition (EMT) or immune cell homing which involve shifts between distinct cellular states[4]. Dependent grossly on tissue architecture and spatial arrangement, cells can adopt a defined repertoire of geometric configurations which in turn influence their patterns of motility[5–8]. While the physiological relevance of distinct migratory modalities has been widely studied *in vitro*, the functional heterogeneity of migration in the context of cell geometry remains poorly understood.

The *in vitro* two-dimensional (2D) scratch assay is routinely used to quantify and examine the effect of exogenous cues on cell migration[9]. In recent years, several alterations to its experimental setup in conjunction with quantitative tools have provided insight into the biophysical properties of migrating cells[10–15]. In a previous study, we coupled time lapse microscopy with the 2D scratch assay and defined three distinct migratory modes in ovarian cancer cell lines, *viz.,* passive collective cell migration (pCCM), active collective cell migration (aCCM) and EMT. We further demonstrated the switching of migration modalities in response to extrinsic cues for these cell lines and observed that cells which undertook aCCM exhibited higher chemoresistance and a greater inherent plasticity[16]. A similar study with breast cancer cell lines associated monolayer migratory patterns with specific mutational signatures, thus providing a clinically relevant readout for metastatic potential[17]. Independent reports have further discerned the leader-follower dynamics, cell polarity and orientation, and directionality of collective cell migration with similar experimental approaches[18–20]. Cell tracking and particle image velocimetry (PIV) approaches have also been adopted in tissue-derived scaffolds[21] and developing embryos[22] to dissect niche dependent migratory dynamics. In addition to the use of time lapse microscopy, the recent development of high-throughput analytical tools has allowed efficient handling of large migration datasets[13]. Microfluidic devices in conjunction with machine learning have been used to discern the shape-shifting dynamics of single cells[23]. However, despite their utility, most of these platforms require high-end computing and are often inaccessible to users with limited coding experience.

75 We developed an open source quantitative approach, utilizing cell geometry features
76 (CGFs), for label-free time lapse microscopy data. Our analysis allowed the identification
77 and characterization of migratory subsets within otherwise homogenous cell systems.
78 Interestingly, CGF trends and functional heterogeneity observed in our study extended
79 across multiple experimental models of cell migration. These results allowed us to speculate
80 the effects of 2D-packing, growth factor treatment and ECM coating on migratory subsets,
81 and emphasized the widespread utility of our pipeline. Our approach to conduct image
82 analysis with Fiji and the availability of well-annotated codes would allow experimental
83 biologists to efficiently integrate this pipeline for analysing cell migration data.

84

85 **Materials and Methods:**

86 *Imaging Datasets, Processing and Quantification of imaging data:*

87 Label-free time-lapse imaging data acquired with phase-contrast microscopy were
88 downloaded from published studies[16,24–26] and processed with Fiji software
89 (https://imagej.net/Fiji/Downloads) as previously described[12]. Briefly, images from individual
90 datasets were converted into an 8-bit format, adjusted for contrast and threshold, and
91 analysed with the 'Analyze Particles' tool and 'MTrack2' plugin in Fiji. CGFs were extracted
92 across the time course for each experimental dataset. Migratory and non-migratory cells
93 were identified by comparing and merging cell position co-ordinates from the 'Analyze
94 Particle' and 'MTrack2' outputs in R version 4.0.2. Codes used for data wrangling have been
95 uploaded to Github and can be found at https://github.com/askhari139/CGF.

96

97 *Cumulative Correlation Scores:*

98 Cumulative correlation scores were calculated as follows:

99 $$Correlation\ Score\ Cell\ Size = \sum rCGF_{size}$$

100 $$Correlation\ Score\ Cell\ Shape = \sum rCGF_{shape}$$

101 where, $rCGF_{size}$ = Pearsons' Correlation coefficient between two Size CGFs

102 $rCGF_{shape}$ = Pearsons' Correlation coefficient between two Shape CGFs

103 Only 'r' values with a $p < 0.05$ were used for this calculation

104

105 *Statistical analysis:*

106 Statistical analyses were performed with Prism software package version 6.0 (GraphPad).

107 *P* values were calculated using two-tailed paired Student's t test and considered significant if

108 *P*<0.05. Pearson's Correlation Co-efficient and Principal Component Analysis were

109 calculated in R version 4.0.2 using the functions *cor.test* and *prcomp* from the Base package

110 respectively.

111

112 **Results:**

113 *In vitro 2-D cell migration exhibits conserved geometric features across cell line models*

114 To investigate the differences in cell geometry during 2D-migration, we first defined

115 quantitative morphological features which could be consistently extracted from *in vitro*

116 imaging data. Due to its open-source availability, multi-operating system compatibility and

117 user-friendly interface, the Fiji software was used in this study. Cell geometry features

118 (CGFs) identified with the 'Analyze Particles' tool in Fiji were broadly categorized into size

119 (Area, Perimeter, Width Height, Feret Diameter), shape (Circularity, Solidity, Roundness,

120 Aspect Ratio) and position descriptors (X/Y co-ordinates, Feret Angle) as detailed in **Figure**

121 **1**. Prior to examining trends during migration, we evaluated the statistical relationships

122 between these CGFs using a migration dataset for high grade serous ovarian cancer

123 (HGSC) cell lines[16]. The HGSC data comprised of six cell lines A4, OVMZ6, OVCA420,

124 PEO14, OVCAR3, CAOV3 which were exposed to four distinct culture conditions during the

125 *in vitro* wound healing assay *viz.,* in the presence of 5% serum, serum starvation, exposure

126 to chemotherapeutic drug paclitaxel (IC_{50} concentration as previously defined for individual

127 cell lines[16]) and 0.1% DMSO as a drug-associated vehicle control. Quantification of wound

128 closure over the course of the time-lapse experiment (16 hours) reiterated previous

129 observations of efficacy in the following order A4> OVMZ6 > OVCA420 > PEO14 > OVCAR3

130 > CAOV3 (**Figure 2a**).

131
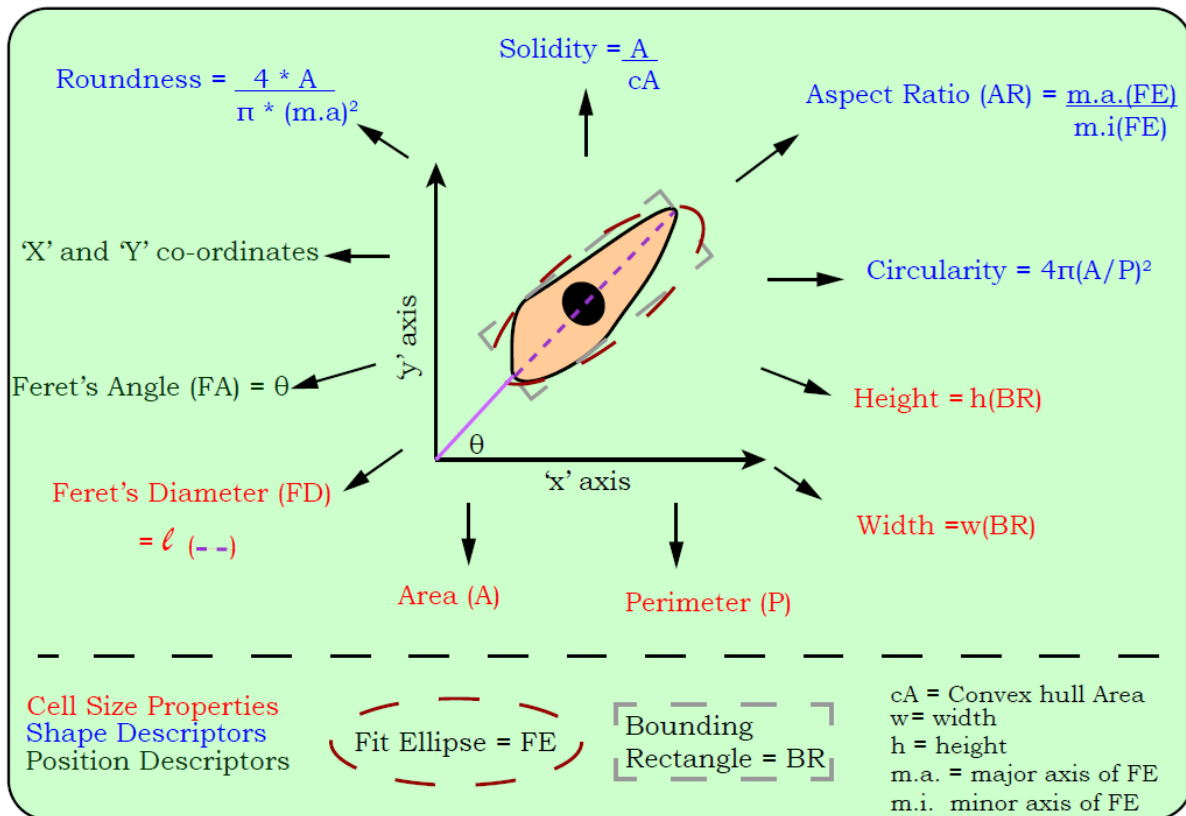
Figure 1. Cell Geometry Features. Schematic depiction of cell geometry features (CGFs) quantified for individual cells in migration datasets. CGFs were broadly classified into three categories *viz.,* cell size, cell shape, and cell position. 'Area', 'Perimeter' and 'Position Co-ordinates' for each cell were directly quantified while other CGFs were derived by an artificial overlay of 'fit ellipse' and 'bounding rectangle' by the Fiji software. Formulae employed for quantifying CGFs by Fiji are outlined in the schematic.

132

133   We previously associated the phenotypic properties of each HGSC cell line with their
134   migration capacity[16]; however, the dynamic changes in cell geometry were beyond the scope
135   of the study. To obtain a preliminary understanding of their statistical inter-dependencies,
136   pairwise correlation co-efficient between the ten CGFs were quantified without temporal
137   segregation of data points. This revealed a significant positive correlation among cell size
138   descriptors and independently, between the two cell shape descriptors 'circularity' and
139   'solidity'. Further, these two CGF subsets (shape and size) negatively correlated with each
140   other, but demonstrated no significant correlations with aspect ratio (AR), roundness and
141   ferret angle (FA). While roundness and AR were negatively correlated, none of the CGFs
142   correlated significantly with FA suggesting that CGF associations may be independent of cell
143   orientation in a migrating monolayer (**Supplemental Figure 1**). To consolidate CGF
144   correlative trends across multiple samples, we developed a cumulative correlation score
145   (CCS) for cell size and shape descriptors separately (*Outlined in Materials and Methods*); FA

146 was excluded from the analysis due to aforementioned observations. Distribution based on

147 CCSs generated a tight cluster of samples with few outliers, despite the distinct phenotypic

148 differences between cell lines (**Figure 2b**). To examine whether these distributions were

149 affected over time during cell migration, we quantified the CCS across 30-minute intervals

150 for each cell line-treatment pair and obtained comparable outcomes to what was seen in the

151 earlier analysis (**Supplemental Figure 2**). These observations were conserved across the

152 entire dataset, irrespective of the cell type or treatment, thus demonstrating the time- and

153 position-independent nature of associations among the CGFs, which provided a robust

154 framework for further analysis.

155
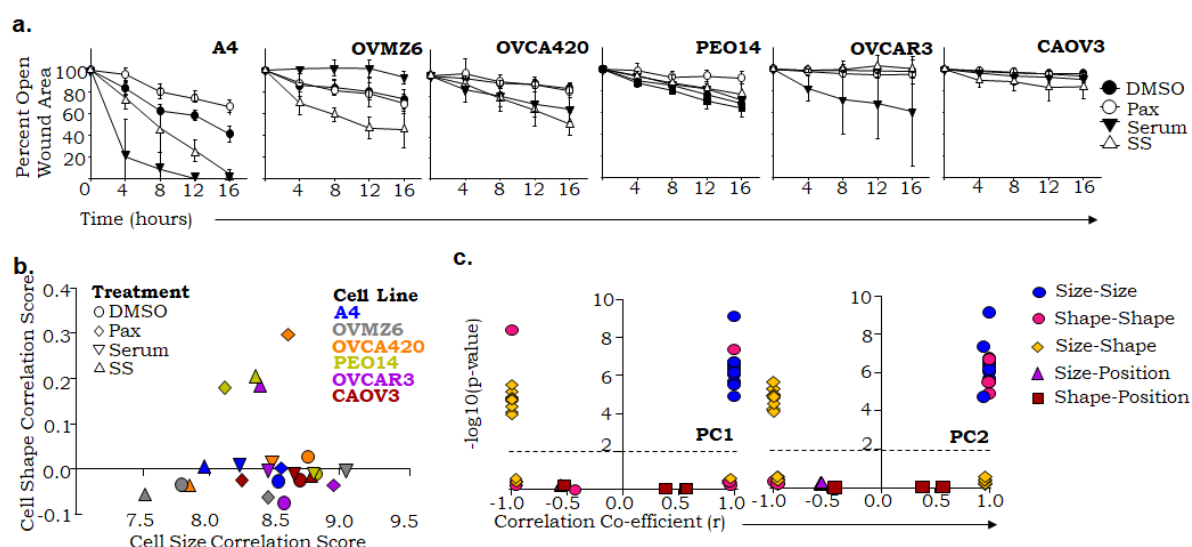
Figure 2. Conserved statistical relationships between cell geometry features are characteristic of 2D-cell migration. a. Percent open wound area quantified for a panel of high grade serous ovarian cancer (HGSC) cell lines in response to distinct assay conditions: DMSO, Paclitaxel (Pax), Serum, Serum Starved (SS). **b.** Cumulative correlation score-based (*detailed in Materials and Methods*) distribution of samples. **c.** Correlation coefficients computed for CGF-specific Eigen vectors for PC1 and PC2. Correlation coefficients were broadly calculated for five types of CGF pairs: (i) Size-size, (ii) shape-shape, (iii) size-shape, (iv) size-position and (v) shape-position. Except for panel (a), the data has not been segregated temporally and includes CGFs for individual cells from the HGSC migration.

156

157 Next, we examined the contribution of individual CGFs to the variance and heterogeneity of

158 cell migration. Principal component analysis (PCA) on temporally unsegregated CGF data

159 reiterated the non-significant contribution of FA, hence cell orientation, to geometry

160 (**Supplemental Figure 3a**). Principal components 1 and 2 (PC1 and PC2) captured

161 approximately 65% of the total variance with the size descriptors, circularity, solidity and AR,

162 roundness contributing to maximum variance in PC1 and PC2, respectively (**Supplemental**

163 **Figures 3b, 3c**). Contribution of individual CGF pairs to PC axes was reminiscent of

164 previous correlative trends wherein FA exhibited no quantitative significance (**Figure 2c**).

165   These observations resulted in the exclusion of FA from subsequent analysis and
166   established the fundamental nature of CGF associations.

167

168   *Cell shape heterogeneity represents the plasticity of migratory cells*

169   While we observed a consistent pattern of CGF associations, it was unclear whether these
170   resulted from an averaging of heterogeneous sub-populations or the absence of phenotypic
171   heterogeneity within cell lines. To examine the first hypothesis, individual cells across the
172   HGSC migration dataset were assigned a motility status *viz.,* non-migratory (NM) and
173   migratory (M), based on the minimal distance covered by them during the experiment. The
174   proportion of 'NM' and 'M' cells across the dataset exhibited cell line and treatment specific
175   differences, with a majority of the cells being assigned the 'NM' phenotype (**Figure 3a**).
176   Correlative trends quantified for 'NM' and 'M' cells were strikingly similar to those observed
177   in the whole population data, thus affirming the absence of normalization effects
178   (**Supplemental Figure 4, Supplemental Table 1**). Next, we tested whether the conserved
179   CGF correlations emerged from the lack of morphological differences between the 'NM' and
180   'M' subsets. Examination of mean CGF values revealed that 'M' cells were larger in size and
181   less circular than the 'NM' subset (**Figure 3b, Supplemental Figures 5 and 6**). While
182   observations pertaining to circularity of 'M' cells were unsurprising, it was interesting to note
183   the relatively small cell sizes of the 'NM' population.

184   To further examine the heterogeneity of our cell systems, we evaluated the co-efficient of
185   variance (CoV) for individual CGFs in the 'NM' and 'M' subsets. 'NMs' displayed a previously
186   unappreciated heterogeneity in cell size which may be an outcome of differential proliferation
187   and/or migratory-cell induced packing. On the other hand, the 'M' subset demonstrated
188   greater heterogeneity of cell shape further reflecting its morphological plasticity (**Figure 3c,**
189   **Supplemental Table 2**). The observed differences in CGF trends between 'NM' and 'M'
190   cells thus, confirm the universal conservation of correlative trends and identify greater
191   heterogeneity in functional subsets during 2-D cell migration.

192
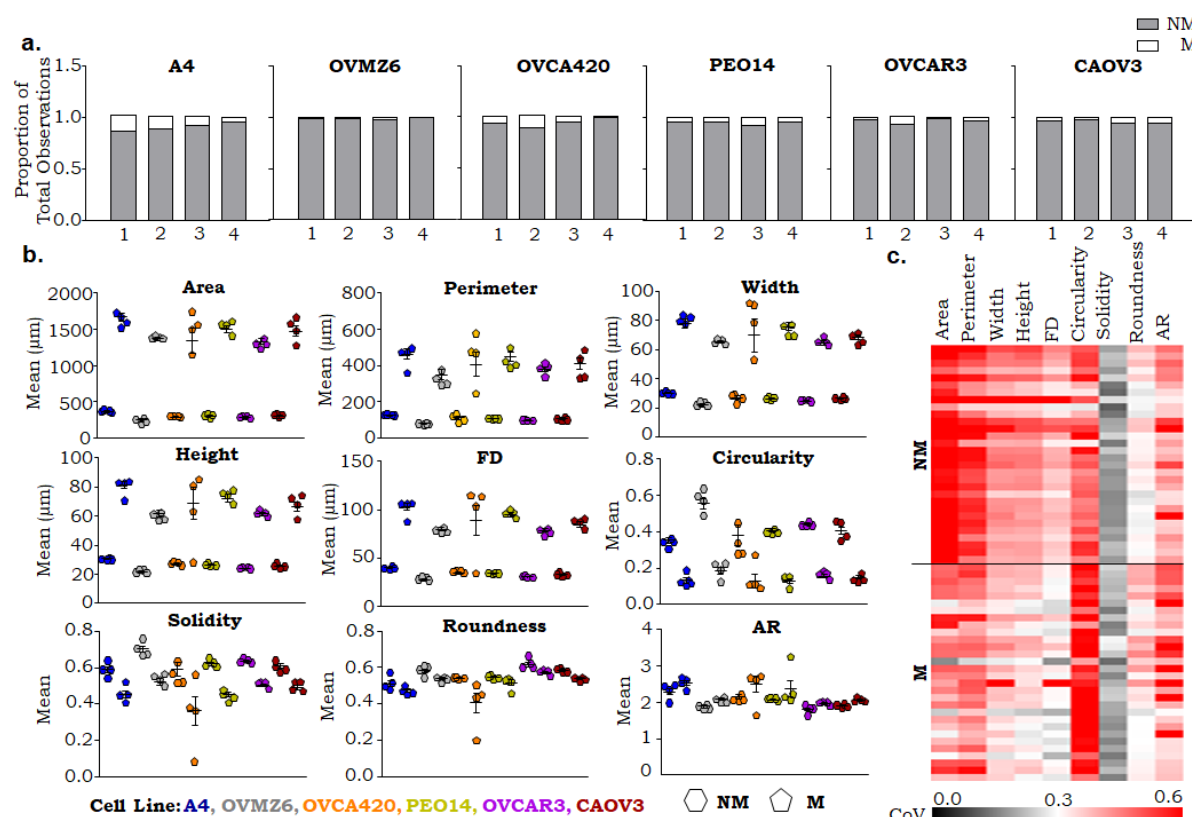
Figure 3. Consistent geometric trends identify heterogeneity in cell size for non-migrators and cell shape for migrators. a. Proportion of non-migratory (NM) and migratory (M) cells across the HGSC migration dataset. 1= DMSO, 2= Paclitaxel, 3= Serum, 4= SS. **b.** Mean distribution of CGFs between NM and M cells depicted for individual cell lines. Each data point depicts a specific treatment for individual cell lines. **c.** Heatmap depicting the co-efficient of variance for individual CGFs between NM and M cells. Data are not temporally segregated.

193

194   *CGF features are conserved in an independent system of 2D monolayer migration*

195   Robust statistical CGF trends prompted us to investigate whether similar observations could

196   be extrapolated to an independent dataset. Literature search for articles with *in vitro* time-

197   lapse migration data, provided a large cohort. Studies were selected based on the

198   availability of good quality imaging data, processability with Fiji, the presence of appropriate

199   controls and existence of relevant technical replicates. To examine CGF trends during *in*

200   *vitro* wound healing, we extracted data from the Zaritsky *et al.* study wherein the wound

201   healing efficacy of a mouse mammary adenocarinoma cell line (DA3) was examined in

202   response to the growth factor HGF-SF. Differences in the cell system and, experimental

203   setup permitted an unbiased assessment of CGF associations[25]. Data were not temporally

204   segregated based on previous observations and analysed between control and HGF-SF

205   treated samples (**Supplemental Figure 2**). Assignment of motility status identified a large

206   population of 'M' cells in the control samples which were significantly enriched in response to

HGF-SF (**Figure 4a**). A similar trend of large, non-circular 'M' cells as compared to the 'NM' subset was observed in both, control and HGF-SF treated samples (**Figure 4b**). Interestingly, increase in cell size for both, 'NM' and 'M' subsets along with reduced circularity for the non-migrators was evident following HGF-SF treatment. Changes in circularity can be attributed to an 'NM' to 'M' transition which explains the enhanced wound healing capacities in response to HGF-SF[25]. Quantifying CoV for DA3 cells revealed greater heterogeneity in 'NM' size as compared to the 'M' subset (**Figure 4c**). These observations can be explained by the existence of 'M' cells which have achieved homogeneity through extensive morphological changes as opposed to 'NMs' which are in the process of acquiring migratory properties. Thus, the Zaritsky, *et al.*, dataset validated our CGF trends and provided an *in vitro* model for examining 'NM' to 'M' transition.
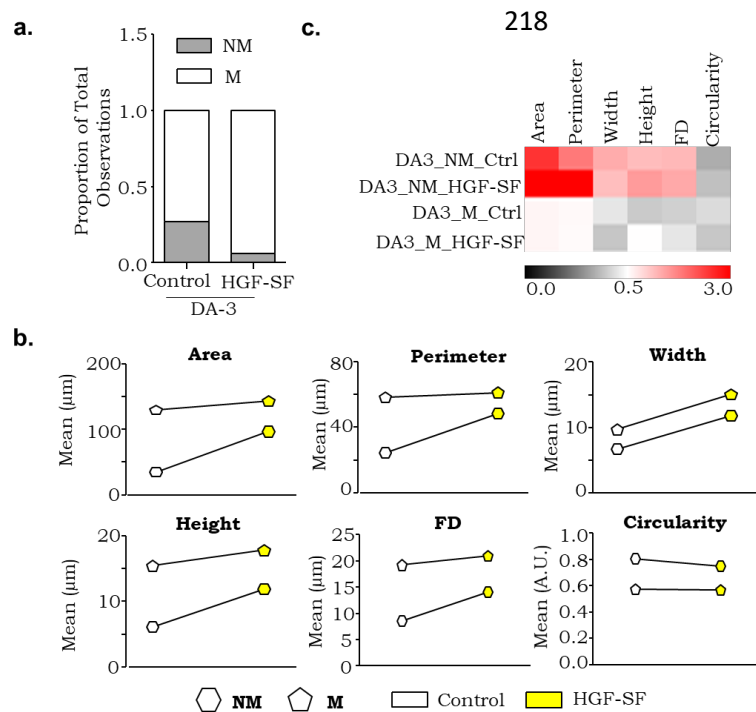
Figure 4. Mouse DA3 cells exhibit extensive non-migrator cell size variability. a. Proportion of non-migratory (NM) and migratory (M) cells in response to HGF-SF treatment. **b.** Mean distribution of CGFs between the NM and M cells. **c.** Heatmap depicting the trend for co-efficient of variance for individual CGFs between the NM and M cells. Data are not temporally segregated. Dataset represented: Zaritsky, et al.

Sparsely seeded cells exhibit greater geometric plasticity in non-migratory as compared to migratory subset

We then investigated the utility of CGF trends in a dataset where directional migration, F-actin organization and focal adhesion maturation were examined in response to anisotropic substrate alignment[26]. Due to differences in image quality, we only extracted data for individually migrating T47D and KPC-derived cells. Annotation of migratory subsets revealed a higher proportion of 'M' cells in KPC as compared to T47D (**Supplementary Figure 7a**). Unlike previous observations, CGF trends demonstrated a cell line specific trend with T47D and KPC exhibiting larger cell size in 'M' and 'NM' subsets, respectively (**Supplementary**

236   **Figure 7b)**. Both systems demonstrated heterogeneity in size with relatively homogenous

237   cell shape features irrespective of the migratory subset analysed (**Supplementary Figure**

238   **7c)**. While these observations may be an outcome of the experimental setup, they could also

239   result from differences in cell packing which is non-existent in sparsely seeded, individually

240   migrating cells.

241

242   *2D packing contributes to CGF variability in a system of confluent monolayer migration*

243   To examine the effects of 2D-packing on cell geometry, we used a dataset where the

244   biophysical dynamics of collective migration were studied on confined PolyDiMethylSiloxane

245   (PDMS) islands [24]. Data were extracted for MDCK cells subject to three experimental setups

246   *viz.*, cytochalasin 'D' treatment (CytoD), high cell density (High) and low cell density (Low).

247   While a higher proportion of 'M' cells was observed in all samples, their frequency increased

248   in the following order CytoD < High < Low (**Figure 5a**). These observations could be

249   explained by the inhibitory effects of CytoD on cytoskeletal re-organization and effects of cell

250   density on 2D-packing and -motility. Unlike previous data, CGF trends were reversed for

251   'NM' and 'M' subsets. Large, less circular 'NMs' were observed for both 'High' and 'Low'

252   experimental groups while 'CytoD' did not significantly affect either migratory subset (**Figure**

253   **5b**). Differences in CGF trends can be explained by the extent of 'NM' spreading, potentially

254   facilitated by collagen coated growth surfaces which were absent in the HGSC and DA3

255   datasets. CoV analysis identified more heterogeneity in 'NMs' versus 'Ms' with 'CytoD'

256   samples exhibiting highest variability in NM size. While cell size variability can be explained

257   as an outcome of CytoD treatment, the greater homogeneity of Ms may be attributed to

258   multiple factors some of which include cell system bias, collagen/ECM-driven uniformity in

259   cell shape and restricted migratory spaces due to NM spreading resulting in limited

260   morphological conformations (**Figure 5c**). Due to the use of a cell monolayer in this data, the

261   CGF trends would principally depend on the extent of cell packing which are influenced by

262   cell division/cell death events, and the frequency of NM-M transitions. These observations

263   highlight the influence of spatial constraints on, driving CGF trends and differentially affecting

264   the migratory subsets.
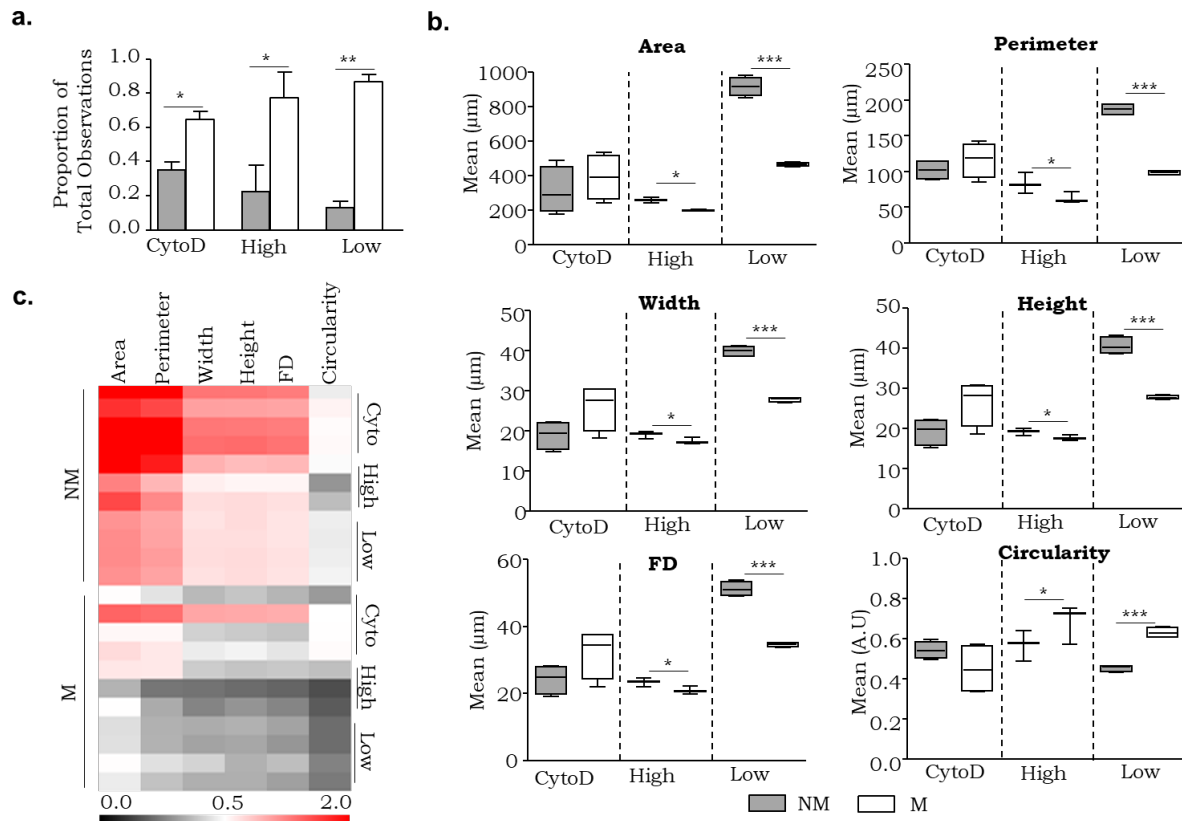
Figure 5. Migration associated CGF trends in a confluent monolayer a. Proportion of non-migratory (NM; grey) and migratory (M; white) cells in response to Cytochalasin D (CytoD) treatment, high density (High) and low density seeding (Low) on micropattern wells. **b.** Mean distribution of CGFs between NM and M cells. **c.** Heatmap depicting the co-efficient of variance for individual CGFs between the NM and M cells. Data are not temporally segregated. Dataset represented: Saraswathibhatla, et al.

Discussion

Decades of research on cell migration and recent developments with *in vitro* assays have been fruitful in understanding several biological processes and for evaluating therapeutic strategies. Due to their widespread utility, analytical pipelines that augment migration read outs have received plenty of attention. In the present study, we have designed a quantitative approach based entirely on open source algorithms for label-free time lapse migration data and applied it for the resolution of functional heterogeneity during migration. Cell geometry was the key parameter evaluated in this analysis and its associated features exhibited consistent trends across experimental systems (**Supplemental Figures 2 and 8**). Our pipeline allows the evaluation of migratory subsets in response to extrinsic cues, wound induction, 2D-packing and provides a foundation to understand migratory transitions.

Our primary dataset for the development of this approach was a panel of HGSC cell lines, where we had previously identified distinct migratory modalities at the population level[27]. Our

280   previous approach had precluded the heterogeneous behaviour of individual cells and

281   exclusively focused on cells that undertook quantifiable displacement during migration.

282   Incorporation of CGFs at a single cell resolution demonstrated the fundamental geometric

283   differences between migratory and non-migratory subsets across phenotypically diverse cell

284   systems including independent datasets; importantly, these trends were unaltered by

285   exogenous cues[25,27,28]. We also observed an unprecedented heterogeneity in the non-

286   migratory subset which may result from one or more of the following reasons: cells existing

287   in different, (i) states of proliferation[29], (ii) functional states while transitioning between

288   migratory subsets[30], or (iii) a greater plasticity of cells with non-specialized functions[31]. The

289   ability of our pipeline to discern proportion of NM to M cells (NM/M) describes the functional

290   composition of cell systems and can prove useful in pharmacological screens. Our

291   observations suggest that the pre-existence of specific migratory subsets and dynamics of

292   their interconversions may be crucial for migration efficiency.

293   Predictive models designed around tissue- and assay-specific variables have suggested the

294   role of cell geometry during *in vitro* migration[29], metastatic invasion[32], and cell polarization for

295   morphogenesis[33]. Experimental evidence linking cell geometry with molecular

296   compartmentalization[34], intercellular crosstalk[35], surface topographies[36] and mechano-

297   reciprocity[35] have further indicated the utility of this parameter in distinguishing functionally

298   distinct cells. Previous reports on colon and breast cancer have also utilized CGFs to predict

299   the chemo-responsiveness[37] and invasive potential[38], respectively, in cell line models. Unlike

300   published studies, our pipeline can be applied not only to scratch assay but also data from

301   confluent monolayers and individual cells. The Saraswathibhatla *et al.*, and Zaritsky *et al.*,

302   datasets were two such distinct studies which upon CGF analysis allowed us to speculate

303   the role of 2D-packing on migratory heterogeneity[24,25]. While we cannot ignore the cell and

304   experimental system specific effects, these trends between migratory subsets do

305   recapitulate jamming-unjamming transitions which are reported during tissue development

306   and disease progression[8,39,40]. While not entirely evident from our analysis, previous reports

307   have provided evidence of cell state transitions in response to spatial packing[41,42].

308   Our analytical pipeline relies on open source free software which can be used across recent

309   computing systems, irrespective of their processing capacity. The application of our

310   approach to label-free systems reduces the dependence on sophisticated labelling or

311   microscopy techniques to capture cell migration. However, due to its reliance on phase

312   differences for distinguishing cells, the quality of input data and presence of debris in the

313   field of view can affect the analysis. We trust that the inclusion of fluorescence microscopy

314   approaches will enhance the sensitivity of our approach and may further resolve CGF

315   trends. We are aware of several image analysis pipelines developed specifically to study cell

316 migration, with each approach presenting a distinct set of advantages[13]. We believe that our
317 pipeline builds on existing quantitative measures and provides a simple analytical platform to
318 dissect the functional heterogeneity during 2D migration. We acknowledge the speculative
319 nature of our conclusions and the need to experimentally validate transitions between NM
320 and M subsets, as well as spatio-temporal specific patterns of NM and M subpopulations.
321 While we initially intended to examine the temporal and spatial heterogeneity influencing
322 migration dynamics, we were unable to capture this information with our current approach,
323 thereby limiting our capacity to investigate the degree of influence of neighbouring cells on
324 migratory behaviour.

325 Overall, we have developed an image analysis pipeline and demonstrated its utility across
326 experimental systems for identifying distinct migratory subsets. While future development will
327 add to its value, we believe that in its current form our analytical approach will greatly benefit
328 biologists interested in the comprehending the process of cell migration.

329

Acknowledgements:

335

Author Contributions:

337 S.S.V. and M.K.J. conceived research. M.K.J. (computational) and S.A.B. (experimental)
338 supervised research. S.S.V. and K.H. performed research and analysed the data. All authors
339 contributed to manuscript writing and have read and agree to this version of the manuscript.

340

Declaration of Interests:

342 The authors declare no conflicts of interest.

343

Data and Code Availability:

Raw data for the HGSC migration dataset is available upon request from S.A.B. Raw data from other datasets used in this study was extracted from the data submitted along with the original publications. Raw data for CGFs is available on the drive link: https://drive.google.com/drive/folders/1gmxfFJyNkyuxmWW3rO623qbgiC8251Vd?usp=sharing. Codes used for structuring, analysing, and visualising the data are available on GitHub at https://github.com/askhari139/CGF.

References:

1. Reig, G., Pulgar, E. & Concha, M. L. Cell migration: from tissue culture to embryos. *Development* **141**, 1999–2013 (2014).

2. Blackley, D. G., Cooper, J. H., Pokorska, P. & Ratheesh, A. Mechanics of developmental migration. *Semin. Cell Dev. Biol.* (2021) doi:10.1016/j.semcdb.2021.07.002.

3. Piacentino, M. L., Li, Y. & Bronner, M. E. Epithelial-to-mesenchymal transition and different migration strategies as viewed from the neural crest. *Curr. Opin. Cell Biol.* **66**, 43–50 (2020).

4. SenGupta, S., Parent, C. A. & Bear, J. E. The principles of directed cell migration. *Nat. Rev. Mol. Cell Biol.* (2021) doi:10.1038/s41580-021-00366-6.

5. Vedula, S. R. K. *et al.* Emerging modes of collective cell migration induced by geometrical constraints. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12974–12979 (2012).

6. Zhang, H. *et al.* Single cell migration dynamics mediated by geometric confinement. *Colloids Surfaces B Biointerfaces* **145**, 72–78 (2016).

7. Fink, A. *et al.* Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns. *Biophys. J.* **118**, 552–564 (2020).

8. Kim, J. H. *et al.* Unjamming and collective migration in MCF10A breast cancer cell lines. *Biochem. Biophys. Res. Commun.* **521**, 706–715 (2020).

9. Liang, C. C., Park, A. Y. & Guan, J. L. In vitro scratch assay: A convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* **2**, 329–333 (2007).

10. Huang, J., Guo, P. & Moses, M. A. A time-lapse, label-free, quantitative phase

375   imaging study of dormant and active human cancer cells. *J. Vis. Exp.* **2018**, 1–8
376   (2018).

377   11. Kramer, N. *et al.* In vitro cell migration and invasion assays. *Mutat. Res. - Rev. Mutat.*
378   *Res.* **752**, 10–24 (2013).

379   12. Varankar, S. S. S. & Bapat, S. A. Migratory Metrics of Wound Healing: A
380   Quantification Approach for in vitro Scratch Assays. *Front. Oncol.* **8**, 1–13 (2018).

381   13. Masuzzo, P., Van Troys, M., Ampe, C. & Martens, L. Taking Aim at Moving Targets in
382   Computational Cell Migration. *Trends Cell Biol.* **26**, 88–110 (2016).

383   14. Das, A. M., Eggermont, A. M. M. & Ten Hagen, T. L. M. A ring barrier-based migration
384   assay to assess cell migration in vitro. *Nat. Protoc.* **10**, 904–915 (2015).

385   15. Wu, S. Y., Sun, Y. S., Cheng, K. C. & Lo, K. Y. A wound-healing assay based on
386   ultraviolet light ablation. *SLAS Technol.* **22**, 36–43 (2017).

387   16. Varankar, S. S. *et al.* Functional balance between Tcf21-Slug defines cellular
388   plasticity and migratory modalities in high grade serous ovarian cancer cell lines.
389   *Carcinogenesis* **41**, (2020).

390   17. Lee, R. M., Vitolo, M. I., Losert, W. & Martin, S. S. Distinct roles of tumor associated
391   mutations in collective cell migration. *Sci. Rep.* **11**, 1–12 (2021).

392   18. Revenu, C. *et al.* Quantitative cell polarity imaging defines leader-to-follower
393   transitions during collective migration and the key role of microtubule-dependent
394   adherens junction formation. *Dev.* **141**, 1282–1291 (2014).

395   19. Reffay, M. *et al.* Orientation and polarity in collectively migrating cell structures:
396   Statics and dynamics. *Biophys. J.* **100**, 2566–2575 (2011).

397   20. Vilchez Mercedes, S. A. *et al.* Decoding leader cells in collective cancer invasion. *Nat.*
398   *Rev. Cancer* **21**, 592–604 (2021).

399   21. Tisler, M. *et al.* Analysis of fibroblast migration dynamics in idiopathic pulmonary
400   fibrosis using image-based scaffolds of the lung extracellular matrix. *Am. J. Physiol. -*
401   *Lung Cell. Mol. Physiol.* **318**, L276–L286 (2020).

402   22. Sampedro, M. F., Miño, G. L., Galetto, C. D. & Sigot, V. Spatio-temporal analysis of
403   collective migration in vivo by particle image velocimetry. *Phys. Biol.* **18**, 066008
404   (2021).

405   23. Elbez, R. *et al.* Cell-morphodynamic phenotype classification with application to

406   cancer metastasis using cell magnetorotation and machine-learning. *PLoS One* **16**,
407   e0259462 (2021).

408 24. Saraswathibhatla, A., Galles, E. E. & Notbohm, J. Spatiotemporal force and motion in
409   collective cell migration. *Sci. Data* **7**, 1–7 (2020).

410 25. Zaritsky, A., Natan, S., Kaplan, D., Ben-Jacob, E. & Tsarfaty, I. Live time-lapse
411   dataset of in vitro wound healing experiments. *Gigascience* **4**, 4–8 (2015).

412 26. Ray, A. *et al.* Anisotropic forces from spatially constrained focal adhesions mediate
413   contact guidance directed cell migration. *Nat. Commun.* **8**, (2017).

414 27. Varankar, S. S. *et al.* Functional Balance between TCF21-Slug defines phenotypic
415   plasticity and sub-classes in high-grade serous ovarian cancer. *bioRxiv* (2018).

416 28. Zaritsky, A., Natan, S., Ben-Jacob, E. & Tsarfaty, I. Emergence of HGF/SF-Induced
417   Coordinated Cellular Motility. *PLoS One* **7**, (2012).

418 29. Matsiaka, O. M., Penington, C. J., Baker, R. E. & Simpson, M. J. Discrete and
419   Continuum Approximations for Collective Cell Migration in a Scratch Assay with Cell
420   Size Dynamics. *Bull. Math. Biol.* **80**, 738–757 (2018).

421 30. Wong, I. Y. *et al.* Collective and individual migration following the epithelial-
422   mesenchymal transition. *Nat. Mater.* **13**, 1063–1071 (2014).

423 31. Huang, B. *et al.* Modeling the Transitions between Collective and Solitary Migration
424   Phenotypes in Cancer Metastasis. *Sci. Rep.* **5**, 1–13 (2015).

425 32. Bratsun, D. A., Krasnyakov, I. V. & Pismen, L. M. Biomechanical modeling of invasive
426   breast carcinoma under a dynamic change in cell phenotype: collective migration of
427   large groups of cells. *Biomech. Model. Mechanobiol.* **19**, 723–743 (2020).

428 33. Stepien, T. L., Lynch, H. E., Yancey, S. X., Dempsey, L. & Davidson, L. A. Using a
429   continuum model to decipher the mechanics of embryonic tissue spreading from time-
430   lapse image sequences: An approximate Bayesian computation approach. *PLoS One*
431   1–23 (2019) doi:https://doi.org/10.1371/journal.pone.0218021.

432 34. Pereira, D., Richert, A., Medjkane, S., Hénon, S. & Weitzman, J. B. Cell geometry and
433   the cytoskeleton impact the nucleo-cytoplasmic localisation of the SMYD3
434   methyltransferase. *Sci. Rep.* **10**, 1–12 (2020).

435 35. Roffay, C., Chan, C. J., Guirao, B., Hiiragi, T. & Graner, F. Inferring cell junction
436   tension and pressure from cell geometry. (2021) doi:10.1242/dev.192773.

36. Vasilevich, A. S. *et al.* On the correlation between material-induced cell shape and phenotypical response of human mesenchymal stem cells. *Sci. Rep.* **10**, 1–15 (2020).

37. Pasqualato, A. *et al.* Shape in migration: Quantitative image analysis of migrating chemoresistant HCT-8 colon cancer cells. *Cell Adhes. Migr.* **7**, 450–459 (2013).

38. Baskaran, J. P. *et al.* Cell shape, and not 2D migration, predicts extracellular matrix-driven 3D cell invasion in breast cancer. *APL Bioeng.* **4**, 1–16 (2020).

39. Park, J. A. *et al.* Unjamming and cell shape in the asthmatic airway epithelium. *Nat. Mater.* **14**, 1040–1048 (2015).

40. Garcia, S. *et al.* Physics of active jamming during collective cellular motion in a monolayer. *Proc. Natl. Acad. Sci.* **112**, 15314–15319 (2015).

41. Fabian Spill, Vivi Andasari, Michael Mak, Roger D Kamm, M. H. Z. Effects of 3D geometries on cellular gradient sensing and polarization. *Phys. Biol.* **176**, 139–148 (2017).

42. Rupprecht, J., Haur, K., Yin, J., Huang, A. & Dinh, H. Geometric constraints alter cell arrangements within curved epithelial tissues. *Mol. Biol. Cell* **8**, 3563–3725 (2017).