An analytical approach differentiates between individual and collective cancer invasion

Elad Katz[a,1,*], Wim Verleyen[b,1], Colin G. Blackmore[c], Michael Edward[d], V. Anne Smith[b] and David J. Harrison[a]

[a]*Breakthrough Research Unit and Division of Pathology, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK*
[b]*School of Biology, University of St Andrews, St Andrews, UK*
[c]*Definiens AG, Munich, Germany*
[d]*Faculty of Medicine, Section of Dermatology, Division of Cancer Sciences, University of Glasgow, Glasgow, UK*

Abstract. Tumour cells employ a variety of mechanisms to invade their environment and to form metastases. An important property is the ability of tumour cells to transition between individual cell invasive mode and collective mode. The switch from collective to individual cell invasion in the breast was shown recently to determine site of subsequent metastasis. Previous studies have suggested a range of invasion modes from single cells to large clusters. Here, we use a novel image analysis method to quantify and categorise invasion.

We have developed a process using automated imaging for data collection, unsupervised morphological examination of breast cancer invasion using cognition network technology (CNT) to determine how many patterns of invasion can be reliably discriminated. We used Bayesian network analysis to probabilistically connect morphological variables and therefore determine that two categories of invasion are clearly distinct from one another.

The Bayesian network separated individual and collective invading cell groups based on the morphological measurements, with the level of cell-cell contact the most discriminating morphological feature. Smaller invading groups were typified by smoother cellular surfaces than those invading collectively in larger groups. Interestingly, elongation was evident in all invading cell groups and was not a specific feature of single cell invasion as a surrogate of epithelial-mesenchymal transition.

In conclusion, the combination of cognition network technology and Bayesian network analysis provides an insight into morphological variables associated with transition of cancer cells between invasion modes. We show that only two morphologically distinct modes of invasion exist.

Keywords: Invasion, imaging, breast cancer, Bayesian networks, EMT

1. Introduction

Spread and metastasis of breast cancer is responsible for the vast majority of patient morbidity and mortality associated with this disease [1]. The invasive properties of cancer cells underlying metastatic disease have been investigated extensively for many years [2–4], with new treatments beginning to enter clinical trials. There are few effective, short term biomarkers. Novel approaches in clinical trials include cytotoxins such as taxanes and biological agents, especially those targeting tyrosine kinases such as Trastuzumab (Herceptin) [5]. Anti-angiogenic approaches such as the use of the anti-VEGF-A antibody Bevacizumab (Avastin) in metastatic breast cancer have failed so far to improve the prognosis of metastatic breast cancer patients [6]. The 5-year survival of patients with metastatic

---

[1] These authors contributed equally.

*Corresponding author: Elad Katz, Breakthrough Research Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. Tel.: +44 131 537 3154. E-mail: elad.katz@ed.ac.uk.

breast cancer patients is currently estimated at 26% [1, 7].

Cancer cells are capable of transitioning from one mode of invasion to another in response to both endogenous cues and the environment [8]. The tumour environment influences the mode of cancer cell invasion which has been labelled individual or collective [9]. Preservation of cell-cell contacts enables a larger number of cells to invade collectively, rather than as single cells which may be undergoing epithelial to mesenchymal transition (EMT) [10]. Christofori and colleagues have suggested a classification of cancer cell invasion, broadly dividing it to "single" cell and collective invasion [2]. Collective invasion was further divided into "cohort invasion" of 6–10 cells loosely attached to the bulk of the tumour [11] and "coordinated invasion" of larger cell groups, maintaining some contact with the bulk of the tumour [12]. The biological and clinical distinction between these groups remains uncertain. Importantly, even with the same genetic background, individually invading cells target distal sites such as the lung and collectively invading cells showed preference for lymph node metastasis [13]. Until now, the sole morphological distinction between invasion modes was the number of cells invading [2]. If morphological type of invasion is to be a useful biomarker it must be reliably and reproducibly measurable using a multitude of morphological parameters. The analysis method also would have to be feasible in the clinical context, where paraffin-embedded primary breast cancer tissue is often the only available material. Previous work has relied heavily on intravital or confocal imaging [14, 15], both of which are unlikely to be implemented in the clinic.

While individual cell invasion is usually associated with down regulation of cell-cell contact molecules such as E-cadherin [16], somewhat paradoxically, it does not imply necessarily that the cells are invading entirely as single cells. In lobular breast carcinomas, despite down-regulation of E-cadherin, small groups of cells invade together in single file formation [17]. Triple negative breast cancers (ER- PR- HER2-) are usually associated with EMT and poor prognosis, but rarely display a complete switch from collective to individual invasion, probably due to up-regulation of N-cadherin [18]. Collective cell invasion has not been investigated as much as individual cell invasion *in vitro*, although it is a very prominent feature in human cancer biopsy samples [10, 19–21]. In squamous cancer cells, several key proteins were implicated in collective inva-

sion type, including a mucin-type glycoprotein called podoplanin [22] and the small GTPase CDC42 [21].

Bayesian networks have been previously applied in cancer biology, usually as a method of revealing mechanism or underlying interaction in high-throughput data [23–25], but they have not been used in morphological analysis. Using Bayesian network analysis, we show that automatically measured morphological variables can characterise two invasion types in organotypic culture, individual and collective, which can be discriminated by a number of morphological variables: cell-cell contact, invading group area and surface roughness.

2. Materials and methods

2.1. Cell lines, collagen invasion assays and tumour materials

H16N-2 is an immortalized cell line derived from normal breast epithelium. H16N-2 cells were cultured in DFCI media [26]. Generation of the C35 transfected pool (C35pool) and the C35.C3 clone (C35hi) is described in detail elsewhere [27].

For collagen invasion assays, rat collagen type I solution (0.3–0.5 mg/ml) was mixed with 10^5 human breast fibroblasts (from reduction mammoplasty) per 3 ml lattice and left to contract in fibroblast media (DMEM (Invitrogen) supplemented with 10% serum, 50 U/ml penicillin, and 50 mg/ml streptomycin) for ~7 d. When the lattices were of the required size (~4-fold contraction), 3×10^5 H16N2 cells from the desired lines were seeded on top and the lattices and incubated as submerged cultures for 4 d in H16N2 media. To induce invasion, the lattices were raised to the air/liquid interface and incubated for a further 7 d before the lattices were fixed in 10% phosphate buffered formalin and wax embedded.

Usage of tumour material was approved by the Lothian Research Ethics Committee (08/S1101/41).

2.2. Image acquisition

Immunofluorescence was performed using methods previously described using the AQUA system (HistoRx) [28]. Pan-cytokeratin rabbit polyclonal antibody (Dako, 1 : 150) was used to identify epithelial cells. The epithelial compartment was then visualised with anti-rabbit-Cy3 (Invitrogen, 1 : 25). DAPI (4′,6-

diamidino-2-phenylindole) counterstain (Invitrogen) was used to identify nuclei.

Monochromatic high-resolution (1024 × 1024 pixel, 0.5-μm resolution) images of each invasion assay were captured at × 20 objective using an Olympus AX-51 epifluorescence microscope.

2.3. Immunohistochemistry

Antigen retrieval for vimentin staining (Sigma, 1 : 400) was performed using sodium citrate buffer (18 μM citric acid, 82 μM sodium citrate, pH 6.0). Standard immunohistochemistry protocol was performed using the REAL EnVision mouse kit (Dako), according to manufacturer's instructions.

2.4. Automated cell delineation

Images were analysed with the Cognition Network Technology (CNT) Definiens Cellenger modular application framework [29]. This framework allows the simultaneous existence of image "*objects*" at different orders of magnitude in a hierarchical organisation

[30] (Fig. 1). In this study, a "Nucleus Level", was first generated. *Nucleus* objects were formed using the DAPI signal (blue image channel) via semi-adaptive background thresholding, followed by an adapted watershed transformation and constrained growth of intermediary image objects. These *nucleus* objects were used as seed points to grow surrounding *cytoplasm* objects from the cellular pan-cytokeratin mask (green image channel, Supplementary Figure 1), using an automatic adaptive threshold. This process resulted in the formation *nucleus* and *cytoplasm* objects at the Nucleus Level. These were combined to create *cell* objects at the "Cell Level", containing of *nucleus* and *cytoplasm* "sub objects". The *cell* objects represented all epithelial (nucleus[+] cytokeratin[+]) cells within a given immunofluorescence image, fully delineated for further morphological analysis. All stromal cells (cytokeratin[−]) were excluded at this stage.

We further developed the CNT hierarchy to define group objects at a top "Group Level" (see Results). These groups were classified into tumour invasion types and were the base unit from which morphological measurements were made.
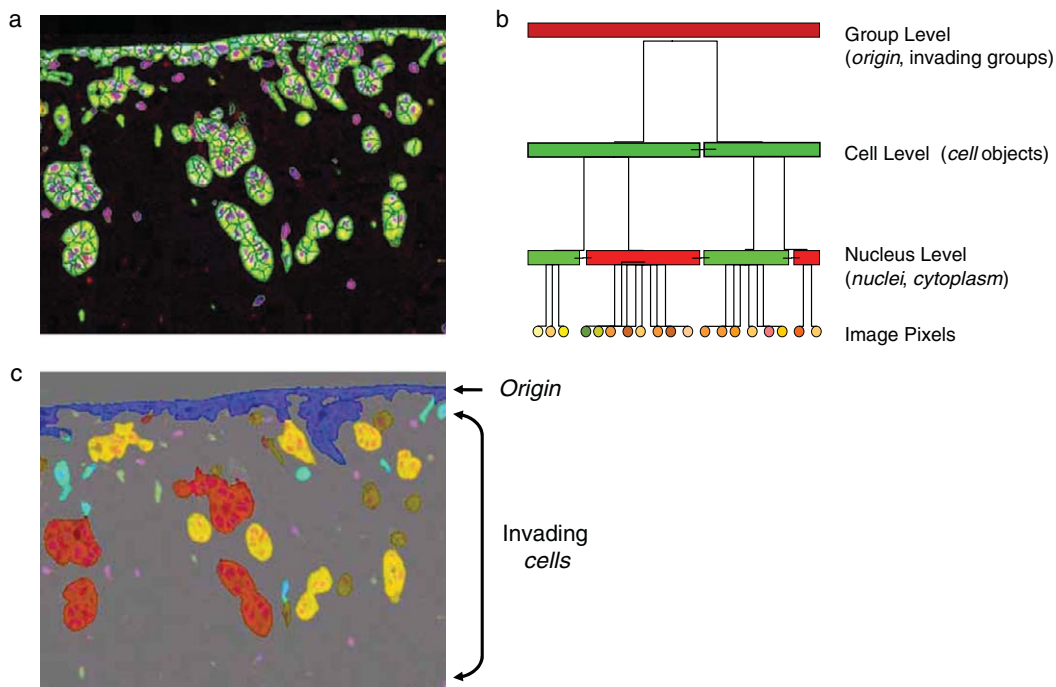

Fig. 1. Cell segmentation and group size analysis using CNT. (a) A typical imaging field used for CNT-based analysis. (b) Flow-chart of cell and group segmentation using CNT. (c) Resulting cell and group delineation, including exclusion of *origin* (in blue).

"Roundness" of an invading cell group describes how similar the group is to an ellipse. A perfectly round object has the value of 0. The roundness value is higher for non-round objects.

"Roughness" (or "shape index") of an invading cell group describes the ratio between the group border length and optimal border length (as defined by the group length). A perfectly smooth object has the value of 1. A group with jagged borders or with cell projections will have a higher roughness value.

2.5. Bayesian network analysis

Bayesian networks are a method to visualise statistical dependencies, drawn as lines between variables. Among a set of highly correlated variables, they will only connect those with direct statistical dependence (i.e., not mediated through other measured variables) [31]. The dependencies can be annotated with information about the type of relationship: positive (meaning high values in one correspondence with high values in another, and low with low), negative (high values in one correspond with low in another), or non-monotonic (such as U- or hump-shaped relationships); relative strengths are available for positive and negative relationships (scaled between 0 and 1; [32]).

We applied the publicly available Bayesian network software BANJO (http://www.cs.duke.edu/~amink/software/banjo/; [33]) to construct Bayesian networks of selected morphological variables and tumour invasion types. The data was first discretised with quantile discretisation using three levels [32]. The software was then applied with the following settings: static
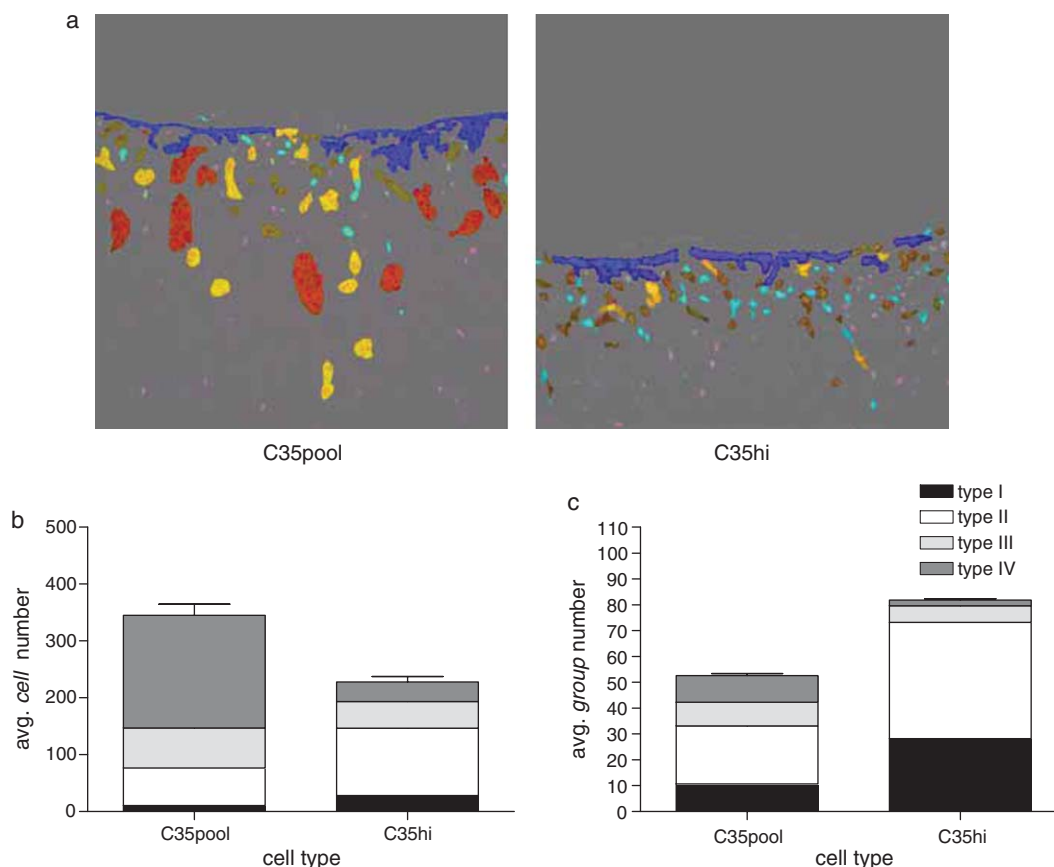

Fig. 2. Detailed analysis of invasion using CNT-based approach. (a) Typical images and group size distribution by CNT-based image analysis software for H16N2 cells: C35 pool and C35hi. Colour coding: type I (single cells) in turquoise, type II (2–5 cells) in brown, type III (6–10 cells) in yellow and type IV (>10 cells) in red. Cell types distributions are show by invading cell numbers per imaging field (b) or invasion type groups per imaging field (c).

networks, equivalent sample size equal to one, greedy search with random resets every 3000 iterations, quantile discretisation with three levels, maximum parent count equal to three (maximal number allowed by data amount [34]), relationships among tumour invasion types disallowed (mutually exclusive categories), and a consensus network generated from the top 100 networks. Because direction of statistical relationship had no biological meaning, results are shown as consensus networks with undirected relationships to ease interpretation.

The Bayesian network provides two types of information based on its structure. First, invasion types sharing links with the same morphological variables are likely to be similar. Second, variables directly connected to invasion types are the most important for those types. To confirm the latter, we applied Student's *t*-test to evaluate whether the means of morphological variables differed for invasion types within each cell line; reported *p*-values have been Bonferroni-corrected over all tests.

3. Results

3.1. Construction of model system

As an *in vitro* invasion model, we used H16N-2 breast cells expressing the C35 protein in a collagen lattice invasion model. C35 is an oncogene co-amplified with HER2 in breast cancers and is capable of inducing mammary epithelial cell invasion *in vitro* [26, 27]. While empty vector expressing cells (Null) did not significantly invade the lattice, expression of C35 in cells led to invasion [27]. The C35 transfectant pool (C35pool), which has variable, intermediate levels of C35 expression, invades mostly as large clusters of cells, maintaining overall levels of cell-cell contact molecules, including E-cadherin [27]. In 3-dimensional (3D) organotypic culture, C35pool cells display an invasive morphology with high nuclear grade, comparable to moderately-differentiated breast cancers (E. Katz and Jeremy S. Thomas, unpublished observations). A clone which expresses C35 at high levels (C35hi) shows complete transition to spindle cell phenotype, with individual cells invading deep into the lattice. Invading C35hi cells are comparable to EMT-enriched claudin-low breast cancers [35]. This model system enables comparison of multiple invasion types in similar genetic, cellular and environmental contexts.

3.2. Developing "Group Level" in CNT analysis

A constrained fusion of *cell* objects was performed to cluster formations of cells into compact groups, forming a "Group Level". Firstly, all *cell* objects with a common border were fused together and small enclosing holes filled to provide contiguous objects. A shape splitting algorithm was then applied to the fused *cells*, which cut the objects satisfying local conditions (constrained by the minimum concave angle over the border of 25 degrees, the length of the border over which the object was to be cut, 20 pixels, and the maximum length over which objects would be cut, 50 pixels), creating *group* objects. This resulted in objects representing *groups* of *cells*. However, it was necessary to differen-
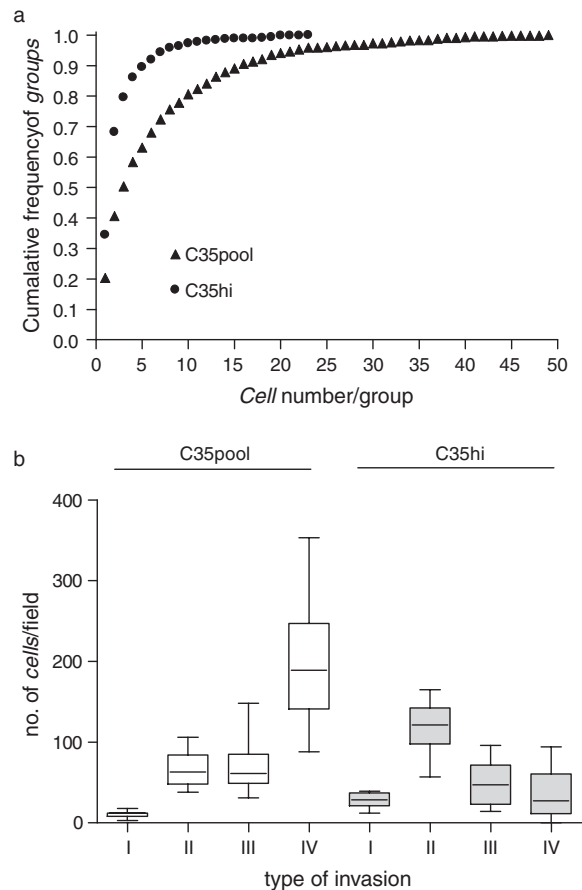

Fig. 3. Profiles of C35-induced invasion. (a) Frequencies of group sizes (by cell number) in C35pool and C35hi cells. (b) Distribution of cells amongst different invasion types (averaged across all imaging fields examined) in C35pool and C35hi cells.

tiate *groups* attaching and/or proliferating on top of the collagen lattice from groups of invading cells. Therefore, we first defined *groups* belonging to an *origin* object (Fig. 1c). The *origin* was identified by finding a characteristic seed *group* object (using a high length to width ratio). Similar *origin cell* objects were then identified using a fuzzy logic approach, using a combination of area, similarity in direction to the *origin* seed and length to width, to identify the remaining *origin cell* objects. A series of local and constrained merging steps fused *group* objects with similar positional and/or border lengths greater than 22% with the *origin* to finalise the *origin* structure. This resulted identification of invading *groups* which were not part of the *origin*. These *groups* could be either completely disjoint from the origin or separated via a short common border.

Invading *groups* were classified into four tentative invasion types by examination of 2-dimensional sec-

tions of cultures, following the Christofori model [2]: single *cells* (type I), 2–5 *cells*/*group* (type II), 6–10 *cells*/group (type III) and >10 *cells*/*group* (type IV). Type II consists of *groups* too small (<6 *cells*) to be cohorts by the Christofori definition, but are not single *cells*. Such small groups have been documented *in vivo* [36].

Examples of group classification using CNT in C35pool and C35hi cells are shown in Fig. 2a. Overall, most *groups* in the C35hi were categorised as type I or II (Fig. 2b). C35pool *groups* were divided into four types (Fig. 2c). Overall profiles of invading group sizes are shown in Fig. 3.

3.3. Choice of morphological variables

The CNT software package offers >100 object-based variables that can be measured. Of those, we used
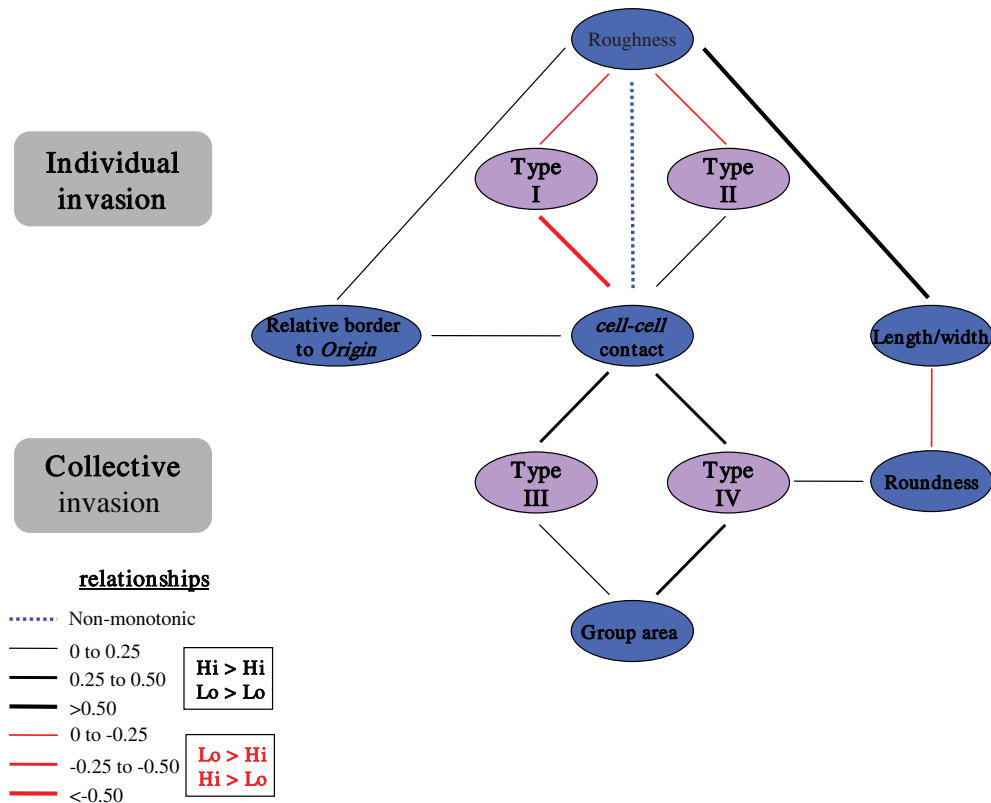

Fig. 4. The Bayesian network learned from the *group* morphological variables measured with CNT analysis together with the four Christofori types of tumour invasion. The relationships in the Bayesian network are illustrated with black for positive, red for negative and blue dotted line for non-monotonic (see Materials and methods). The strength of positive and negative relationships is reflected by line thickness as in legend shown. The grey boxes on the left indicate interpretations of Types I and II as individual invasion, and types III and IV as collective invasion.

six candidate variables with clear potential biological implications. Variables were calculated on the invading *group* objects and allow comparison of the behaviour of invasive cells regardless of invasion type employed. Five variables were intrinsic properties of the invasive groups: their area (in pixels), length/width ratio, roundness and surface roughness. The first two parameters are thought to be detrimental in the epithelial to mesenchymal transition underlying individual cell invasion [37]. Another variable calculated was the mean of *cell-cell* contact (the border between *cell* objects) found in the invasive *group*. An additional variable was "relative border to *origin*", an indication of how much contact exists between the invading *group* and the *origin*, that is *cells* seeded on the surface of the

gel. Christofori et al. suggested that the value for this variable is high in coordinated invasion, but low in other invasion modes such as cohort invasion [2].

3.4. Bayesian network analysis of invasion types and morphological variables

Bayesian network analysis was applied to the output of the image analysis in order to examine which morphological variables could be important for categorisation of tumour invasion types. We originally attempted to classify invading *groups* into four categories based on *group* size (types I, II, III and IV). A consensus Bayesian network was created containing the six selected parameters as described above and
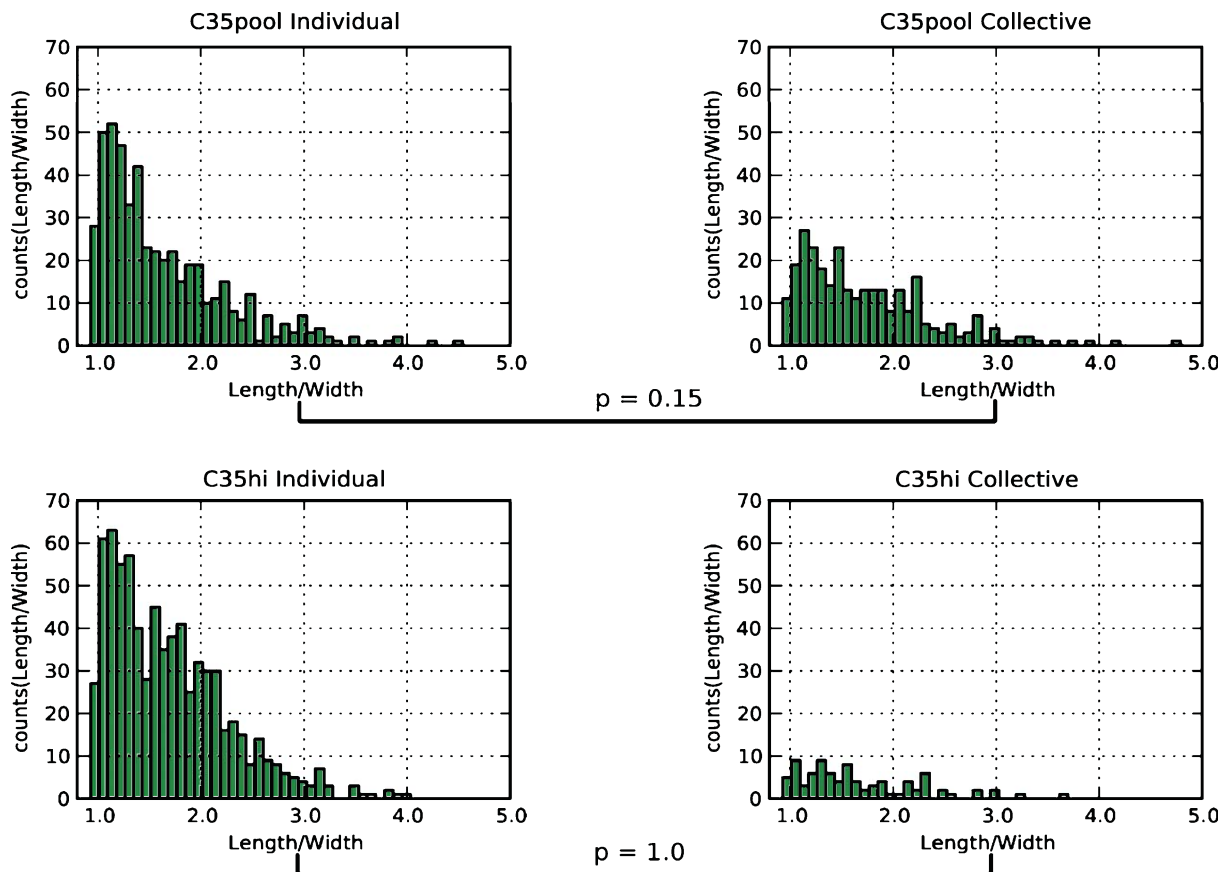

Fig. 5. Length/width ratio does not differ between individual and collective invasion. Distributions of length/width ratio value shown for C35pool (*top panels*) and C35hi cells (*bottom panels*): shown are histograms depicting counts of measurements taken which fall inside the range of the bar on *x*-axis. *p*-values show signficance of tests between individual and collective for each cell line, corresponding to Bonferroni-corrected *t*-test (corrected for the 8 tests in Figs 5–8).

four Christofori types (Fig. 4). Neither length/width nor relative border to *origin* were directly related to any invasion type suggesting they are less characteristic than other variables of invasion type. The relationships to other morphological variables, however, revealed that Christofori types I and II could be classified as individual and types III and IV as collective: individually invading *groups* (type I and II) have a negative relationship with surface roughness, while collective invasion (types III and IV) had a positive relationship to *group* area. All types were highly related to *cell-cell* contact.

Next, we examined how these variables revealed as important in the Bayesian network analysis on data from both cell lines combined can be generalised to

differences within individual cell lines. Some morphological variables are more connected to tumour invasion types (Fig. 4) and we expect this to correspond to a significant difference in their distributions between invasion types within a cell line (Supplementary Table 1). As a control, we first examined length/width ratio, which was not directly connected to any invasion type. Individual cell invasion is usually associated with mesenchymal spindle-like morphology [37, 38] implying a length/width ratio significantly greater than 1. However, the Bayesian network analysis on this dataset suggested that this morphological variable is rather marginal. Indeed, it showed no significant difference between individual and collective invasion for the two cell lines (Fig. 4; $p \geq 0.15$). We found ratios
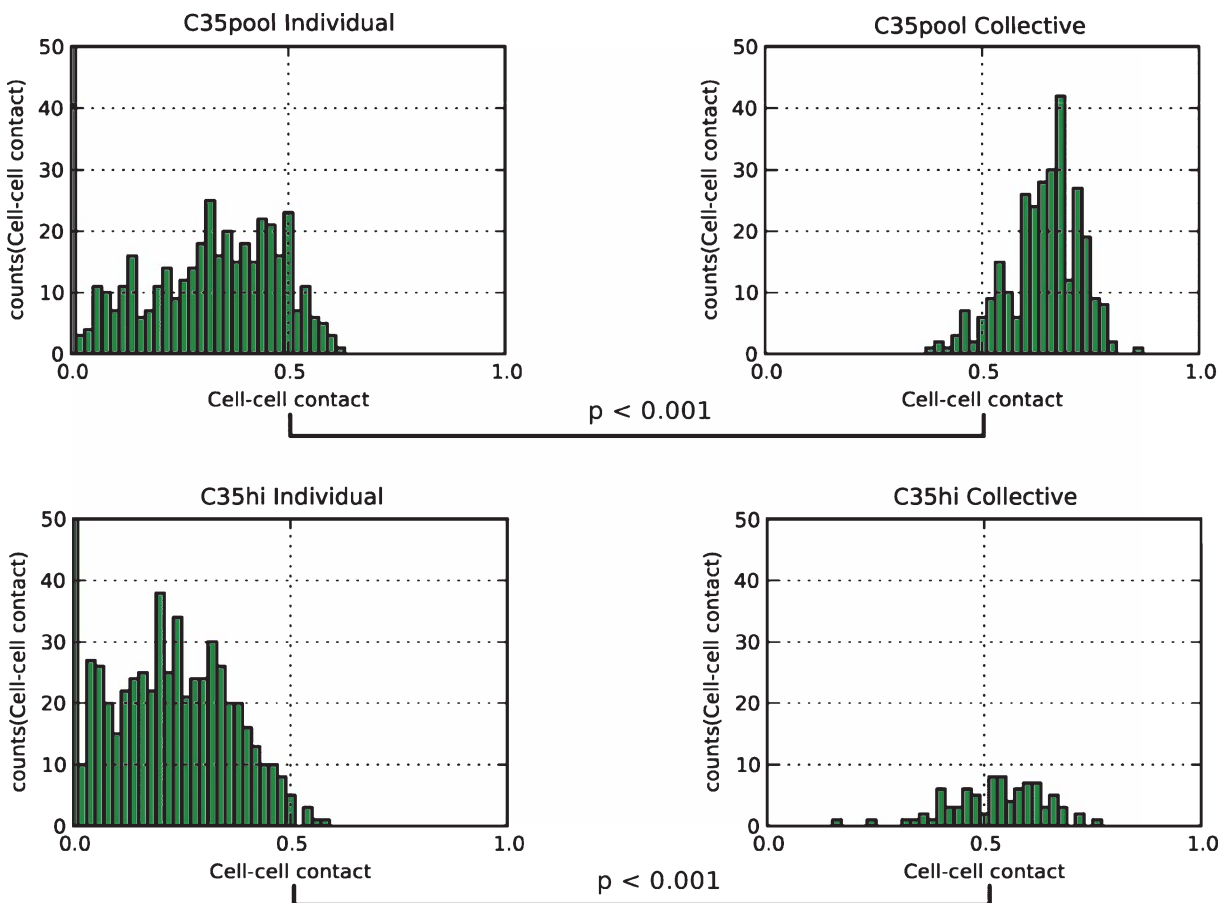

Fig. 6. *Cell-cell* contact is greater in collective than individual invasion. Distributions of *cell-cell* contact value shown for C35pool (*top panels*) and C35hi cells (*bottom panels*): shown are histograms depicting counts of measurements taken which fall inside the range of the bar on *x*-axis. *p*-values show significance of tests between individual and collective for each cell line, corresponding to Bonferroni-corrected *t*-test (corrected for the 8 tests in Figs 5–8).

as high as 5 in some of the invading *groups*, but these were not restricted to any particular *group*, either individual or collective. Recent findings in border cells of the Drosophila ovary also suggest that invading group elongation (i.e. high length/width ratio) can also be found in collectively invading cells [39]. In contrast to length/width ratio, the three variables prominent in the Bayesian network analysis across both cell lines were significantly different between individual and collective invasion also within the same cell line: *cell-cell* contact, *group* area and surface roughness. The high statistical significance observed in all of these variables (all *p* values <0.0001) strongly supports the notion that these morphological variables can be used to define invasion modes in cancer.

Typically, individually invading cells, down-regulate cell-cell contact molecules part of their epithelial to mesenchymal transition [16]. Collectively invading cells maintain their cell-cell contacts as they invade [8]. As expected, *cell-cell* contact was significantly greater in collectively invading *groups* than individually invading *groups* (Fig. 6). The maintenance of cell-cell contacts in collectively invading cells is not dependent on E-cadherin levels, as they are very low in all C35hi *group* types (data not shown). *Group* area was significantly greater in collective invasion compared to individual invasion (Fig. 7). This is sensible as more cells make for a greater area, however, it is not necessarily trivial (as cell size could have been smaller [37]), and thus provides useful mor-
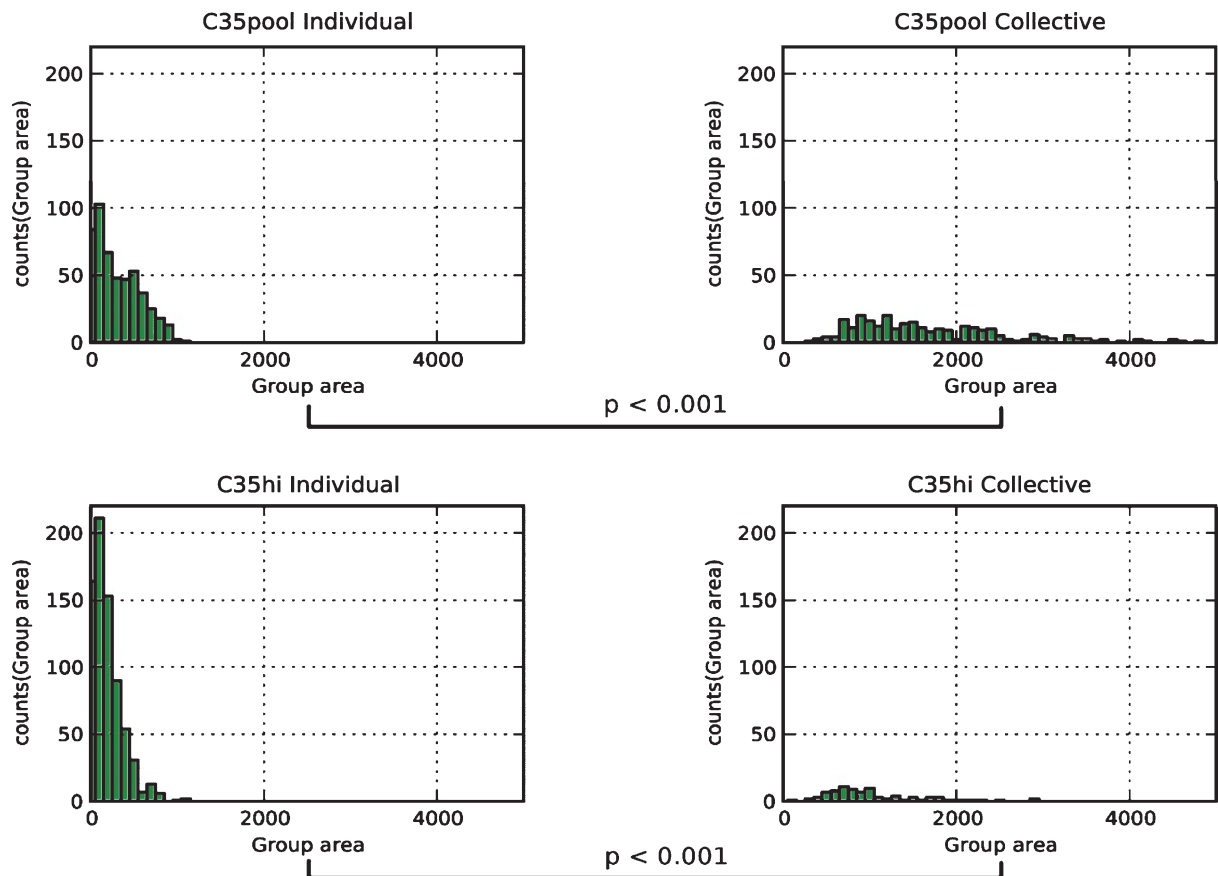

Fig. 7. Group area is greater in collective than individual invasion. Distributions of group area value shown for C35pool (*top panels*) and C35hi cells (*bottom panels*): shown are histograms depicting counts of measurements taken which fall inside the range of the bar on *x*-axis. *p*-values show signficance of tests between individual and collective for each cell line, corresponding to Bonferroni-corrected *t*-test (corrected for the 8 tests in Figs 5–8).
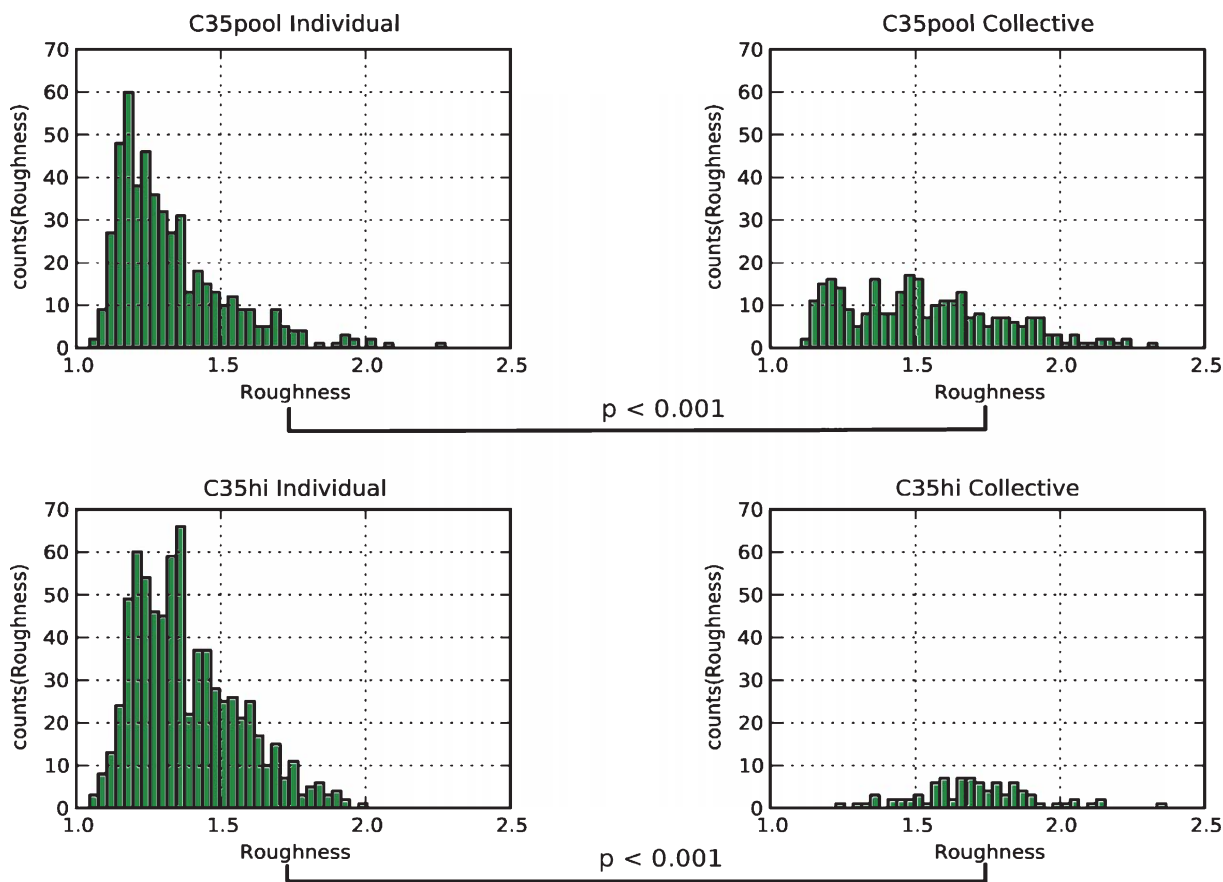
Fig. 8. Roughness is greater in collective than individual invasion. Distributions of roughness value shown for C35pool (*top panels*) and C35hi cells (*bottom panels*): shown are histograms depicting counts of measurements taken which fall inside the range of the bar on *x*-axis. *p*-values show significance of tests between individual and collective for each cell line, corresponding to Bonferroni-corrected *t*-test (corrected for the 8 tests in Figs 5–8).

phological information about the size of invading groups.

3.5. Roughness as a morphological characteristic of collectively invading cells

The surface roughness of *groups* was significantly less in individual compared to collective invasion (Fig. 8). This implies that single or small *groups* of cells (1–5 *cells*, type I/II) may invade in *groups* with smoother surface than those invading in larger collective *groups* (> 5 *cells*, type III/IV).

In order to better understand the source of roughness, we probed C35pool cells for the invadopodia component vimentin. Invadopodia are most common

mechanism employed by cells invading 3D matrices [14]. Vimentin is a critical component of these invasive structures [40] and it is expressed in cells at the front of collective invading groups (Fig. 9a). Interestingly, a similar pattern of vimentin expression was found in a metastatic breast tumours ($n = 5$; a representative tumour shown in Fig. 9b). In both cases, these structures were also positive for MT1-MMP (data not shown), consistent with invadopodia as described elsewhere [14].

4. Discussion

Studying invasion in human cancer is difficult, and biomarkers are few. In some animal studies *in vivo* imaging has been used, but there is a dearth of robust
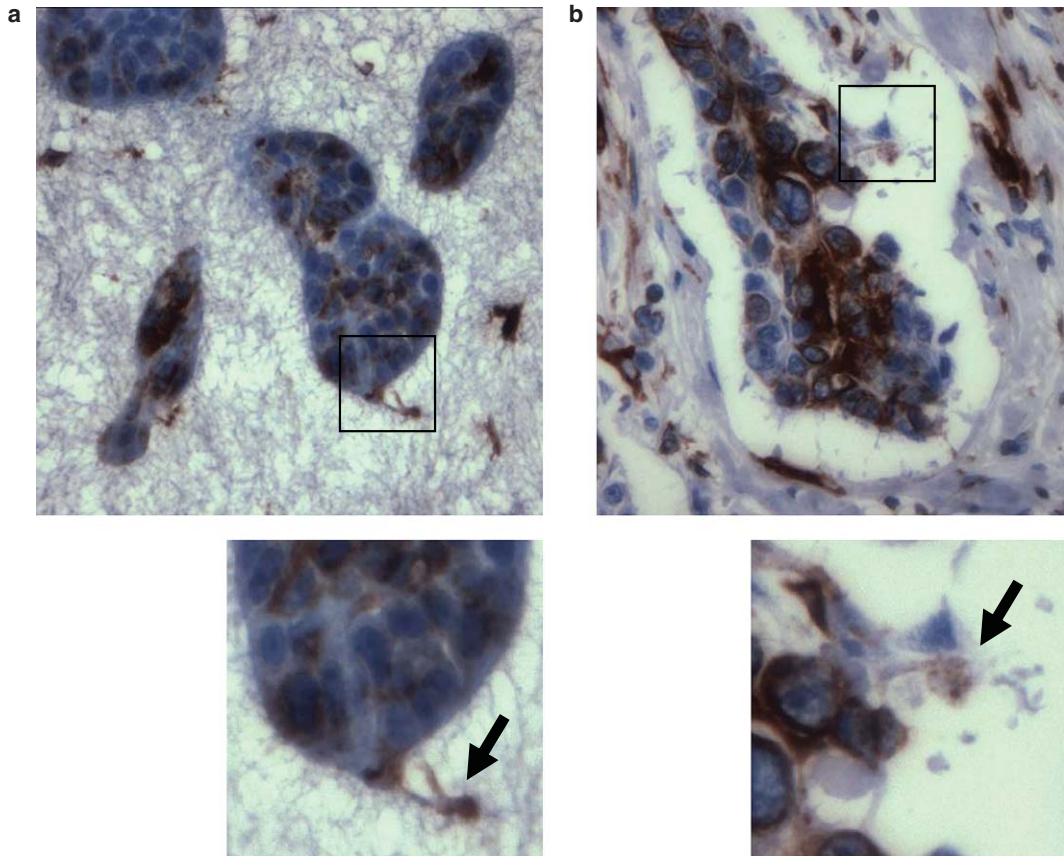
Fig. 9. Invadopodia as a source of surface roughness in collectively invading groups. Examples of vimentin staining in (a) C35pool and (b) high grade breast tumour. In *lower panels*, arrows indicate prominent vimentin$^+$ protrusions. Original magnification: ×40.

biomarkers for quantifying effects of new treatments on invasion. Mathematical models factoring common events such as proliferation and cell death predict uneven tumour border, associated with tumour invasiveness [41–43]. Many 3D culture approaches also suggest that the behaviour of cells in ECM is predictive of their *in vivo* behaviour [44–46]. Nonetheless, even extensive studies rely in their interpretation on the overall appearance of cell clusters without direct correlation to an invasion measurement [41, 45]. Ultimately, these observations led to a proposal of four types of invasion: single cells, small groups of cells, cohorts and collective invasion [2]. This study analyses invasion using cell morphology, discriminating between possible different modes of invasion and reliably identifies two categories rather than four. This approach is particularly pertinent due to our growing understanding that tumour cells are capable of switching between different invasion modes, even within the same tumour [8, 13]. We evaluated a total of six morphological variables, most intrinsic to the invading *group* (e.g. *cell-cell* contact), another showing the relationship to surrounding cells or the *origin* from which the invasion started. Our Bayesian network analysis determined that morphologically two modes of invasion exist. These modes, individual and collective, are distinct by three biomarkers: cell-cell contact, invading group size and group surface roughness.

We observed here that collectively invading *groups* have significantly rougher surfaces than those invading individually (1–5 *cells*). Mathematical modelling suggests roughness in collective invasion could be stemming from tumour cell interactions with the ECM [47], generating to invadopodia or similar cellular processes [48, 49]. We found that some of C35pool cells express vimentin, an important component of

invadopodia [40]. These processes were found around collectively invading cell groups. We are currently exploring what mechanisms may drive roughness in collective invasion. Interestingly, we did not find that these include generation of peripheral actomyosin forces by Rho kinases ([50] and data not shown).

It is possible that down-regulation of cell junctions and cytokeratins which typifies individually invading cells paradoxically enables them to maintain a more smooth morphology than those invading collectively. Some evidence exists that motile, individually invading cancer cells have a smoother appearance than non-motile cells _in vivo_ [51]. Furthermore, invasive lobular carcinomas of the breast also contain round and smooth invading single cells [17].

An additional variable highlighted by Bayesian network analysis is that only collectively invading _groups_ (>5 _cells_) are similarly linked to the overall _group_ area. This implies that unlike what is predicted in the Christofori model, division into cohorts of 6–10 cells and bigger coordinated groups has no morphological grounds. This conclusion is strengthened by the absence of any direct links between both type III (presumed "cohorts") or IV (presumed "coordinated") and the variable describing relative border to _origin_. This would not be the case if coordinated groups were indeed keeping contact with the tumour origin.

Morphological examination of cancer cells is regularly used in clinical pathological practice to determine tumour grade. Until now, morphological analysis has not being used extensively in the study of cancer invasion _in vitro_. We show here that using our approach only two distinct modes of invasion are consistently found. Morphological differences between those invasion modes could now be exploited in both organotypic cell lines models and human cancer specimens to measure changes in tumour progression or drug response. Recent studies suggest that findings from organotypic models such as this used here could be directly translated to the _in vivo_ setting [46].

Acknowledgments
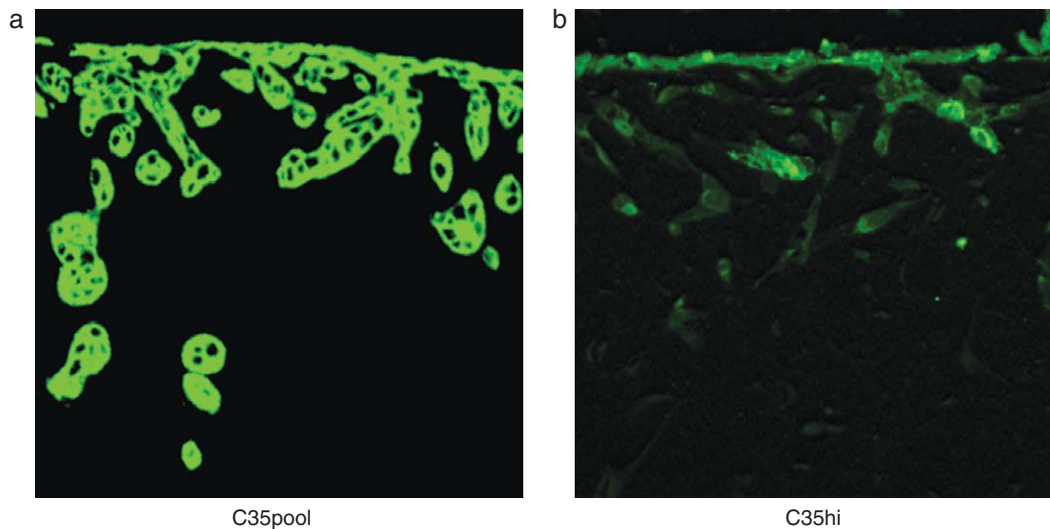
Supplementary Figure 1. Examples of pan-cytokeratin outlining of invasion in (a) C35pool cells and (b) C35hi cells. The cytokeratin mask (green) was used to determine cell morphology in subsequent image analysis.

Supplementary Table 1

Summary of statistical analysis for Figs 4–7. ns, not significant

		n			*n*		
Numbers of group object analysed							
C35pool	individual	498		collective	290		
C35hi	individual	732		collective	86		
		Mean	SD		Mean	SD	Corrected *p* value
Length/width							
C35pool	individual	1.68	0.64	collective	1.79	0.67	0.15 (ns)
C35hi 0	individual	1.73	0.57	collective	1.72	0.58	1.00 (ns)
Cell-cell contact							
C35pool	individual	0.25	0.19	collective	0.65	0.08	*p* < 0.001
C35hi	individual	0.17	0.15	collective	0.53	0.11	*p* < 0.001
Group Area							
C35pool	individual	355.81	255.78	collective	1956.52	1243.44	*p* < 0.001
C35hi	individual	236.40	176.51	collective	1085.98	575.03	*p* < 0.001
Roughness							
C35pool	individual	1.335	0.189	collective	1.543	0.272	*p* < 0.001
C35hi	individual	1.392	0.179	collective	1.712	0.201	*p* < 0.001

References

[1] J. Lu, P.S. Steeg, J.E. Price, et al., Breast cancer metastasis: challenges and opportunities, *Cancer Res* **69**(12) (2009), 4951–4953.

[2] M. Yilmaz, G. Christofori and F. Lehembre, Distinct mechanisms of tumor invasion and metastasis, *Trends Mol Med* **13**(12) (2007), 535–541.

[3] T. Blick, E. Widodo, H. Hugo, et al., Epithelial mesenchymal transition traits in human breast cancer cell lines, *Clin Exp Metastasis* **25**(6) (2008), 629–642.

[4] P. Friedl and K. Wolf, Tube travel: the role of proteases in individual and collective cancer cell invasion, *Cancer Res* **68**(18) (2008), 7247–7249.

[5] Y. Fernandez, J. Cueva, A.G. Palomo, et al., Novel therapeutic approaches to the treatment of metastatic breast cancer, *Cancer Treat Rev* **36**(1) (2010), 33–42.

[6] D. Miles, N. Harbeck, B. Escudier, et al., Disease course patterns after discontinuation of bevacizumab: pooled analysis of randomized phase III trials, *J Clin Oncol* **29**(1) (2011), 83–88.

[7] A. Thompson, K. Brennan, A. Cox, et al., Evaluation of the current knowledge limitations in breast cancer research: a gap analysis, *Breast Cancer Res* **10**(2) (2008), R26.

[8] P. Friedl and K. Wolf, Plasticity of cell migration: a multiscale tuning model, *J Cell Biol* **188**(1) (2010), 11–19.

[9] O. Ilina and P. Friedl, Mechanisms of collective cell migration at a glance, *J Cell Sci* **122**(18) (2009), 3203–3208.

[10] P. Friedl and D. Gilmour, Collective cell migration in morphogenesis, regeneration and cancer, *Nat Rev Mol Cell Biol* **10** (2009), 445–457.

[11] Y. Shimao, K. Nabeshima, T. Inoue, et al., Complex formation of IQGAP1 with E-cadherin/catenin during cohort migration of carcinoma cells. Its possible association with localized release from cell-cell adhesion, *Virchows Arch* **441**(2) (2002), 124–132.

[12] A. Wicki and G. Christofori, The potential role of podoplanin in tumour invasion, *Br J Cancer* **96**(1) (2007), 1–5.

[13] S. Giampieri, C. Manning, S. Hooper, et al., Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility, *Nat Cell Biol* **11**(11) (2009), 1287–1296.

[14] K. Wolf and P. Friedl, Mapping proteolytic cancer cell-extracellular matrix interfaces, *Clin Exp Metastasis* **26**(4) (2009), 289–298.

[15] S. Giampieri, S. Pinner and E. Sahai, Intravital imaging illuminates transforming growth factor beta signaling switches during metastasis, *Cancer Res* **70**(9) (2010), 3435–3439.

[16] M. Yilmaz and G. Christofori, Mechanisms of motility in metastasizing cells, *Mol Cancer Res* **8**(5) (2010), 629–642.

[17] G. Arpino, V.J. Bardou, G.M. Clark, et al., Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome, *Breast Cancer Res* **6**(3) (2004), R149–R156.

[18] D. Sarrio, S.M. Rodriguez-Pinilla, D. Hardisson, et al., Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype, *Cancer Res* **68**(4) (2008), 989–997.

[19] K. Wolf, Y.I. Wu, Y. Liu, et al., Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion, *Nat Cell Biol* **9**(8) (2007), 893–904.

[20] I.R. Macpherson, S. Hooper, A. Serrels, et al., p120-catenin is required for the collective invasion of squamous cell carcinoma cells via a phosphorylation-independent mechanism, *Oncogene* **26**(36) (2007), 5214–5228.

[21] C. Gaggioli, S. Hooper, C. Hidalgo-Carcedo, et al., Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells, *Nat Cell Biol* **9**(12) (2007), 1392–1400.

[22] A. Wicki, F. Lehembre, N. Wick, et al., Tumor invasion in the absence of epithelial-mesenchymal transition: podoplanin-mediated remodeling of the actin cytoskeleton, *Cancer Cell* **9**(4) (2006), 261–272.

[23] O. Gevaert, F. De Smet, D. Timmerman, et al., Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks, *Bioinformatics* **22**(14) (2006), e184–e190.

[24] U. Guha, R. Chaerkady, A. Marimuthu, et al., Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS, *Proc Natl Acad Sci U S A* **105**(37) (2008), 14112–14117.

[25] S.Y. Kim, S. Imoto and S. Miyano, Inferring gene networks from time series microarray data using dynamic Bayesian networks, *Brief Bioinform* **4**(3) (2003), 228–235.

[26] E.E. Evans, A.D. Henn, A. Jonason, et al., C35 (C17orf37) is a novel tumor biomarker abundantly expressed in breast cancer, *Mol Cancer Ther* **5**(11) (2006), 2919–2930.

[27] E. Katz, S. Dubois-Marshall, A.H. Sims, et al., A gene on the HER2 amplicon, C35, is an oncogene in breast cancer whose actions are prevented by inhibition of Syk, *Br J Cancer* **103**(3) (2010), 401–410.

[28] R.L. Camp, G.G. Chung and D.L. Rimm, Automated subcellular localization and quantification of protein expression in tissue microarrays, *Nat Med* **8**(11) (2002), 1323–1327.

[29] M. Baatz, J. Zimmermann and C.G. Blackmore, Automated analysis and detailed quantification of biomedical images using definiens cognition network technology (R), *Comb Chem High Throughput Screen* **12**(9) (2009), 908–916.

[30] M. Athelogou, G. Schmidt, A. Schäpe, et al., Cognition network technology – a novel multimodal image analysis technique for automatic identification and quantification of biological image contents, in: *Imaging Cellular and Molecular Biological Functions*, S.L. Shorte and F. Frischknecht, eds, Springer, 2007, pp. 407–422.

[31] N. Friedman, Inferring cellular networks using probabilistic graphical models, *Science* **303**(5659) (2004), 799–805.

[32] J. Yu, V.A. Smith, P.P. Wang, et al., Advances to Bayesian network inference for generating causal networks from observational biological data, *Bioinformatics* **20**(18) (2004), 3594–3603.

[33] V.A. Smith, J. Yu, T.V. Smulders, et al., Computational inference of neural information flow networks, *PLoS Comput Biol* **2**(11) (2006), e161.

[34] J. Yu, Developing Bayesian network inference algorithms to predict causal functional pathways in biological systems, *Duke*, 2005.

[35] E. Katz, S. Dubois-Marshall, A.H. Sims, P. Gautier, H. Caldwell, R.R. Meehan and D.J. Harrison, An *in vitro* model that recapitulates the epithelial to mesenchymal transition (EMT) in human breast cancer, *PLoS ONE* (2011) 6:e17083.

[36] P. Auguste, L. Fallavollita, N. Wang, et al., The host inflammatory response promotes liver metastasis by increasing tumor cell arrest and extravasation, *Am J Pathol* **170**(5) (2007), 1781–1792.

[37] C.M. Nelson, D. Khauv, M.J. Bissell, et al., Change in cell shape is required for matrix metalloproteinase-induced epithelial-mesenchymal transition of mammary epithelial cells, *J Cell Biochem* **105**(1) (2008), 25–33.

[38] M. Yilmaz and G. Christofori, EMT, the cytoskeleton, and cancer cell invasion, *Cancer Metastasis Rev* **28**(1–2) (2009), 15–33.

[39] X. Wang, L. He, Y.I. Wu, et al., Light-mediated activation reveals a key role for Rac in collective guidance of cell movement *in vivo*, *Nat Cell Biol* **12**(6) (2010), 591–597.

[40] M. Schoumacher, R.D. Goldman, D. Louvard, et al., Actin, microtubules, and vimentin intermediate filaments cooperate for elongation of invadopodia, *J Cell Biol* **189**(3) (2010), 541–556.

[41] A. Sottoriva, J.J. Verhoeff, T. Borovski, et al., Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity, *Cancer Res* **70**(1) (2010), 46–56.

[42] H. Enderling, L. Hlatky and P. Hahnfeldt, Migration rules: tumours are conglomerates of self-metastases, *Br J Cancer* **100**(12) (2009), 1917–1925.

[43] H. Enderling, A.R. Anderson, M.A. Chaplain, et al., Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics, *Cancer Res* **69**(22) (2009), 8814–8821.

[44] A. Rizki, V.M. Weaver, S.Y. Lee, et al., A human breast cell model of preinvasive to invasive transition, *Cancer Res* **68**(5) (2008), 1378–1387.

[45] P.A. Kenny, G.Y. Lee, C.A. Myers, et al., The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression, *Molecular Oncology* **1**(1) (2007), 84–96.

[46] P. Timpson, E.J. McGhee, J.P. Morton, et al., Spatial regulation of rhoa activity during pancreatic cancer cell invasion driven by mutant p53, *Cancer Res* **71**(3) (2011), 747–757.

[47] K.J. Painter, N.J. Armstrong and J.A. Sherratt, The impact of adhesion on cellular invasion processes in cancer and development, *J Theor Biol* **264**(3) (2010), 1057–1067.

[48] G. Caldieri and R. Buccione, Aiming for invadopodia: organizing polarized delivery at sites of invasion, *Trends Cell Biol* **20**(2) (2010), 64–70.

[49] R. Poincloux, F. Lizarraga and P. Chavrier, Matrix invasion by tumour cells: a focus on MT1-MMP trafficking to invadopodia, *J Cell Sci* **122**(Pt 17) (2009), 3015–3024.

[50] C. Hidalgo-Carcedo, S. Hooper, S.I. Chaudhry, et al., Collective cell migration requires suppression of actomyosin at cell-cell contacts mediated by DDR1 and the cell polarity regulators Par3 and Par6, *Nat Cell Biol* **13**(1) (2011), 49–58.

[51] S. Pinner, P. Jordan, K. Sharrock, et al., Intravital imaging reveals transient changes in pigment production and Brn2 expression during metastatic melanoma dissemination, *Cancer Res* **69**(20) (2009), 7969–7977.