

## Single-cell analysis of EphA clustering phenotypes to probe cancer cell heterogeneity

Andrea Ravasio<sup>1,2</sup>, Myint Z. Myaing<sup>1</sup>, Shumei Chia<sup>3</sup>, Aditya Arora<sup>1</sup>, Aneesh Sathe<sup>1</sup>, Elaine Yiqun Cao<sup>3</sup>, Cristina Bertocchi <sup>1,2</sup>, Ankur Sharma <sup>3</sup>, Bakya Arasi<sup>1</sup>, Vin Yee Chung <sup>1</sup>, Adrienne C. Green<sup>4</sup>, Tuan Zea Tan <sup>5</sup>, Zhongwen Chen <sup>1</sup>, Hui Ting Ong<sup>1</sup>, N. Gopalakrishna Iyer<sup>4</sup>, Ruby YunJu Huang<sup>6,7</sup>, Ramanuj DasGupta <sup>3</sup>, Jay T. Groves<sup>1,8</sup> & Virgile Viasnoff <sup>1,9</sup>✉

The Eph family of receptor tyrosine kinases is crucial for assembly and maintenance of healthy tissues. Dysfunction in Eph signaling is causally associated with cancer progression. In breast cancer cells, dysregulated Eph signaling has been linked to alterations in receptor clustering abilities. Here, we implemented a single-cell assay and a scoring scheme to systematically probe the spatial organization of activated EphA receptors in multiple carcinoma cells. We show that cancer cells retain EphA clustering phenotype over several generations, and the degree of clustering reported for migration potential both at population and single-cell levels. Finally, using patient-derived cancer lines, we probed the evolution of EphA signalling in cell populations that underwent metastatic transformation and acquisition of drug resistance. Taken together, our scalable approach provides a reliable scoring scheme for EphA clustering that is consistent over multiple carcinomas and can assay heterogeneity of cancer cell populations in a cost- and time-effective manner.

<sup>1</sup>National University of Singapore, Singapore, Singapore. <sup>2</sup>Pontificia Universidad Católica de Chile, Santiago, Chile. <sup>3</sup>Genome Institute of Singapore, ASTAR, Singapore, Singapore. <sup>4</sup>National Cancer Centre Singapore, Singapore, Singapore. <sup>5</sup>National University Health System, Singapore, Singapore. <sup>6</sup>School of Medicine & Graduate Institute of Oncology, College of Medicine, National Taiwan, Taiwan. <sup>7</sup>University Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. <sup>8</sup>University of California, Berkeley, CA, USA. <sup>9</sup>Centre National de la Recherche Scientifique, Singapore, Singapore. ✉email: [dbsvnr@nus.edu.sg](mailto:dbsvnr@nus.edu.sg)

Receptor Tyrosine Kinases (RTK) are critical for tissue homeostasis<sup>1–3</sup> and their dysfunction is associated with cancer progression<sup>4,5</sup>. Eph receptors, which comprise the largest family of RTK, interact with ephrin ligands expressed on the surfaces of neighboring cells. In particular, an increase in EphA activity has been implicated in many cancer types<sup>6</sup> including 40% of breast cancers<sup>7,8</sup>. This juxtacrine receptor–ligand signaling creates significant opportunities for spatial and mechanical constraints at the cell–cell interface to become intermingled with signaling behavior<sup>9–11</sup>. Recent studies on breast cancer cell lines reported that cytoskeleton remodeling and EphA2 signaling interact in a feedback loop<sup>12</sup> to control the clustering of EphA2, which specifically binds to ephrinA1 ligands<sup>13–16</sup>. Alterations in this feedback loop result in changes in receptor aggregate morphologies, which, in turn, lead to phenotypic changes in cell morphologies and behaviors<sup>17–19</sup>. Seminal reports strongly suggest that the activities of EphA2, and of other Eph receptors<sup>20,21</sup>, are regulated by their ability to spatially cluster in a cytoskeleton-dependent way. The cluster morphologies correlate with the invasion potentials of breast cancer cell lines at the population level<sup>13,15</sup>. Therefore, we hypothesized that developing a generic scoring scheme for individual cells could provide a robust phenotypic assay to ascertain the heterogeneity of cellular functional states within a cancer cell population, which could be compared across cancers of different origin. In particular, it could probe intra-tumor functional heterogeneity of EphA signaling in a cost- and time-effective manner. Such assays could also complement single-cell transcriptomic approaches by integrating gene expression profiles with essential information about cell behavior and cell-state in an unbiased way at the single-cell level.

In this paper, we developed a simple phenotypic assay based on a general scoring scheme to quantify EphA clustering induced by ephrinA1 ligation that we standardize amongst a variety of human cancer cell lines of different origins. We demonstrated that cluster morphology is an inheritable trait that propagates over cell divisions for several generations, and that it correlates with the migratory potential of single cells. Finally, we validated that our assay predicts alterations in EphA activation pathway in patient-derived cells (PDCs) from primary, metastatic, and drug-resistant tumor cell populations.

## Results

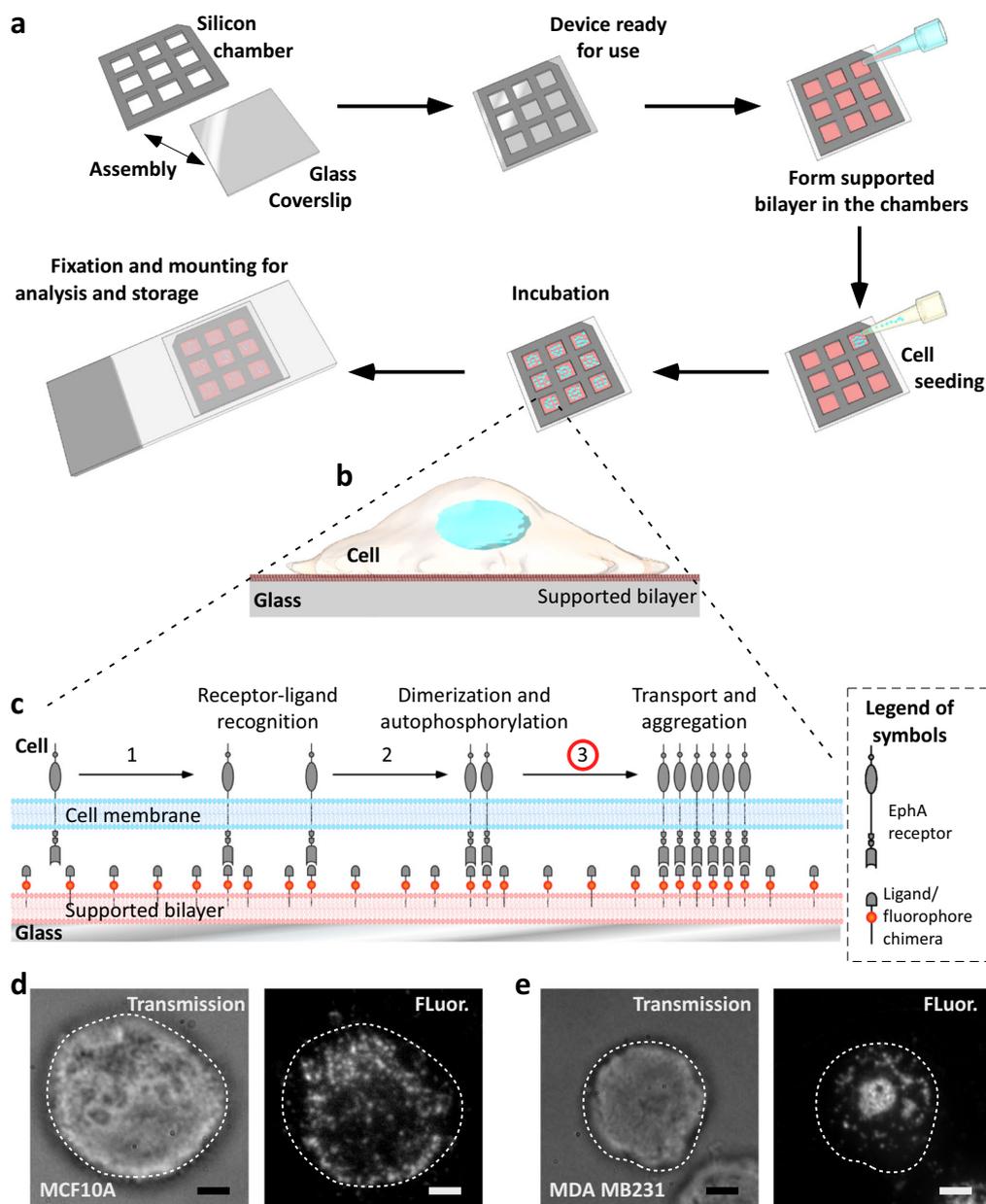
**Working principles and work flow of the assay.** In this assay, we used supported lipid bilayers functionalized with fluorescently-tagged (Alexa 568) ephrinA1 to allow for cytoskeleton-dependent clustering of EphA receptors. Upon ligand binding, the receptors oligomerize and get trans-phosphorylated. The mobile ligand/receptor pairs then aggregate into clusters driven by cytoskeleton rearrangements<sup>15</sup>. Using single-cell image analysis, we quantified cluster morphologies based on scalar scores that reflected their degree of aggregation. The distribution of individual scores provided information about intra-tumor heterogeneity. Figure 1a summarizes the experimental scheme that is detailed in “Methods” section. In brief, a silicon gasket (20 × 20 × 0.3 mm), comprising of nine independent recording chambers (3 × 3 mm), was layered on a clean, hydrophilic glass coverslip (Supplementary Fig. 1a). We generated supported bilayers (96 mol% DOPC and 4% DOGS-NTA-Ni) by standard small unilamellar vesicle (SUV) deposition (“Methods” section) in millimeter-sized chambers. Polyhistidine–ephrinA1 labeled with Alexa 568 were conjugated with the NTA lipid moieties at saturating densities (4%). We optimized the size of the chambers and the spatial arrangement of the duplicates to achieve high reproducibility and to allow

internal quality controls (Supplementary Fig. 1b, c). Following sedimentation, suspended cells (~100 cells per chamber) establish contact with the bilayer within 30–60 s (Fig. 1b). Upon establishment of contact, EphA receptors on the cell surfaces cluster and engage with the ligands freely diffusing on the bilayer (Fig. 1c). For all the cell types we tested, receptor aggregates reached a steady morphology within a few minutes to half an hour. Thereafter, these structures remained largely unaltered for up to 1 h. Our samples were fixed after 1 h to capture statistically significant differences in EphA clustering morphologies at steady state. Individual imaging of 150–200 cells typically took an hour. In this study, a manual implementation of the assay proved sufficient. Optimizing automated imaging and analysis of larger cell populations allowed by the scalable format of the assay (Fig. 1a–c) was left for further development.

**EphA clustering in cancer cells of different origins.** We first tested whether cells originating from different cancer types displayed variations in EphA clustering, as seen in breast cancer cell lines<sup>15</sup> (Fig. 1d, e). Figure 2a displays the cluster morphologies obtained for seven ovarian (PEO1, SKOV3, OVCA420, OVCA429, A2780, HeyA8, and OVCAR10), one lung (A549), one gastric (MKN28), and three breast cancer cell lines (MCF10A, MCF7, MDA MB231). Reflection interference contrast microscopy images confirmed that the EphA clusters localized in regions where cells established intimate contact with the bilayer (Fig. 2b).

Interestingly, the steady state variations in cluster morphologies between cell types resulted from differences in receptor–ligand radial transport (Fig. 2b–d and Supplementary Movies 1 and 2). Some cell types displayed small aggregates with very little cell-to-cell variability, whereas other cell types presented larger and nonhomogeneous cluster morphologies. For example, PEO1 cells presented small EphA clusters with limited mobility over 30 min of observation (Fig. 2c, left and kymograph in Fig. 2d, top). In contrast, EphA puncta in HeyA8 cells were quickly transported toward the inner regions of the cells (Fig. 2c, right), where they fused into aggregates. These aggregates further merged with other aggregates (kymograph in Fig. 2d, bottom) to form a series of larger clusters (Fig. 2b). In these experiments, we have also determined that for all cell types steady state was reached at 30 min after cell/bilayer contact and clustering morphologies remained unchanged from that point onwards. Thus, we performed all the following EphA experiments by fixing the cells at 60 min.

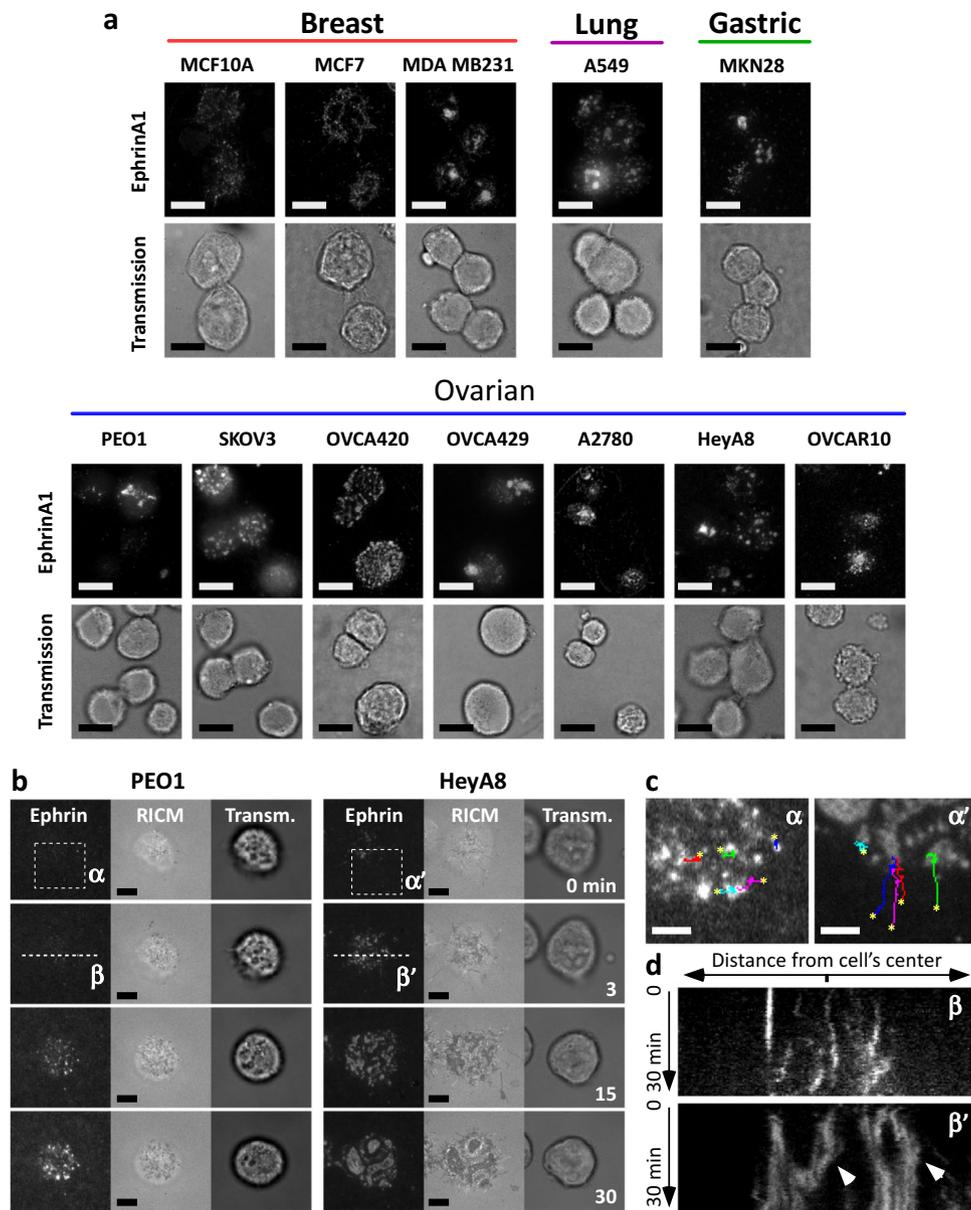
**Scoring cluster morphologies.** We hypothesized that some morphological features of the EphA clusters could reflect a generic cell response to ephrinA1 binding that was independent of cancer type, whereas a different set of morphological features would be cancer-type specific. We thus probed 75 morphological descriptors of the cluster patterns for their specificity, using a combination of t-distributed stochastic neighbor embedding (t-SNE) algorithm and a machine-learning method (Random Forests) (see Supplementary Note, Supplementary Tables 1 and 2 and Supplementary Figs. 2 and 3). This unbiased study singled out the intensive descriptor  $S_{\text{EphA}} = \int \rho \ln \rho$ , where  $\rho = \frac{I}{I_{\text{mean}}}$ , is the ratio of the fluorescence intensity value  $I$  of each pixel divided by the average fluorescence intensity  $I_{\text{mean}}$  under each cell (Supplementary Fig. 4).  $S_{\text{EphA}}$  emerged as the best parameter to score the scattering degree of EphA clusters while ignoring their specific spatial arrangements that mostly reflected the cancer cell type of origin (Supplementary Fig. 4 for justifications). Morphologies with scattered small puncta resulted in low  $S_{\text{EphA}}$



**Fig. 1** Illustration of multi-well ephrin clustering assay. **a** Schematic representation of the experimental procedure. A silicon gasket comprising of nine chambers is placed on a clean coverslip. The coverslip is biofunctionalized with small unilamellar vesicles to form a supported lipid bilayer. The bilayer presents ephrinA1 ligands. Cells are seeded in each chamber, left for sedimentation, and incubation for an hour. The system is fixed and mounted on a coverslip for subsequent imaging. **b, c** Cartoons illustrating the EphA clustering assay. By sedimentation, cells contact the supported bilayer functionalized with fluorescent chimera of the ephrinA1 ligand. Upon receptor/ligand recognition (step 1), active receptors dimerize and phosphorylate (step 2), initiating the signaling cascade. In response to the initial signaling, active transport of receptors created signaling clusters. Their aggregation is imaged using ephrinA1 tagged with Alexa 568. **d** Examples of images of noninvasive breast cancer cells (left, MCF10A) and an (**e**) invasive breast cancer cells (right, MDA MB231) presenting different cluster morphologies (Scale bar: 5  $\mu\text{m}$ ). Scoring of individual cells based on quantification of the cluster morphologies is used as a proxy to define the heterogeneity of cellular states within a population.

scores while morphologies with large aggregates resulted in higher scores. Figure 3a shows the distribution of single-cell scores obtained for the epithelial-type PEO1 carcinoma cells and the mesenchymal-type HeyA8 cells. We linked individual scores to the corresponding cluster images. Whereas PEO1 cells showed homogeneously limited clustering, HeyA8 cells displayed a high variability in cluster sizes, which correlated with high variability in  $S_{\text{EphA}}$  scores. In this heterogeneous population, cells with scattered small clusters obtained low  $S_{\text{EphA}}$  values (consistently with PEO1) and cells with large aggregated clusters obtained high  $S_{\text{EphA}}$  values.

**Quantitative assessment of  $S_{\text{EphA}}$  distribution.** We first tested the reproducibility of the determination of the  $S_{\text{EphA}}$  score distribution over several cell types. Repeated measurements for the same cell population on different chips resulted in qualitatively identical salient features of the distributions (narrow vs. spread distribution). However, inter-chip variability limited the reproducibility of quantitative data. We noticed that with every chip, the experimental variations for all the tested cell lines were correlated. This suggested that quantitative changes arose from day-to-day fluctuations in the detection system rather than from cell-intrinsic noise. We alleviated these variations by implementing an

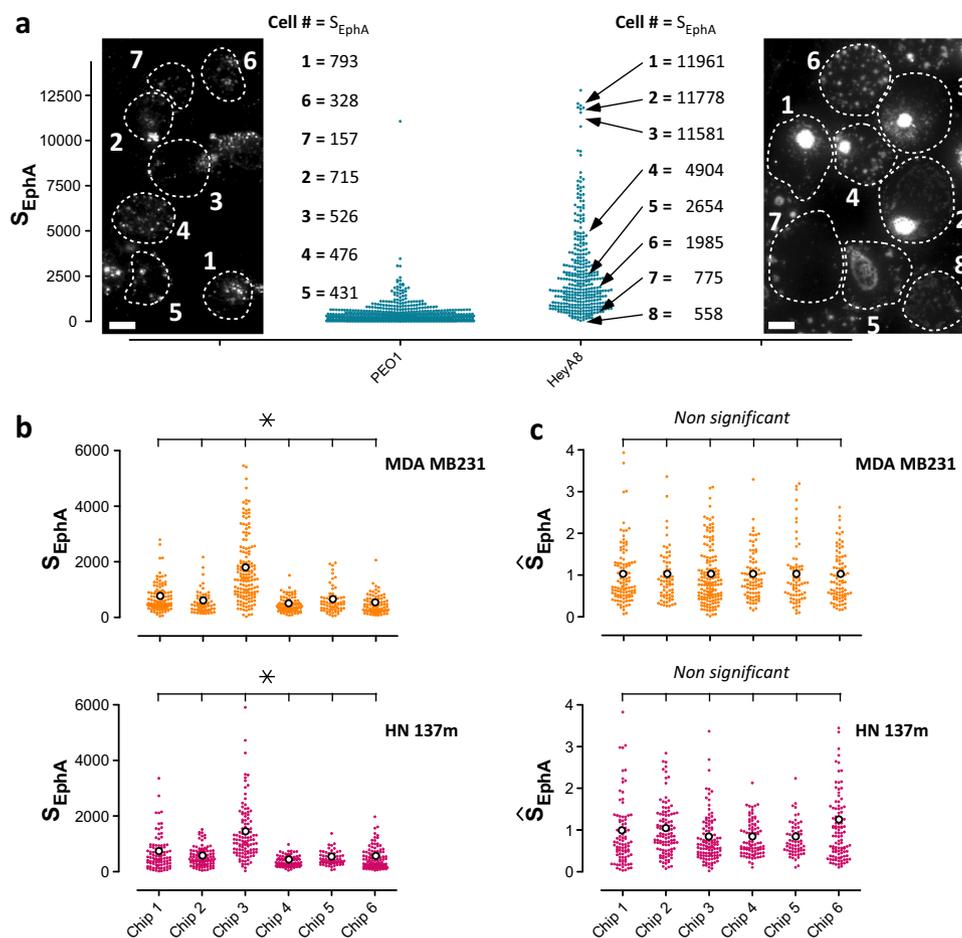


**Fig. 2** Cancer cells from various origins show morphological variability in receptor clustering. **a** Various Eph clustering morphologies obtained across a library of cancer cell lines from various cancer origins (Scale bar: 20  $\mu$ m). **b** Time-lapse imaging of the clustering process in PEO1 and HeyA8 cells: noninvasive and epithelial-like cancer cells (PEO1) showed limited ability to form clusters as compared with the more invasive and mesenchymal-like cancer cells (HeyA8) (Scale bar: 5  $\mu$ m). **c** Enlarged view from boxes ( $\alpha$ ) and ( $\alpha'$ ) in **b**. Trajectories of fluorescent puncta are shown by the colored lines. Initial location of the puncta is indicated by an asterisk (Scale bar: 2.5  $\mu$ m). **d** Time evolution (kymograph) of fluorescence along dashed lines ( $\beta$ ) and ( $\beta'$ ) in **b**. Whereas Eph clusters remained scattered in PEO1, further aggregation and merging of large clusters is observed in HeyA8 cells (arrow heads).

intrinsic normalization scheme. On every chip, we measured triplicates of MDA MB231 cells that were used as a reference cell line (Supplementary Fig. 1b). Supplementary Fig. 1c shows an example of the intra-chip and inter-chip variability for each triplicate in three independent chips. As a quality control measure for device preparation, we considered only those individual chips for which the MDA MB231 triplicates showed no significant differences ( $p > 0.05$  using a Kruskal–Wallis nonparametric test with Dunn’s multiple comparison applicable to non-Gaussian distributions). When this criterion was not met (in less than 20% of chips), the entire chip was excluded from analysis. For each selected chip, we then combined the distribution obtained for the MDA MB231 triplicates into a single distribution and used its mean value to normalize all the distributions measured on the entire chip. Figure 3b illustrates the combined distribution of

triplicates of MDA MB231 cells (breast cancer) and HN137m cells (patient-derived squamous head-and-neck cancer cell) run over six different chips. Pairwise comparison demonstrates the inter-chip variability observed experimentally. Due to the proportionality of the pairwise inter-chip variability, we could normalize all the scores for a given chip by the averaged value of the MDA MB231 distribution ( $N > 150$  cells). Normalized values are represented as  $\hat{S}_{EphA}$ , where  $\hat{S}_{EphA} = \frac{S_{EphA}}{S_{EphA,MDA,MB231}}$ . The use of a reference cell line is a simple strategy that provides dual advantage: a method for controlling chip quality and a scheme to reduce inter-chip variability below 5% (Fig. 3c).

**Heterogeneity of clustering within diverse cell populations.** We then used the above approach to quantify the distribution of



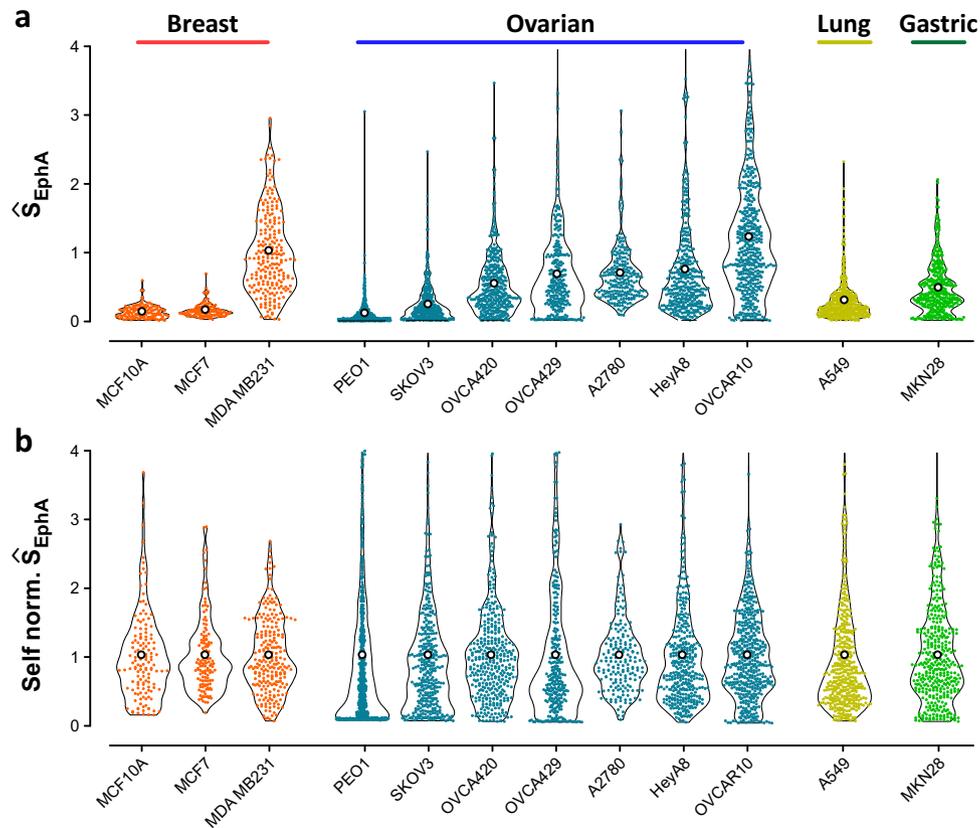
**Fig. 3 Reproducible measurement of  $S_{\text{EphA}}$  scores.** **a** Distributions of  $S_{\text{EphA}}$  scores for PEO1 and HeyA8 lines. Every PEO1 cell displays very scattered puncta. It is reflected by the low average  $S_{\text{EphA}}$  score for the population and a distribution with a small spread. HeyA8 population presents a large intercellular variability in clustering morphologies that is reflected in the high population score and a distribution with a large spread. The matching between clustering morphologies and  $S_{\text{EphA}}$  scores is exemplified for PEO1 (seven cells) and HeyA8 (eight cells) lines (Scale bar: 5  $\mu\text{m}$ ). **b** The inter-chip (six chips) variability of the score distribution for MDA MB231 (top) and HN137m (bottom) cells shows significant differences between replicates. **c** Normalization of all distributions on individual chips by the average scores of triplicates of the reference cell line MDA MB231 abolishes significant inter-chip variability. It allows quantitative comparison between chips and cell types. Each distribution is based on  $N > 200$  single-cell analysis. We established significance using the two-sample Kolmogorov–Smirnov test that probes for changes in the shape of the distribution.

single-cell  $\hat{S}_{\text{EphA}}$  scores for the 12 cancer cell lines of breast, ovarian, lung and gastric origins (Fig. 4a). Independent of the cancer type, epithelial-like cell displayed lower, narrowly distributed  $\hat{S}_{\text{EphA}}$  scores, whereas more mesenchymal-like cell showed higher, widely-distributed  $\hat{S}_{\text{EphA}}$  scores. Independent of their mean values, the shape of the distributions differed between the cell lines (Fig. 4b). The combinatorial statistical differences between different populations are presented in Supplementary Table 3. While PEO1 cells predominantly displayed very low scores with a relatively long tail toward higher scores, A549 cells and MDA MB231 cells mostly displayed a symmetric distribution around their average scores. Kolmogorov–Smirnov tests revealed that considering distributions shapes together with their average values provided characteristics of each cell population with 90% accuracy (Supplementary Table 3). Taken together, the results shown in Figs. 3a and 4b suggest that the shapes of  $\hat{S}_{\text{EphA}}$  distribution, in addition to average scores, could be used to predict single-cell heterogeneity in each population, rather than being regarded as measurement noise.

**EphA clustering ability is an inheritable trait.** So far, we established that the distribution of EphA clustering could

characterize the cell population. We then tested whether individual cell response was an intrinsic property of individual cells that could be retained even upon cell division. To this end, we adopted the following strategy. We used patient-derived primary cultures from head-and-neck tumor (specifically HN137p)<sup>22</sup>. We isolated and expanded 30 individual cells into clonal colonies for 10 days, following which we tested half of the cells from every colony with our assay. The other half was left to expand for another 10 days, and cells from every colony were tested again after 20 days. We then compared the distribution of scores for the original population (before cell isolation) with the distribution after 10 and 20 days for each clonal expansion. Nine colonies survived this screening process (Fig. 5a).

During the first 10 days (seven to nine cell divisions), we monitored the expansion of clonal colonies using phase contrast microscopy. Some colonies displayed very homogeneous morphologies (e.g., colony 1) with well-defined cell–cell contacts and smooth edges, a typical characteristic of epithelial-like cells. Other colonies, however, appeared much more heterogeneous (e.g., colony 9) with large variations in their individual cell phenotypes, a lower degree of contact inhibition, and more diffuse edges (Fig. 5b).



**Fig. 4** Measuring  $\hat{S}_{Epha}$  distributions across cell types. **a**  $\hat{S}_{Epha}$  distributions for a panel of breast, ovarian, lung and gastric cancer cell lines. All cell types displayed cluster morphologies that could be quantitatively assessed in our assay, leading to significantly different average value ( $N > 200$  for each line). **b** When the scores were normed by their own average values, they were still differently distributed around 1. For the sake of clarity, the combinatorial assessment of statistical relevance between all distributions is presented in Supplementary Table 3.

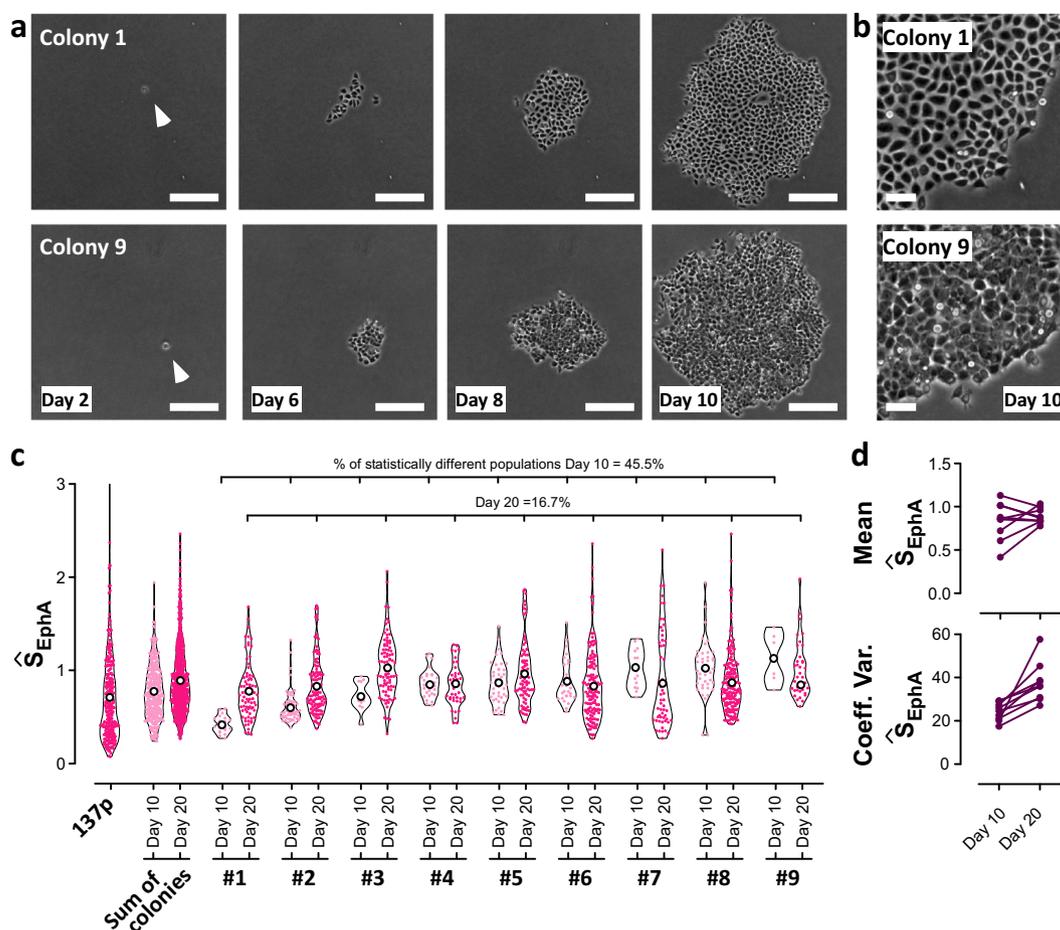
Figure 5c compares the evolution of EphA scores for the unsorted population, and for the nine colonies at 10 and 20 days. After 10 days, colonies with homogeneous epithelial morphologies (e.g., colony 1 or 2) displayed  $\hat{S}_{Epha}$  scores significantly lower than their unsorted population average. Conversely,  $\hat{S}_{Epha}$  scores higher than the population average characterized the colonies with more irregular appearances (e.g., colonies 8 or 9). At day 10, each colony displayed a different average  $\hat{S}_{Epha}$  score (Fig. 5d, top). However, intra-colony variability was small as shown by the narrowly distributed scores around their respective average values (Fig. 5d, bottom). After 20 days of culture, the colonies evolved toward similar  $\hat{S}_{Epha}$  averages (Fig. 5d, top), which converged toward the average value measured for the entire population. Concomitantly, intra-colony variations increased substantially as shown by the larger coefficients of variation of  $\hat{S}_{Epha}$  distributions (Fig. 5d, bottom). The growth of clonal colonies limits the number of cells available for this analysis. However, confidence analysis (Supplementary Note Information and Supplementary Fig. 5) demonstrates that the distribution of scores obtained for each colony cannot be interpreted (with 99.5% confidence) as a random subset of the unsorted distribution, instead they must be attributed to the phenotypes of the initial clones. In addition, after 10 days of growth, 45.5% of the colonies significantly differed from each other (Kruskal–Wallis with Dunns posttest). This ratio dropped to 16.7% at 20 days demonstrating the “reconstitution” of the original population diversity from each clonal expansion. Of note, reconstituted populations after 10 and 20 days lacked cells with very low  $\hat{S}_{Epha}$  scores (below 0.2). Instead, an increasing fraction of cells presented  $\hat{S}_{Epha}$  larger than 0.6. This observation was consistent with the idea that expanding clonal colonies did

not favor the growth of the most epithelial-like cellular types (only nine cells could be expanded out of the initial 30 cells, with most of them being quiescent or apoptotic). During proliferation, most epithelial phenotypes were not reconstituted, and cells appeared to drift toward the more mesenchymal cell types.

Altogether our results, suggest that EphA cluster morphology of a single cell is an inheritable trait that gradually diverges upon sequential divisions to progressively regenerate the initial population diversity.

**$\hat{S}_{Epha}$  correlates with cell migration potential.** In breast cancer cells, different EphA activation levels were associated with different invasion potentials at the population level<sup>15</sup>. This correlation was attributed to pathways modulating the actin cytoskeleton, that could induce EphA clustering as well as promote cell migration. We therefore tested how phenotypic characterizations of EphA clustering overlapped with other standard phenotypic assay, including the migratory potential of cancer cells.

Using microchip arrays, we first measured the transcriptional levels of all EphA receptors in all our cell lines. EphA2 proved to be the most abundantly expressed receptor in 8 out of the 11 cell lines (Supplementary Table 4).  $\hat{S}_{Epha}$  scores at the population level (Supplementary Fig. 6) did not correlate with the mRNA levels of EphA2 and also with all EphA pulled together. The correlation greatly improved when EphA2 protein expression levels (Supplementary Fig. 6b, c, Western blot) were considered. We then tested if the cellular phenotypes defined by our assay correlated with cancer hallmarks<sup>23–25</sup>, epithelial mesenchymal

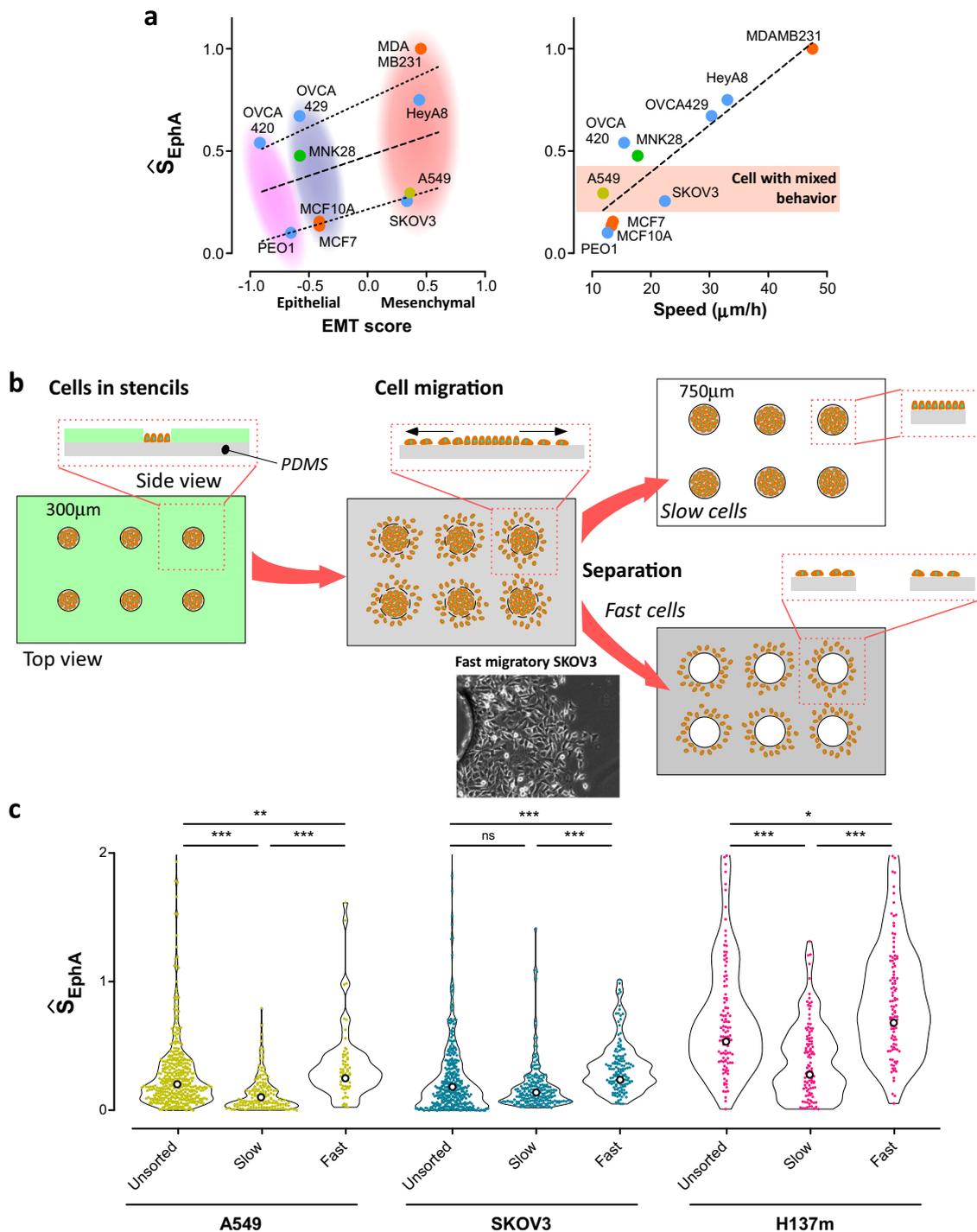


**Fig. 5 The clonal expansion of HN137p colonies demonstrates that EphA clustering is an inherited trait passed over many progenies.** Thirty individual cells were picked from the unsorted population of HN137p cells, out of which nine clonal colonies could grow for 20 days (a). Phase contrast images of the clonal colonies over time. Some colonies display a very homogeneous aspect (e.g., colony 1), some others show a noticeable spread in cell morphologies (e.g., colony 9) (Scale bar: 100  $\mu$ m). b Close-up of colonies 1 and 9 after 20 days (Scale bar: 25  $\mu$ m). c  $\hat{S}_{EphA}$  distributions for the unsorted cell population, for populations reconstituted from the nine colonies at 10 and 20 days, and for each individual colony at both time points. Note the absence of low scores in the reconstituted populations and the significant shift toward larger score between 10 and 20 days. After 10 days, each colony retains a distribution well-centered around its average value, indicative of the conservation of the clustering properties upon cell proliferation. After 20 days though, all colony distributions converge toward the same distribution, indicating the genomic drift of the cellular states over time. d Convergence of the mean score for every colony between day 10 and day 20. The spreads of all distributions increase simultaneously, indicating a larger variability in cell behavior. Statistical analysis of populations (Kruskal-Wallis with Dunn’s posttest) and confidence test (Bootstrapping) are discussed in Supplementary Note.

transition scores (EMT) and their migration potential. Interestingly,  $\hat{S}_{EphA}$  did not correlate with any of the listed cancer hallmarks (Supplementary Table 5). We then computed EMT scores for all cell types according to Tan et al.<sup>26</sup> by measuring the transcription levels of a list of 118 genes. Overall, EMT scores poorly correlated with  $\hat{S}_{EphA}$  (Fig. 6a, left; dash line, Pearson coefficient = 0.34). However, we obtained a much higher correlation by subdividing the cell lines into two groups using  $\hat{S}_{EphA} = 0.4$  as threshold (Fig. 6a, left; dotted lines, Pearson coefficients = 0.83 for group with  $\hat{S}_{EphA} > 0.4$ , and Pearson coefficients = 0.99 for group with  $\hat{S}_{EphA} < 0.4$ ). We then systematically performed wound-model experiments to assess the migration potentials of all cell lines. Interestingly, cell motility speed highly correlated with  $\hat{S}_{EphA}$  (Fig. 6a, right; Pearson coefficient = 0.89). Furthermore, we observed that groups of cell lines with equivalent EMT scores, but different  $\hat{S}_{EphA}$  score (colored ovals in Fig. 6a, left) had as a common discriminator their cell migration potential and cell lines in the group with  $\hat{S}_{EphA} > 0.4$  were fast and  $\hat{S}_{EphA} < 0.4$  were slow migratory. These results suggest that the strong phenotypic link between EphA clustering and migration potential established for breast

cancer cell lines may remain valid for cancer cell lines of different origin.

Some of our cell lines clearly showed heterogeneity in migratory behaviors within the cell population as it often occurs in cancer cells with intermediate EMT score (Fig. 6a, right). We thus explored whether this heterogeneity was also reflected in  $\hat{S}_{EphA}$  distribution patterns. In the three cell lines that we used—A549 (lung), SKOV3 (ovarian), HN137m (head-and-neck PDCs<sup>22</sup>)—migration speeds were substantially heterogeneous, and, correlatively, their  $\hat{S}_{EphA}$  distributions were also scattered. We then used a modified wound-model assay to isolate the fast-migratory cells from the slow ones within each cell population and compared the  $\hat{S}_{EphA}$  distributions between the different subpopulations (fast-migratory cells, slow migratory cells, unsorted populations) (Fig. 6b). We initially confined the cells on 300-micron islands on a PDMS sheet coated with fibronectin (see “Methods” section) and, after removal of the confinement, we allowed migration in 2D for 3 days. We then punched-out the central part of the colonies with a 750-micron punch, thereby sorting cells that had migrated more than 250 microns away from their initial position. In all cell types, fast migrating cells displayed

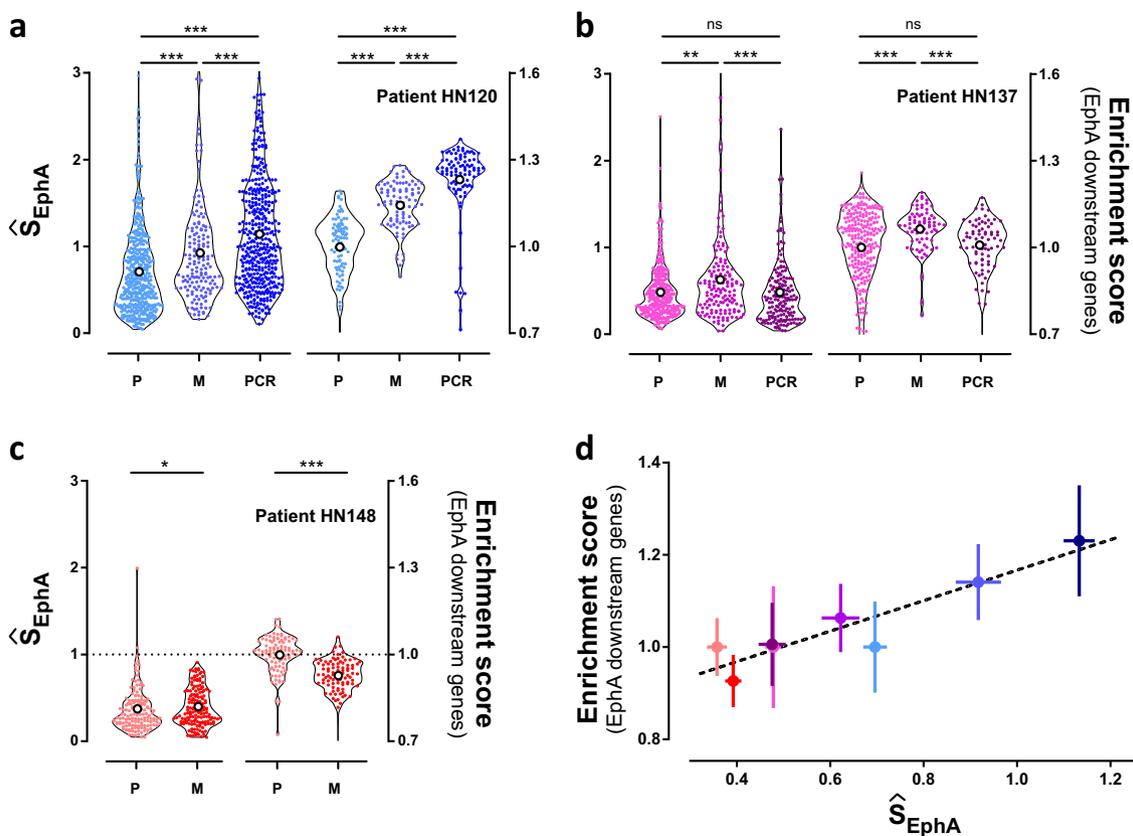


**Fig. 6** Correlation between  $\hat{S}_{EphA}$  scores EMT score and migration speed. **a** Correlation of the average  $\hat{S}_{EphA}$  with the EMT scores (based on 118 genes). The correlation is globally poor, but it largely increases (Pearson coefficient  $\sim 0.85$ ) if cell types are split into slow and fast migrating cells. The correlation between the average  $\hat{S}_{EphA}$  with the mean mobility of cells in the migration front of the gap-closure assay for each population is relatively high (Pearson coefficient  $\sim 0.89$ ). **b** Schematic drawing of our separation assay based on 2D migration. Cells are plated on 300- $\mu\text{m}$  islands obtained by protein printing. The cells are left to migrate away from the initial island. Using a 750- $\mu\text{m}$  punch, we then separated the fast migrating cells that migrated more than 750  $\mu\text{m}$  away from the initial island center, from the rest. **c**  $\hat{S}_{EphA}$  distributions for unsorted, slow and fast populations from A549 (lung), SKOV3 (ovarian), H137m (head-and-neck) cancer cell types. Note that the fast (respectively slow) subpopulations systematically higher (respectively lower)  $\hat{S}_{EphA}$  scores than the unsorted population.  $N > 200$  cells.

a  $\hat{S}_{EphA}$  distribution with a higher average value and higher spread (Fig. 6c).

These results demonstrated the correlation between Epha clustering and migration potential at the population level across diverse cancer cell types. Furthermore, enrichment of cells with

high  $\hat{S}_{EphA}$  in subpopulations selected for their migration potential support the idea that  $\hat{S}_{EphA}$  correlate with single-cell migration potential.  $\hat{S}_{EphA}$  scores hence appear as a fast and reliable measurement of the migration potential of cells and heterogeneity of their invasion potential within a given population.



**Fig. 7 Comparison of single-cell phenotyping by EphA clustering for three different patients.** **a** Left axis: distribution of  $\hat{S}_{\text{EphA}}$  for patient-derived cell lines HN120 from patient 120. HN120P (P) was derived from a primary tumor, HN120M (M) from proximal lymph node, and HN120pcr (PCR) is an induced, cisplatin resistant cell line derived from HN120p. Right axis: transcriptomic analysis (enrichment score) obtained from the single-cell analysis of 71 genes positively correlated with EphA2 activation. Values are normalized by the average value of P. **b, c** Identical studies for patient 137 (HN137) and 148 (HN148). **d** Correlative analysis of EphA clustering ( $\hat{S}_{\text{EphA}}$ ) (mean  $\pm$  SE) and EphA transcriptomic downstream signalling (enrichment score) (mean  $\pm$  SD). Strong correlation (Pearson coefficient = 0.93,  $p = 0.001$ ) between  $\hat{S}_{\text{EphA}}$  and transcriptomic scores demonstrate correspondence between EphA clustering and EphA downstream signaling.

**$\hat{S}_{\text{EphA}}$  as a measure of evolution of disease.** Finally, we tested whether our assay could follow changes in the cellular phenotypes during cancer progression in patient-derived models. We have previously reported the generation of multiple primary cancer cell lines from three different head-and-neck patient tumor samples<sup>22,27</sup>. These include HN137, HN120, and HN148 cell lines, that were generated from the primary site of tumors (HN120p, HN137p, HN148p) or the paired lymph node metastatic sites (HN120m, HN137m, HN148m). In addition, we also generated in vitro acquired-resistance models for cisplatin for those cell lines derived from the primary site (HN137pcr, HN120pcr)<sup>27</sup>. Extensive characterization of these cells demonstrated that they all presented very distinct transcriptomic profiles and migratory phenotypes<sup>22,27</sup>. We therefore tested whether these previously reported characteristics were reflected in the distribution of  $\hat{S}_{\text{EphA}}$  for these distinct eight PDCs.

Figure 7a shows that for patient HN120, the  $\hat{S}_{\text{EphA}}$  distributions from primary, metastatic and cisplatin resistant cell lines displayed increasing average values. This observation correlated with the increased migration potential observed across these cell lines<sup>22,27</sup>. In addition, HN120p and HN120pcr cells displayed very different  $\hat{S}_{\text{EphA}}$  scores, corroborating the previous observation that HN120pcr cells are transcriptomically distinct from HN120p. This conclusion was also supported by comparing the  $\hat{S}_{\text{EphA}}$  scores to the single-cell transcriptomic profiles ( $N > 100$  cells) established based on the expression levels of 71 genes, which have been reported to be downstream of EphA2 signaling

cascade<sup>15</sup> (Supplementary Table 6). The transcriptional scores confirmed that all PDCs lines were distinct. Importantly, EphA2 expression profile of the HN120 lines ranked similar to and correlated with the matched  $\hat{S}_{\text{EphA}}$  scores.

Our results from patient HN137 (Fig. 7b) showed that HN137p and HN137pcr cells did not display any significant differences in  $\hat{S}_{\text{EphA}}$  scores. However, scores in HN137m cells were slightly higher than that observed in HN137p, in line with the observed invasion potentials<sup>22,27</sup>. A similar trend was reflected in the transcriptomic scores. Patient HN148 (Fig. 7c) displayed equal scores in HN148p and HN148m cells, although the transcriptional scores were slightly smaller in HN148m cells than in HN148p cells. Furthermore, average population values of  $\hat{S}_{\text{EphA}}$  scores and transcription score displayed strong correlation (Pearson coefficient = 0.93,  $p = 0.001$ ), thereby indicating that our clustering score provides a reliable measure for EphA associated gene expression profile and phenotypes (Fig. 7d).

Overall, single-cell analysis of EphA clustering behavior and transcriptomic response in PDCs showed that our assay provided a reasonable assessment of the evolution of cell-states within these patient samples in a fast and cost-effective manner.

## Discussion

Intra-tumor heterogeneity poses a major challenge in therapeutic outcomes for cancer treatment in the clinic. Different clonal or sub-clonal cell populations can exhibit differential response to a

given drug, and utilize distinct mechanisms to evolve therapy resistance, such as cell-state change like EMT or dormancy induced quiescence. Therefore, it is important to develop methodologies that can probe this heterogeneity to be able to identify novel genetic and therapeutic vulnerabilities. While existing single-cell genomic/transcriptomic platforms have been shown to be extremely powerful in defining intra-tumor heterogeneity and the identification of new cell types and cell states, methodologies that can uncover functional heterogeneity at the level of single cells remain largely under-explored. In our study, we show that EphA cluster morphologies, which are associated with distinct migratory state of cells, constitute a single-cell-inherited trait of cell populations. We developed a common analysis scheme to quantitatively correlate cluster morphologies with the intrinsic migration potentials of cells. In addition, results from this study extend to multiple cancer types our previous conclusions based on breast cancer cell lines<sup>15</sup> that cluster morphologies are indicative of the degree of transcriptional activation of genes involved in EphA signaling cascade. This signaling pathway largely involves cytoskeleton regulatory genes, which can explain the correlation between cluster morphologies and migration properties. Oppositely the other gene pools that we tested (Supplementary Table 5) hardly involve cytoskeleton components. We thus believe that the absence of correlation with these pathways is additional evidence that our scoring approach targets EphA signaling pathways specifically. By using PDCs from various cancer types, our study reinforces the growing body of evidence<sup>13–16</sup> that clustering properties of juxtacrine receptors can be used to define the “phenotypic state” of single cells. We conclude that this assay provides a technically simple, cost-effective and effective strategy to follow the gradual buildup of intracellular heterogeneity either upon cell division, (Fig. 5) or across genomic drifts and drug selection, (Fig. 7). Such an approach is nonetheless restricted to ligands that foster cell adhesions and spreading on fluid bilayers. If EphB4 ligands have already been tested<sup>28</sup>, the extension to other non-membrane tethered RTK need to be explored. In this case, alternative approaches such as coupling several ligands together, could be devised<sup>29</sup>. The assay can also be performed in a 4 × 4 mm well that are the standard size of the compartments in a 384 well plate. We thus provided proof of concepts that the assay can be scaled up. However, this would entail the parallel formation of lipid bilayer in each well with quality control. We foresee that only automated dispensing can provide sufficient reproducibility in lipid deposition and rinsing to ensure high enough quality bilayer. The reproducibility we achieved here relied on the normalization of each batch data by the scores of MDA MB231 cell line measured on the same day on the same chip. Using a reference line was handy to check the fluidity of the bilayer, to check the quality of the ligand and to perform a normalization of the acquisition conditions. One could argue that a physical calibration of the system (e.g., using fluorescent beads and Fluorescence Recovery After Photobleaching) would alleviate the constraints of using a reference cell line that might differ from lab to labs or drift over time. This process could indeed be implemented on a calibrated microscope equipped with a FRAP module. Such calibration could be performed systematically before each experiment. However, we found no significant drifts in MDA MB231 used no more than 12 passages from the commercial batch. They are also suited for calibration due to the large variability of the distribution of their scores as compared with epithelial cell lines such as PEO1. A systematic comparison of data acquired in different lab using our scheme remains to be done. Furthermore, the procedure to determine cell contours and cluster morphology can already be performed without any user assistance. It could thus be potentially incorporated in a detection pipeline provided that an automated imaging routine is developed.

Lastly, the rationale to develop parallelizable phenotypic assays stems from the observation that the expression levels of genes related to a certain pathway do not always directly correlate with the behavioral response of cells. It is therefore important to develop cost-effective and simple behavioral assays that allow phenotypic probing of response to an activated pathway. We believe that this work is a step forward in this direction.

## Methods

**Preparation of SUV.** A clean, 5 ml round-bottom flask (PYREX) was rinsed with 100% chloroform (Acros Organics) twice. A small amount of 100% chloroform was left in the flask after rinsing. To generate SUVs using the sonication approach, 96 mol% 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC; Avanti Polar Lipids, cat. no. 850375) and 4 mol% 1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt) (DOGS-NTA-Ni; Avanti Polar Lipids) were mixed in chloroform in the flask. A water bath apparatus from a rotary evaporator (EYELA, Japan) was warmed at 50 °C to prevent the solvent from freezing during the evaporation process. The mixture of liposome and NTA-Ni was dried, first by using the rotary evaporator and then under a gentle stream of N<sub>2</sub> for about 1 min, to prevent the exposure of lipid film to air. Dried lipids were resuspended in 1 ml of ultrapure (Milli-Q) water at the room temperature as milky suspensions of multilamellar vesicles. To release physically adsorbed lipids from the walls of the flask, the flask containing the lipid solution was immersed in a warm water bath (60 °C) for 10 s. The lipid suspension was then transferred to a round-bottom polypropylene tube and sonicated for 20 min until the suspension became clear. After sonication, the clear lipid suspension was transferred to a 1.5 ml Eppendorf tube and centrifuged at maximum speed at 4 °C, for at least 4 h. Finally, the supernatant containing SUVs was aliquoted and stored at 4 °C.

**Preparation of cover glass and silicon chamber device.** Microscope cover glasses (22 × 22 mm) from high precision were carefully cleaned to remove impurities by using a series of washing steps. Briefly, cover glasses were washed with dish-washing soap and rinsed with tap water. Thereafter, they were placed in a ceramic rack and sonicated in a sonicator bath (Elma) for 10 min, submerged in 100% acetone (Fisher Chemical). After rinsing with Milli-Q water, they were sonicated for 30 min, submerged in 1:1 isopropanol/Milli-Q water solution (isopropanol from Fisher Chemical). After thorough rinsing with deionized water, the cover glasses were incubated overnight submerged in a 1:1 sulfuric acid/Milli-Q water solution (sulfuric acid from Sigma) to remove all traces of organic solvent. Finally, they were rinsed with Milli-Q water, blow-dried using nitrogen gas, and placed in an 80 °C dry oven for at least 2 h. Meanwhile, a PDMS silicon foil of 0.25 mm thickness (STERNE SAS, France) was cut into square chips of 20 × 20 mm sizes and further subdivided into nine recording chambers (3 × 3 mm) using a cutting plotter (Graphtec Corporation, Japan). Importantly, one of the corners of the square was also cut to break the double symmetry and allow easy identification of wells (see Fig. 1a and Supplementary Fig. 1a, b). As the PDMS foil was polymerized in between two plastic films, the PDMS cuts were free from dust and contaminants. Finally, PDMS cuts were mounted on the glass coverslips and bounded using air plasma for at least 15 min at high power in a plasma cleaner (Harric Plasma). This step also provided a final cleaning of the surfaces and activation of the glass to facilitate adsorption of the lipids bilayer as described in the next paragraph.

**Preparation of ephrinA1 functionalized supported bilayer.** The clean coverslip was placed on the lid of a 35 mm petri dish with the PDMS gasket facing up. Four microliters of SUV TBS solution (1:1 ratio of SUV and 2× TBS solution, TBS form Sigma) was placed in each hole of the device. Upon contacting the activated surface of the plasma-treated cover glass, SUV spontaneously adsorb and formed a supported bi/multi-layer. From this point onwards, care must be taken to avoid direct contact of the supported bilayer with any air–liquid interface. After incubation for 5 min at the room temperature, the device was submerged in a large beaker containing Milli-Q water and it was washed by vigorous sideways shaking to remove excess phospholipids from the bilayer. Thereafter, the device was placed in a 35 mm petri dish containing 1 ml of Milli-Q water. One milliliter of 1% BSA (Sigma) dissolved in 2× TBS solution was then added to the dish to block the unspecific binding of cells onto the glass or the PDMS gasket (final concentration of BSA about 0.5%). The supported lipid membrane on the device was incubated in the blocking solution on a rotary shaker (gentle rotation) for 30 min at the room temperature. After incubation, the device was washed three times with 1× TBS solution, and was then left submerged in 2 ml of 1× TBS solution. Meanwhile, ephrinA1 protein was purified and conjugated with Alexa 568 with NHS ester according to the protocol we developed in<sup>15</sup>. It was dissolved in 1× TBS solution at a dilution ratio of 1:2000. Hundred microliters of protein solution was then added to the device in the petri dish. The device was then incubated with protein solution for 1 h on a rotary shaker (gentle rotation) at the room temperature, in the dark. Conjugation of protein to the supported bilayer occurs during this step. After incubation, excess protein solution on the membrane was washed using imaging buffer (150 mM NaCl, 30 mM KCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 10 mM Hepes, 10

mM D-Glucose, pH 7.4). The device was then placed in a petri dish and submerged in 2 ml of imaging buffer. Devices were freshly prepared and long-term storage was avoided. Devices were kept covered in the dark at the room temperature while waiting for the cells to be seeded. The devices were acclimated inside a 37 °C incubator for at least 1 h before cell seeding.

**Cell culture, dissociation, seeding, and fixation.** All the cells used in this study were tested for mycoplasma using Mycoprobe Mycoplasma Detection kit from R&D system. Cells were passaged no more than 10 times from the initial stocks. The lung, ovarian, gastric and breast cancer cell lines were obtained commercially or were kindly donated by Dr. Ken Yamaguchi from Kyoto University Graduate School of Medicine<sup>30</sup>. The isolation and derivation of the PDCs are extensively described in<sup>22</sup>. Briefly, tumors were minced and dissociated enzymatically using 4 mg/ml collagenase type IV (Thermo Fisher, cat. no. 17104019) in DMEM/F12, at 37 °C for 2 h. The cell suspensions were strained through 70 µm cell strainers (Falcon, cat. no. 352350), prior to pelleting and resuspension in RPMI (Thermo Fisher, cat. No. 61870036), supplemented with 10% foetal bovine serum (Biowest, cat. no. S181B) and 1% penicillin-streptomycin (Thermo Fisher, cat. no. 15140122). The identities of the cell lines were checked comparing the STR profile (IDexx BioResearch) of each cell line to its original tumor. A standard cell culture incubator at 37 °C with 5% CO<sub>2</sub> was used to expand all cells. The cells were sub-cultured at 80% confluence in T25 flasks (Thermo Fisher Scientific) and reseeded into fresh flasks at ~30–50% confluence. Prior to the experiments, cells were grown at 50–70% confluence and dissociated with an enzyme-free dissociation buffer (Gibco) for about 15–30 min. In the meantime, the device was kept submerged in imaging media in a 35 mm petri dish and was acclimated inside an incubator at 37 °C. After removing the imaging buffer from the petri dish, the chambers of the device were left filled with imaging media (~2 µl). Care was taken in this very critical step to avoid exposing the lipid bilayer to air-liquid interface at all times. Immediately after removal of the media, 5 µl of cell suspension was directly deposited on top of each chamber of the device, forming small drops with high contact angle due to the hydrophobicity of the silicon-based device. Spacing between the chambers kept cells contained within their respective chambers. The three cell types labeled with different fluorescent markers (stably transfected GFP and RFP cell lines and one with Hoechst live nuclear stain) were used to ensure the absence of overspill of cells from adjacent chambers. Due to the low volume and height of the chambers, seeded cells nearly immediately contacted the bilayer. About 200 µl of sterile Milli-Q water was pipetted along the walls of the petri dishes to provide extra humidity and prevent evaporation of water from the chambers. Petri dishes were placed in a secondary container (typically a 140 mm petri dish with humidifying pads soaked in sterile Milli-Q water) previously acclimated at 37 °C. The secondary container was then placed inside a 37 °C incubator for 30 min to allow EphA clustering to occur at nearly physiological conditions. After 30 min, clustering was stopped by PFA fixation: devices inside the 35 mm petri dish were flooded with 2 ml of 4% PFA solution (Thermoscientific). Cells were fixed at 37 °C for 15 min. Thereafter, devices were washed with 1× PBS solution for five times and incubated in 10 mM NH<sub>4</sub>Cl solution for 10 min at the room temperature. Devices were then washed with 1× PBS solution for three times and mounted on glass slides using fluorescent mounting medium (Dako) and the coverslips were sealed using nail polish. Short-term storage of the devices was done at 4 °C overnight to a maximum of 2 days before imaging.

**Microscopy and image analysis.** Fixed samples were imaged using an IX81 microscope (Olympus) equipped with an ×100 oil-immersion objective (UAPON 100XTIRF, NA 1.49, Olympus), X-Cite light source (Excelitas Technologies), a CoolSNAP EZ CCD camera (Photonic) and an appropriate filter set to detect Alexa 568 conjugated to ephrinA1. Pairs of bright-field and fluorescence images were acquired for each field of view. The focus of the objective was adjusted to visualize the supported bilayer. Images were analyzed using a custom MATLAB code (can be provided upon request). The code performed bunch analysis of images found in the input folder. The analysis was divided into two separate steps: first, cell segmentation was performed by manually outlining cell borders on the composite image generated from the bright-field and corresponding fluorescence images. This process defines the regions of interests for subsequent analysis and generates a third masking image where the ROI are coded with ascending natural numbers. In the second step, image intensity distribution from the ROIs was analyzed and the image entropy for each cell quantified according to  $S_{\text{EphA}} = -\int \rho \ln \rho$ , where  $\rho = \frac{I}{I_{\text{max}}}$ . For live-cell images, trajectory analysis shown in Fig. 2c has been performed using ImageJ manual tracking plugin.

**Western blotting.** Protein lysates (30 µg) from individual cell lines were firstly dissolved in SDS-PAGE loading buffer and loaded onto 10% SDS-PAGE gel. Using Bio-Rad Mini-PROTEAN electrophoresis system, protein samples were run at 110 V, until samples ran completely through the stacking gel and into the resolving gel. The voltage was then increased to 150 V, until dye front ran off the bottom of the gel. After protein separation by SDS-PAGE, proteins were transferred from gel to a PVDF membrane (Merck, cat. no. IPSN07852). Transfer was carried out for 45 min at 110 V in transfer buffer (cold room). Following this, the PVDF membrane was blocked with 5% skim milk powder in TBS-Tween20 (0.1% v/v) at the room

temperature for ~1 h. Anti-EphA2 (Cell Signaling, cat no. 6997) and anti-beta actin (Pierce, cat. no. MA515739) were used as primary antibodies and diluted in blocking solution. The membrane was first probed with primary antibodies at 4 °C overnight. Anti-rabbit HRP (Invitrogen, cat. no. G21234) and anti-mouse HRP (Invitrogen, cat. no. G21040) were used as secondary antibodies and diluted in blocking solution. The membrane was then probed with secondary antibodies at the room temperature for 1 h. The membrane was washed three times for 10 min in TBS-Tween20 (0.1% v/v) on an orbital shaker after blocking with milk and after each probing step. Finally, the membrane was treated with chemiluminescent substrate (ThermoScientific, cat. no. 34080) and protein bands were detected by Bio-Rad Chemidoc imaging system.

**Wound-model assay.** Classical wound-model assays have been performed<sup>31,32</sup>. Briefly, a rectangular cut of PDMS was placed in the middle of a 6-well plate and let to adhere for 2 h at 37 °C before cell seeding was performed. Cancer cells were seeded around the PDMS cut and allowed to grow to full confluence inside a cell culture incubator for about 1 day. After removal of the PDMS cut, the 6-well plate was placed inside a Biostation CT (NIKON) and time lapses of cell migration by phase contrast were acquired for 12 h (time interval between image acquisitions was 15 min). Cell movement within the first 50 µm from the wound edge was analyzed using Particle Image Velocimetry in MATLAB.

**Single-cell transcriptomic analysis.** We previously generated single-cell gene expression matrix for 682 cells from PDCs from head-and-neck (HN) carcinomas from primary site (P), secondary/metastatic tumor (M), and primary with acquired cisplatin resistances (PCR) from three patients (patients 120, 137, and 148). Thus, PDC nomenclature is HN120p, HN120m, HN120pcr, HN137p, HN137m, HN137pcr, HN148p, and HN148m<sup>22</sup>. The method and the raw data to obtain the transcriptomic analysis are fully available from the references above. The expression levels were quantified as  $E = \log_2(\text{TPM} + 1)$ . We selected only the subset of genes putatively involved in the EphA2 signaling cascade (Supplementary Table 6)<sup>15</sup>. Potential markers with average expression levels lower than 1 across all the cells were excluded. We defined single-cell genomic score as the linear average of the expression levels of the remaining genes. Transcriptomic scores were normalized by average score of primary samples of each patient to account for individual genetic background. Five cells with low EphA2 score (<3) were discarded as outliers.

**Statistics and reproducibility.** Statistical analysis was performed using GraphPad Prism software. For all statistical analysis data from at least three biological repeats performed on separate days was used.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Previously generated single-cell transcriptomics data from patient-derived cell lines are available in GEO, accession number GSE117872. Raw data behind the graphs are available in Supplementary Data 1. All original images, data, and MATLAB codes are available from the corresponding author upon reasonable request. All the data are currently stored in data storage facility at Mechanobiology Institute, Singapore.

## Code availability

Custom made MatLab code for image analysis is available from the corresponding author upon reasonable request. Python codes used for image feature extraction and unbiased machine-learning method are available at Github: [https://github.com/aneeshsathe/Ephrin\\_cluster\\_analysis](https://github.com/aneeshsathe/Ephrin_cluster_analysis). Data analysis, statistics, and presentation have been made using GraphPad Prism 6 Version 6.01.

Received: 19 September 2019; Accepted: 29 June 2020;

Published online: 06 August 2020

## References

- Battle, E. et al. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263 (2002).
- Poliakov, A., Cotrina, M. & Wilkinson, D. G. Diverse roles of eph receptors and ephrins in the regulation of cell migration and tissue assembly. *Dev. Cell* **7**, 465–480 (2004).
- Holmberg, J. et al. EphB receptors coordinate migration and proliferation in the intestinal stem cell niche. *Cell* **125**, 1151–1163 (2006).
- Pasquale, E. B. Eph-ephrin bidirectional signaling in physiology and disease. *Cell* **133**, 38–52 (2008).
- Astin, J. W. et al. Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells. *Nat. cell Biol.* **12**, 1194–1204 (2010).

6. Nasreen, N., Mohammed, K. A., Lai, Y. & Antony, V. B. Receptor EphA2 activation with ephrinA1 suppresses growth of malignant mesothelioma (MM). *Cancer Lett.* **258**, 215–222 (2007).
7. Lackmann, M. & Boyd, A. W. Eph, a protein family coming of age: more confusion, insight, or complexity? *Sci. Signal* **1**, re2 (2008).
8. Brantley-Sieders, D. M. et al. The receptor tyrosine kinase EphA2 promotes mammary adenocarcinoma tumorigenesis and metastatic progression in mice by amplifying ErbB2 signaling. *J. Clin. Invest.* **118**, 64–78 (2008).
9. Nikolov, D. B., Xu, K. & Himanen, J. P. Eph/ephrin recognition and the role of Eph/ephrin clusters in signaling initiation. *Biochim. et. Biophys. Acta* **1834**, 2160–2165 (2013).
10. Janes, P. W., Nievergall, E. & Lackmann, M. Concepts and consequences of Eph receptor clustering. *Semin. Cell Dev. Biol.* **23**, 43–50 (2012).
11. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat. Rev. Cancer* **10**, 165–180 (2010).
12. Macrae, M. et al. A conditional feedback loop regulates Ras activity through EphA2. *Cancer Cell* **8**, 111–118 (2005).
13. Lohmüller, T., Xu, Q. & Groves, J. T. Nanoscale obstacle arrays frustrate transport of EphA2—Ephrin-A1 clusters in cancer cell lines. *Nano Lett.* **13**, 3059–3064 (2013).
14. Xu, Q., Lin, W.-C., Petit, Rebecca, S. & Groves, Jay, T. EphA2 receptor activation by monomeric Ephrin-A1 on supported membranes. *Biophys. J.* **101**, 2731–2739 (2011).
15. Salaita, K. et al. Restriction of receptor movement alters cellular response: physical force sensing by EphA2. *Science* **327**, 1380–1385 (2010).
16. Greene, Adrienne, C. et al. Spatial organization of EphA2 at the cell-cell interface modulates trans-endocytosis of EphrinA1. *Biophys. J.* **106**, 2196–2205 (2014).
17. Seiradake, E., Harlos, K., Sutton, G., Aricescu, A. R. & Jones, E. Y. An extracellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nat. Struct. Mol. Biol.* **17**, 398–402 (2010).
18. Seiradake, E. et al. Structurally encoded intraclass differences in EphA clusters drive distinct cell responses. *Nat. Struct. Mol. Biol.* **20**, 958–964 (2013).
19. Himanen, J. P. et al. Architecture of Eph receptor clusters. *Proc. Natl Acad. Sci. USA* **107**, 10860–10865 (2010).
20. Ojosnegros, S. et al. Eph-ephrin signaling modulated by polymerization and condensation of receptors. *Proc. Natl Acad. Sci.* **114**, 13188–13193 (2017).
21. Schaupp, A. et al. The composition of EphB2 clusters determines the strength in the cellular repulsion response. *J. Cell Biol.* **204**, 409–422 (2014).
22. Chia, S. et al. Phenotype-driven precision oncology as a guide for clinical decisions one patient at a time. *Nat. Commun.* **8**, 435 (2017).
23. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
24. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
25. Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
26. Tan, T. Z. et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **6**, 1279–1293 (2014).
27. Sharma, A. et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat. Commun.* **9**, 4931 (2018).
28. Dong, M. et al. Spatiomechanical modulation of EphB4-Ephrin-B2 signaling in neural stem cell differentiation. *Biophys. J.* **115**, 865–873 (2018).
29. Chen, Z. et al. Spatially modulated ephrinA1:EphA2 signaling increases local contractility and global focal adhesion dynamics to promote cell motility. *Proc. Natl Acad. Sci.* **115**, E5696–E5705 (2018).
30. Yamaguchi, K. et al. Identification of an ovarian clear cell carcinoma gene signature that reflects inherent disease biology and the carcinogenic processes. *Oncogene* **29**, 1741–1752 (2010).
31. Ravasio, A. et al. Gap geometry dictates epithelial closure efficiency. *Nat. Commun.* **6**, 7683 (2015).
32. Tarle, V., Ravasio, A., Hakim, V. & Gov, N. S. Modeling the finger instability in an expanding cell monolayer. *Integr. Biol. Quant. Biosci. Nano Macro.* **7**, 1218–1227 (2015).

### Acknowledgements

V.V. and J.T.G. acknowledge the support of the collaborative NRF CRP11-2012-02 grant, and MBI seed funds. A.R. and C.B. acknowledge financial support from seed funds of the Pontificia Universidad Católica de Chile. R.D. acknowledges funding from partly the National Medical Research Council (Singapore) (NMRC/CIRG/1434/2015). Proof reading was kindly provided by Andrew Wong and Sruthi Jagannatha. The cell lines were kindly given by Dr. Ken Yamaguchi from Kyoto University Graduate School of Medicine. We would like to extend our gratitude to all patients and families involved in this study.

### Author contributions

A.R., M.Z.M. performed the experiments. A.R., S.C., A.A., E.Y.C., C.B. performed the data analysis, A.S. devised the unsupervised clustering. S.C., A.S., N.G.I. produced the PDC and analyzed the genomic data. B.A., V.Y.C., A.C.G., T.Z.T., Z.C. provided wet lab technical support for the lab. H.T.O. developed the automatic cluster analysis routine, R.Y.H., R.D., J.T.G., V.V. supervised the students and postdocs. J.T.G. and V.V. devised the experiments. A.R., A.A., and V.V. wrote the paper. V.V. supervised the work.

### Competing interests

The authors declare no competing interests.

### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-01136-4>.

Correspondence and requests for materials should be addressed to V.V.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020