Metastatic potential in clonal melanoma cells is driven by a rare, early-invading subpopulation

Amanpreet Kaur[1], Karun Kiani[2], Dylan Fingerman[3], Margaret C. Dunagin[1], Jingxin Li[2], Ian Dardani[1], Eric M. Sanford[4], Jordan Pemberton[1,5], Yogesh Goyal[1,6,7], Ashani T. Weeraratna[3,8,9], Meenhard Herlyn[3], Arjun Raj[1,10,*]

[1] Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA
[2] Genetics and Epigenetics, Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[3] The Wistar Institute, Philadelphia, PA
[4] Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[5] College of Education, University of Houston, Houston, TX
[6] Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[7] Center for Synthetic Biology, Northwestern University, Chicago, IL, USA
[8] Department of Biochemistry and Molecular Biology, Johns Hopkins School of Public Health, Baltimore, MD, USA
[9] Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins School of Medicine, Baltimore MD, USA
[10] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
* Corresponding author: arjunrajlab@gmail.com

Abstract
Metastasis occurs when tumor cells leave the primary tumor site and disseminate to distal organs. Most cells remain in the primary tumor, however, the circumstances that allow certain cells to drive this dissemination remain unclear. Here, we show that rare, highly invasive melanoma cells can appear even within clonal cell lines due to non-genetic fluctuations. These differences were intrinsic to the cells independent of their external context, and were marked by transiently high levels of *SEMA3C* expression. The invasive subpopulation drove dissemination of tumor cells to distal locations in a mouse model of melanoma. The transcription factor *NKX2.2* emerged as a regulator of the proportion of invasive cells. There was an overall tradeoff between proliferation and invasion in single cells. Our results suggest that phenotypes like metastasis can be the result of intrinsic differences between single cells.

Introduction
Metastasis is the stage of cancer progression in which tumor cells take root in other tissues far away from the primary tumor. The metastatic process consists of multiple steps: cells leave the primary tumor, arrive in a new tissue, and proliferate to form a new tumor. Each step is accompanied by various molecular changes—in particular, cells must first acquire the invasive behavior required to leave the primary tumor, and after arriving at the site of metastasis, revert to a proliferative state in order to form the metastatic tumor (Mittal, 2018; Polyak and Weinberg, 2009). Each transition is characterized by selection: typically, only a very few of the cells will undergo the transition (Francí et al., 2006; Mani et al., 2008). A major question, however, is whether these rare cells are selected based on their intrinsic, cell-autonomous differences, or whether they are selected based on effectively random environmental factors. Concretely, are there rare cells primed within the primary tumor that are intrinsically primed for leaving the primary tumor (invasion (Quinn et al., 2021)), or do those rare cells leave the primary tumor because of external factors such as their local tumor microenvironment (Kaur et al., 2019; Olmeda et al., 2017)?

At the single cell level, the molecular differences between the cells primed for invasion and the bulk of the primary tumor could take a number of forms. Classically, genetic differences (i.e., mutations) had been thought to be the driver behind the transitions (Nataraj et al., 2021; Nguyen et al., 2022); however, in recent years, it has become clear that non-genetic changes in regulatory programs and pathways can also drive the switch to the invasive phenotype (Arozarena and Wellbrock, 2019; Quinn et al., 2021). In melanoma, these changes have been referred to as phenotype switching, which is largely driven by changes in Wnt signaling and the microenvironment (Webster et al. 2020; Kim et al. 2017; Widmer et al. 2013). What remains unclear is the degree to which these programs are initiated by single cells within the primary tumor that then are able to invade and disseminate (intrinsically driven (Hapach et al., 2021)), or whether cells in the right physical location would undergo this transition (extrinsically driven) regardless of their initial intrinsic phenotype. There are certainly examples of intrinsic non-genetic differences that lead to distinct behaviors throughout biology (Raj and van Oudenaarden, 2008; Symmons and Raj, 2016), and in the context of cancer, such differences have been found in many cases to drive therapy resistance (Chang et al., 2021; Emert et al., 2021; Goyal et al., 2021; Gupta et al., 2011; Shaffer et al., 2017, 2020; Sharma et al., 2010; Spencer et al., 2009; Torre et al., 2021). However, far less is known about whether non-genetic differences are the primary intrinsic differences that drive cells towards invasiveness. Suggestively, the rare cells that drive therapy resistance through non-genetic mechanisms often show some signature of the epithelial-mesenchymal transition even before the application of therapy (Quinn et al., 2021; Travnickova et al., 2019), raising the possibility similar non-genetic mechanisms can drive invasiveness in single cells.

Here, we show that even clonal cell lines have rare subpopulations that are highly invasive. This invasive phenotype is transient and hence non-genetic in origin. This rare subpopulation, when injected into mice, drives the bulk of cell metastasis, demonstrating that the subpopulation is intrinsically primed for invasiveness. Molecularly, this invasive subpopulation is marked by the expression of *SEMA3C*, and the transcription factor *NKX2.2* appears to negatively regulate the formation of the subpopulation. Our results establish that non-genetic variability can drive important cancer phenotypes like cellular invasion.

Results

We wanted to determine whether there were highly invasive subpopulations within populations of melanoma cells. In particular, we wanted to see whether differences in invasiveness could arise due to non-genetic fluctuations within clonal populations. We used two patient-derived melanoma cell lines, FS4 and 1205Lu, both of which have *BRAF* mutations (V600K for FS4, V600E for 1205Lu), that are known to be highly invasive *in vitro* and *in vivo* (Alexaki et al., 2010). In order to minimize the effects of genetic variability within the lines, we performed single cell bottlenecks. (We isolated multiple clones and selected clones that matched the overall phenotypic properties of the parental line; Supp. Fig. 1A,B.)

We tested the invasiveness of cells using transwell assays. In these assays, cells are placed in a transwell that is itself placed within a cell culture dish. The transwell is made from polytetrafluoroethylene (PTFE) membranes perforated with 8μm-wide pores. The membranes are coated with matrigel (basement membrane extract). Cells invade through the transwell pores and fall to the bottom of the enveloping dish. A gradient of serum is established between the transwell and the surrounding dish, with the dish having high serum to encourage the invasion. The number of cells that invade through the pores are a measure of the invasiveness of the population.

We wondered whether there were differences in invasive potential between individual cells in the population. To measure such differences, we measured the number of cells that invaded through the transwells over time (Fig. 1A,B). We found that as time elapsed between 0 and 24 hours, progressively more cells went through the transwells, with a small percentage of cells (~1%) rapidly invading the transwells in under 8 hours, well before the bulk of the rest of the population invaded. We obtained similar results for 1205Lu cells (Supp. Fig. 1C).

It is possible that the differences in invasiveness are associated with intrinsic differences in the cells before putting them on the transwell, or alternatively that these differences are just due to the quasi-random differences in the environmental context of individual cells as they land in the transwell. In order to discriminate between these possibilities, we wanted to find a prospective marker of which cells were more invasive, which would establish that intrinsic differences determined invasive potential. To identify candidate markers of this putative intrinsic state, we isolated early invaders (invaded within 8 hours) and late invaders (invaded between 8 and 24 hours) and subjected those cells to RNA sequencing to measure the transcriptomic profile of the various subpopulations of cells (Fig.1C). Overall, we found 692 genes were upregulated and 849 were downregulated in the early vs. late invading subpopulation in FS4 cells ($\log_2$ fold change > 0.5, $p_{adjusted} < 10^{-5}$); these genes were enriched in the cell cycle, cell response to stress, and DNA replication gene ontology categories. Amongst the most differentially expressed genes was the surface marker *SEMA3C*. It was possible, though, that *SEMA3C* expression did not mark the initial highly invasive subpopulation, but rather was itself upregulated by the process of invading rapidly through the transwell. To establish that *SEMA3C* did indeed mark the intrinsically invasive subpopulation, we used antibodies against SEMA3C to label the parental population, then sorted cells with the top 1.5% of staining as SEMA3C-high cells (Fig.1D). We then loaded these subpopulations on the transwells and looked at the rate of invasion. We found that the SEMA3C-high cells were far more invasive, intrinsically, than SEMA3C-low cells and the population overall, thus demonstrating that cells vary intrinsically in their invasiveness, and the very invasive subpopulation is marked by the expression of *SEMA3C* (Fig. 1E). (Overexpression of *SEMA3C* in FS4 cells revealed no changes in invasiveness, suggesting that SEMA3C is a marker with no functional relevance to invasiveness *per se*; Fig. 1D, Fig. 2A-B)

Invasive potential is thought to be critical for the process of metastasis. In order to test whether the single cell differences in invasiveness as measured by transwell passage are important for metastasis, we turned to a mouse model of metastasis. In this model, human melanoma cells are injected into NOD *scid* gamma mouse (NSG) immunocompromised mice (that enable engraftment of cross-species tissue), where they form a primary tumor, and these cells can disseminate to distant sites in the mouse, such as the lung, in a process that is thought to be akin to metastasis. To look for differences in the metastatic potential of the early invading subpopulation, we mixed together equal numbers of early and late invaders and injected them into mice. We only used 1205Lu cells for these experiments because FS4 cells have not been tested for tumor formation in NSG mice, whereas 1205Lu cells have been used extensively to form tumors in mice under a variety of conditions (Alexaki et al., 2010). To distinguish the two populations *in vivo*, we first labeled the early invaders with green fluorescent protein (GFP) via a lentivirus and the late invaders with mCherry, then mixed the cells together before injecting them into the mouse. (We labeled the cells with sufficient virus so that over 50% of the early invaders were labeled with GFP and 98% of the late invaders were labeled with mCherry; Supp. Fig. 2A) We then sampled lungs from mice at various times post-injection to look for metastatic cells (Fig.1F, Suppl. Fig. 2B,C).

Technically, the mCherry cells were not detectable due to the fluorescence of the mCherry protein not being visible in the mouse sections. Nevertheless, we were able to detect late invaders in the population by using a human-specific *MALAT1* RNA FISH probe that binds only to human *MALAT1* RNA and not mouse *MALAT1* RNA (Haimovich et al., 2017). Thus, that probe would show us the location of all human cells, and then GFP positivity would allow us to estimate how many of the human cells were from the initial early-invading subpopulation (Fig 1G-J). We imaged cells from both the primary tumor and the lungs at the time of sacrifice (it was hard to detect the primary tumor in mice sacrificed at 19 and 25 days post-injection due to small tumor size). As expected, in the primary tumor, there was a roughly equal mix of human melanoma cells that were GFP-positive and negative (Fig. 1I). In the lung, however, we saw predominantly GFP-positive cells, showing that the vast majority of cells that migrated from the primary tumor site were initially early invading cells (Fig.

1J). The number of GFP cells in the lung was variable, but generally increased with time. Thus, we established that the highly invasive subpopulation was able to drive metastasis *in vivo*.

NKX2.2 is a transcription factor that promotes the invasive subpopulation

We wondered whether certain regulatory factors could change the fraction of cells that have the early-invading phenotype. In order to find such regulators, we used Assay for Transposase-Accessible Chromatin sequencing (ATAC-sequencing), a method that reveals regions of chromatin that are accessible, and thus thought to indicate areas in which transcription factors bind. We performed ATAC-sequencing on FS4 cells that either went quickly through the transwell (early-invaders) or went through the transwell slowly (late-invaders) or did not go through the transwell at all (non-invaders). ATAC-sequencing reads often come in clusters ("peaks") that harbor transcription factor binding sites of putative regulators; we looked for peaks with different intensities between conditions to identify such regulators. Overall, the number of differential peaks was 1107 with a threshold of minimal fold change of 1.5 and val threshold of 1e-15, which is a relatively small number as compared with many perturbations (Kiani et al., 2022), and may reflect the shallow chromatin accessibility changes associated with single cell fluctuations within a homogeneous population (Shaffer et al., 2017). Consequently, differential peak analysis and identification of transcription factor binding motifs within those differential peaks revealed very few regulators that rose to statistical significance; in the comparison between early invaders and non-invaders, we found *NKX2.2*, *PBX1* and *CHOP* binding sites as being more accessible in the early invaders, and *NKX3.2* as being more accessible in the non-invaders. (Overall, few factors were identified in any comparison between any conditions.)

We wanted to check that these factors regulated the proportion of early-invading cells by knocking them out. We attempted to knock out the transcription factors *NKX2.2* and *PBX1* in both FS4 and 1205Lu cells using the CRISPR/Cas9 system with multiple guides targeting the genes; of these, we were only able to validate knockdown of the target mRNA for *NKX2.2* expression in 1205Lu cells by single molecule RNA FISH (Raj et al., 2008), which revealed a reduction from 1.4 to 0.5 mRNA per cell (Supp. Fig. 3A; averages across 4 control and 4 knockout guides). We then measured the invasiveness and proliferation of these *NKX2.2* knockout 1205Lu cells. We found that both invasiveness and proliferation were starkly increased (Fig. 2C,D), with an increase of 10.36-fold in the number of cells invading through the transwell (averaged across 3 control and 4 knockout guides), and an increase in growth rate from 0.013 per hour to 0.026 per hour (averaged across 3 control and 4 knockout guides).

It is worth noting that the direction of the effect is the opposite of what one might have expected, given that *NKX2.2* binding appeared to be increased (as opposed to decreased) in the early vs. late invading FS4 cells as measured by ATAC-sequencing—thus, knocking out *NKX2.2* might be expected to decrease invasiveness. One could attribute the difference to the difference in cell line. In order to compare the effects of the *NKX2.2* knockout in the 1205Lu cells to our previous results, we performed RNA sequencing on the *NKX2.2* knockout cells and compared the effects on gene expression to the gene expression differences between early vs. non-invaders across the two cell lines. We found that, as expected, *NKX2.2* knockout led to a large number of changes in gene expression, some of which overlapped with early vs. non-invaders, and some of which were unique to the *NKX2.2* knockout condition (Fig. 2E). Of those that overlapped, however, we found that the expression difference in the *NKX2.2* knockout cells were generally more concordant with those of the early invading cells than the late invading cells for both FS4 and 1205Lu cells (Fig. 2F). Overrepresentation analysis of overlapping genes between those differentially expressed between *NKX2.2* knockout guides and control guides and those downregulated in early invaders vs non-invaders was significant for the following GO categories: cell migration, ECM organization, locomotion, cell motility (Fig. 2G). Together, we concluded that *NKX2.2* is a regulator of the size of the invasive subpopulation, and the knockout of *NKX2.2* leads to changes in cellular state that are more concomitant with the early than the late invading subpopulation.

Early invaders are a transient subpopulation
The early invading subpopulation appeared even in cells grown out from a single cell, thus suggesting that cells can transition from a late-invading phenotype to an early-invading one and vice versa. To establish that these transitions occur, we wanted to show that early invading cells would eventually revert to a late invading phenotype. We collected early invaders by isolating FS4 cells that went through the transwell within 8 hours, then culturing those cells for either 3 days or 14 days (alongside parental cells not subjected to the transwell as a control). We then took those cells and plated them into transwells for an additional day, after which we measured the number of cells that passed through the transwell (Fig. 3A). We found that after 3 days, the originally early invading population was still quite invasive as compared to the parental control cells (Fig. 3B). However, after 14 days, the originally early invading population showed much less invasiveness, with a similar number of cells going through the transwell as the parental population (Fig. 3B). Thus, we concluded that early-invading cells were able to revert over time to the less invasive phenotype of the original population.

To corroborate these findings, we also used the fact that *SEMA3C* was a positive marker of the early-invading subpopulation. We performed the same early/late isolation experiments using transwells as described above in both FS4 and 1205Lu cells, culturing them for 1 or 14 days. After this culture period, we took the early invading, late invading, and parental populations and performed single molecule RNA FISH to measure expression of *SEMA3C* mRNA in these conditions (Raj et al., 2008). As expected from the phenotypic reversion experiments above, we found that *SEMA3C* mRNA counts were higher in the originally early-invading population after 1 day of culture than in the late invading and parental populations, but these differences were largely gone after 14 days of culture (Fig. 3C).

As another confirmation, we sorted cells (FS4 and 1205Lu) by SEMA3C expression by fluorescence assisted cell sorting into high and low SEMA3C-expressing subpopulations (and parental), then cultured those subpopulations for 1 and 14 days as above. We then looked for *SEMA3C* expression in those subpopulations by single molecule RNA FISH. In 1205Lu, we found that the expression of *SEMA3C* reverted back to the population baseline by 14 days (Fig. 3D). However, in FS4 cells, we saw persistence of *SEMA3C* levels even after 14 days in culture. It is unclear why *SEMA3C* levels did not revert even though the phenotype reverted in FS4 cells; it does suggest that while *SEMA3C* expression is a marker for the early-invading subpopulation in naive cells, *SEMA3C* does not itself force cells to adopt a early-invading phenotype, consistent with our results showing that *SEMA3C* is likely a marker for invasiveness with no functional effect *per se* (Fig. 2A,B).

Proliferation and motility anti-correlate at the single cell level
As is well documented in the literature on phenotype switching in melanoma, there is a strong anti-correlation between proliferation and invasion, with cell lines often showing an increase at one behavior at the expense of a decrease in the other (Hoek et al., 2006, 2008). Given that we observed a subpopulation of early-invading cells, we wondered whether we could observe this same tradeoff at the single cell level. We separated cells by how quickly they went through the transwell into early, late, and early+late invaders combined. We then measured the growth rate of these subpopulations by live-imaging cells for 7-10 days (Fig. 4A). We found that the early invaders grow the most slowly, with growth rates that were 20% lower in FS4 cells and 8.6% lower in 1205Lu cells (Fig. 4B-C). We found similar results sorting FS4 cells by SEMA3C expression, with SEMA3C-high cells showing a 19.6% lower growth rate (Fig. 4D). (Interestingly, *NKX2.2* knockout cells showed markedly increased invasion and proliferation (Fig. 2A,B), suggesting a change in regulation of both processes.)

We also looked at the tradeoff of proliferation with motility. Upon isolating early and late invaders and subjecting cells to time lapse analysis, we found that the average speed of early invaders was 20.9% higher than late invaders (Fig. 4E), while the average growth rate was lower by 20% (Fig. 4F). At the single cell level, tracking both speed and growth rate, we observed a modest anticorrelation between the two in single cells (Fig. 4G). Similar effects were observed when comparing SEMA3C-high vs. low cells (14.5% increase in speed

and 16.5% decrease in growth rate; Supp. Fig. 4A-C). We also looked at the speed and growth rate of sibling cells and cousin cells, noting a strong correlation between the speeds and growth rates of related cells (Fig. 4H,I, K,L), whereas unrelated cells showed little correlation (Fig. 4J,M). Similar effects were observed when comparing speeds and growth rates of related and unrelated SEMA3C-high vs. low cells (Supp. Fig. 4D-I). These results suggest there is a tradeoff between invasiveness and growth rate as well as motility and growth rate, and that these phenotypes are heritable over at least a couple cell divisions.

Discussion

We here report the existence of a highly invasive subpopulation that can arise even in clonal populations of melanoma cells. Cells can transition between the invasive and non-invasive phenotypes, leading to a small but steady percentage of invasive cells within the population. Critically, these cells were the ones that metastasized most readily to the lung away from the site of the primary tumor.

Our findings show that some cells can have an intrinsic, but non-genetic, ability to invade. In principle, the fact that only a small percentage of the cells invade means that there must be something different about that small percentage of cells. These differences could be intrinsic, meaning from the cell regardless of environment, or extrinsic, meaning from the environment. Our work shows that intrinsic differences can lead to some cells being more invasive regardless of their local environment, as demonstrated by the prospective isolation of this invasive subpopulation by SEMA3C protein levels. Moreover, our reversion experiments show that these differences do not persist indefinitely, suggesting that they are not of genetic origin. Thus, our results suggest that metastasis may, at least in part, be driven by non-genetic intrinsic differences between tumor cells. At the same time, there is ample evidence that microenvironmental factors can also influence whether cells become metastatic (Kaur et al., 2016; Wu et al., 2021). Determining the interactions between intrinsic and extrinsic cues for invasion and metastasis is an ongoing challenge for the field.

These results in many ways mirror those we and others have obtained for therapy resistance (Yang et al., 2021). In our model for resistance (Emert et al., 2021; Goyal et al., 2021; Shaffer et al., 2017, 2020; Torre et al., 2021), there is a rare subset of cells that are primed for resistance, and then once drug is applied, those primed cells progress towards becoming fully resistant. Analogously, here we are describing a rare subset of cells that are primed for invasion. A first natural question is whether it is the same subset of cells that drive both resistance in invasion, especially given the abundant data suggesting that cells that are more metastatic tend to be more drug resistant (Dratkiewicz et al., 2019; Goyal et al., 2021). We found some such signature, with *AXL* and *EGFR* showing increased expression in early invaders in 1205Lu cells (not in FS4). However, SEMA3C expression was unchanged in primed WM989 cells (Shaffer et al., 2017). Overall, there is not a strong overlap between these two subpopulations. It is thus possible that these two biological processes also fluctuate along somewhat separate axes of priming (Travnickova et al., 2019). How, then, could one explain the general correlation between these properties? It could be that some cell lines and tumors are just more prone to priming in general, in which case cells would more often be in the state of being primed for either invasion or resistance. A more general and extensive analysis would be required to make these determinations.

A second question raised by the analogy between priming for resistance and invasion is whether invasive cells adapt to challenge in the same way that cells primed for resistance adapt when faced with prolonged exposure to drug (Emert et al., 2021; Goyal et al., 2021; Shaffer et al., 2017). In the case of resistance, cells that are transiently primed for resistance become permanently resistant over prolonged exposure, but here, cells that are invasive are able to fluctuate back to a non-invasive state over time. This lack of permanence may reflect the fact that the invasive cells are not subjected to stress—in our case, cells merely pass through a transwell, which may be the reason for the “burning in” of the phenotype in the case of resistance. In the full context of metastasis, it could be that the processes of seeding cells in distant places, along with subsequent reversions

to more proliferative phenotypes, could induce more permanent changes in the phenotypes of these cell subpopulations.

Acknowledgements:
We thank Sam Reffsin for assistance with the lentivirus CRISPR, Ian Mellis for help with the ImageJ code for cell counting, Ally Cote for help with R code for lineage tracing experiments, Ben Emert for providing lentivirus plasmids and Ophir Shalem for providing the CRISPR KO plasmid. We thank the Genomics Facility at the Wistar Institute, especially Sonali Majumdar and Sandy Widura, for assistance with sequencing. We thank the Flow Cytometry Core Laboratory at the Children's Hospital of Philadelphia Research Institute for assistance with flow cytometry and fluorescence-activated cell sorting. We thank the Penn Center for Musculoskeletal Disorders Histology Core (P30 AR069619) for their guidance on tissue cryo-sectioning. AK acknowledges support from NIH K00 CA-212437-02; KK acknowledges support from NIH T32 GM008216; YG acknowledges support from the Burroughs Wellcome Fund Career Awards at the Scientific Interface, the Jane Coffin Childs Memorial Fund, and the Schmidt Science Fellowship; DF and MH acknowledge support from NIH grants RO1 CA238237, U54 CA224070, PO1 CA114046, P50CA174523 and the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation; ATW is supported by a Team Science Award from the Melanoma Research Alliance, R01CA174746 and R01CA207935 and P01 CA114046; AR acknowledges support from NIH Director's Transformative Research Award R01 GM137425, NIH R01 CA238237, NIH R01 CA232256, NIH P30 CA016520, NIH SPORE P50 CA174523, and NIH U01 CA227550.

Author Contributions
AK and AR conceived and designed the project. AK designed, performed, and analyzed all experiments, supervised by AR. EMS assisted AK in designing and performing ATAC-sequencing experiments and KK performed analysis of all sequencing data. DF assisted AK in mouse experiments with inputs from MH and AR. MCD assisted AK in smFISH analysis. JL assisted AK in live imaging using incucyte. ID assisted AK in single molecule RNA FISH imaging and cryosectioning tissue samples. YG assisted AK in the design of RNA sequencing experiments. JP assisted AK in single-cell lineage tracking. ATW provided FS4 cells for experimental purposes. AK and AR wrote the manuscript with inputs from all authors.

Data and Code Availability
All raw and processed data as well as code for the analyses in this manuscript can be found at: https://www.dropbox.com/sh/u1jnfbh8m8vbm5y/AAAPwqNiWd4AZu7vzULs3BL_a?dl=0

Competing Interests
AR receives royalties related to Stellaris RNA FISH probes. All other authors declare no competing interests.

Figure 1: A rare, early invading subpopulation of cells is primed for invasion.
A. FS4 melanoma cells were added to transwells and allowed to invade. The invading cells were imaged at the bottom of the tissue culture plate and the cumulative number of cells were quantified. Error bars represent standard error across 3 biological replicates.
B. Schematic showing the setup of transwell invasion assay used to isolate early, late and total invading melanoma cells. Graph shows percentage of invading cells from different subpopulations in FS4 cells.
C. Heatmap showing the differentially expressed genes between early-invading and non-invading FS4 melanoma cells.
D. Schematic showing the staining and isolation of SEMA3C-expressing cells from the parental population.
E. FS4 cells were isolated based on SEMA3C expression as described. Cells were seeded on the transwells and invading cells were counted.
F. Schematic showing the setup of *in vivo* studies to determine invasion potential of different subpopulations. 1205Lu melanoma cells were sorted into early- and late-invading subpopulations and tagged using GFP (early-

invading subpopulation) and mCherry (late-invading subpopulation) using lentiviral transduction. Tagged cells were mixed in equal proportion and injected in the left flank of NSG mice. Mice were sacrificed at various predetermined intervals (timepoint 1 - palpable tumor, timepoint 2 - 1 week after palpable tumor, timepoint 3 and 4 - tumor burden >1000mm^3).
G. Primary tumor and lung samples collected from mice injected with early and late subpopulations were imaged via single molecule RNA FISH for the expression of GFP and *MALAT1*. A representative sample shows expression of GFP protein fluorescence, GFP mRNA and *MALAT1* by single molecule RNA FISH in various fluorescence channels.
H. Schematic showing the analysis of 1205Lu derived mouse primary tumor and lung samples to determine number of early and late invading cells in the tissue sections. Mouse samples (n=1-3 as shown) were imaged using single molecule RNA FISH for GFP and human specific *MALAT1*. Samples were run through the DENTIST2 pipeline and RNA counts of each cell were calculated. Cells with human specific *MALAT1* were categorized as human. If GFP protein and single molecule RNA FISH expression was found, cells were labeled as coming from the early invading subpopulation, if only *MALAT1* expression was observed, cells were labeled as coming from the late invading subpopulation.
I. Primary tumor and lung samples collected from mice injected with 1205Lu early and late subpopulations were imaged for single molecule RNA FISH expression of GFP and *MALAT1*. In the tissue sections, cells were classified and labeled as early and late invading as shown as in D. Each dot represents a cell identified in the image and the color represents the classification of each cell. Representative plots show the imaging area of sectioned samples and the corresponding analysis over different time intervals.
J. The graphs show the total number of early or late invading cells identified in various samples across different experimental timepoints.

Figure 2: *NKX2.2* is a transcription factor that promotes the invasive subpopulation
A. SEMA3C ORF and empty vector control clones were added on the transwells and allowed to invade. Cells collected at the bottom of the well were counted as a measure of invasion.
B. SEMA3C ORF and empty vector control clones were seeded at equal numbers in well plates and allowed to proliferate. Cells were imaged every 24 hours, counted, and growth rate was calculated.
C. 1205lu melanoma cells expressing either *AAVS* or *NKX2.2* knockout were seeded on the transwell and the number of invading cells was calculated. Error bars represent standard error across 3 replicates.
D. 1205lu melanoma cells expressing either *AAVS* or *NKX2.2* knockout were seeded in tissue culture plates and cells were allowed to grow for 10 days. Cells were imaged every 24 hours and cell counts at different times were determined and used to calculate growth rate of the cells. Error bars represent standard error across 3 replicates.
E. FS4 and 1205Lu cells were used to isolate early invaders, late invaders and non-invaders. Differentially regulated genes were determined across cell lines and different subpopulations. The differentially regulated genes were also compared against 1205Lu *AAVS* versus *NKX2.2* knockout cells.
F. Odds ratio showing expression difference in the 1205Lu *NKX2.2* knockout cells with the early invading cells for both FS4 and 1205Lu cells
G. Overrepresentation analysis of overlapping genes between those differentially expressed between *NKX2.2* knockout guides and control guides and those downregulated in early invaders vs non-invaders is shown.

Figure 3: Early invaders are a transient subpopulation
A. Schematic showing the experimental design for testing reversion of invasion phenotype. Early and late invading subpopulations were collected from FS4 parental melanoma cells and allowed to grow for 14 days. At different time intervals, early invaders, late invaders and parental cells were added to the transwell and the number of invading cells were measured.

B. Graph showing the number of invading cells when FS4 early invaders, late invaders and parental cells were allowed to proliferate for a different number of days prior to testing for invasion. Error bars represent standard error across 3 replicates.
C. FS4 and 1205Lu melanoma cells were used to isolate early and late invading cells. Cells were allowed to grow for 1 or 14 days and fixed and imaged for *SEMA3C* RNA using single molecule RNA FISH. Data from 250-2000 cells is shown.
D. FS4 and 1205Lu melanoma cells were sorted based on SEMA3C expression levels and allowed to grow for 1 or 14 days. Cells were fixed and imaged for *SEMA3C* RNA using single molecule RNA FISH. Data from 250-2000 cells is shown.

Figure 4: Proliferation and motility anti-correlate at the single cell level
A. Schematic showing how different invasive subpopulations of melanoma cells were isolated. Cells were allowed to proliferate for 7 days and imaged every 24 hours to obtain cell counts.
B. FS4 cells were isolated into early, late, and early+late invading cells and allowed to proliferate over 7-10 days. Cells were imaged every 24 hours and cell counts were obtained. Cell counts were used to calculate growth rate of the cells and are shown. Error bars represent standard error across 3 replicates.
C. 1205Lu cells were isolated into early,late and early+late invading cells and allowed to proliferate over 7-10 days. Cells were imaged every 24 hours and cell counts were obtained. Cell counts were used to calculate growth rate of the cells and are shown.Error bars represent standard error across 3 replicates.
D. FS4 cells were sorted into SEMA3C high and low expression cells. Cells were cultured for 14 days and imaged every 24 hours to obtain cell counts. Growth rate was calculated from cell counts and are shown. Error bars represent standard error across 3 replicates.
E. Early invaders, late invaders and parental cells were live-imaged for ~10 days hourly and single cells were tracked manually for cell position, cell division and lineage. Cell speed was calculated for each cell using average distance traveled over time.
F. Growth rate was calculated for each cell from the time taken by the cell to divide.
G. For each cell, speed and growth rate was calculated and compared.
H. Early invaders, late invaders and parental cells were live-imaged for ~10 days hourly and single cells were tracked manually for cell position, cell division and lineage. Lineages were traced manually from single cells. Sister, cousin and unrelated cells were identified across 10 generations. Graph shows correlation between speed between sister cells.
I. cousin cells.
J. unrelated cells.
K. Early invaders, late invaders and parental cells were live-imaged for ~10 days hourly and single cells were tracked manually for cell division and lineage. Sister, cousin and unrelated cells were identified across 10 generations. Graph shows correlation between growth rate between sister cells.
L. cousin cells.
M. unrelated cells.

Star Methods
Cell lines:
FS4 cells were a gift from Dr. Ashani Weeraratna. 1205Lu cells were a gift from Dr. Meenhard Herlyn. Cells were cultured in DMEM with glutamax (Invitrogen #10569010) supplemented with 10% serum (Gibco #16000044) at 37°C and 5% CO_2, unless otherwise noted. Cells were single cell bottle-necked by diluting and plating single cells in 96 well plates. Colonies selected for further analyses are labeled in the manuscript. Cells were verified to be mycoplasma free by bi-annual testing using MycoAlert (Cambrex) assay offered by University of Pennsylvania Cell Center Services.

DNA and plasmids:
SEMA3C ORF plasmid was obtained from Sino Biologicals and cloned into lentiviral backbone LV067 under the control of EF1α promoter. Plasmid without open reading frame (ORF) expression were used as control plasmids. LentiEFS-H2b-GFP and LentiEFS-H2b-mCherry were prepared in-house and used for GFP and mCherry overexpression in the 1205lu melanoma cell line. LentiCRISPR v2 (addgene #52961) and LentiCRIPSR v2 expressing *AAVS1* guides were a gift from Dr. Ophir Shalem at the University of Pennsylvania and were used for generating CRISPR knockout cell lines. Single stranded oligos against *PBX1* and *NKX2.2* gene targets were selected from Human Brunello CRISPR knockout pooled library (Doench et al., 2016) and purchased from IDT, annealed into double stranded sgRNA oligos and cloned into LentiCRISPRv2 plasmid using one step BsmBI (NEB #R0580) digestion and T4 ligase (NEB #M0202S) reaction. LentiCRISPRv2 plasmid without cloned sgRNAs was used for cas9 only transduction.
Lentiviral production was carried out as described in the protocol developed by the TRC library (Broad Institute). Briefly, 293T cells were cotransfected with lentiviral gene vector and lentiviral packaging plasmids (pCMV-dR8.74psPAX2, pMD2.G) using lipofectamine 2000 (Life Technologies). The supernatant containing virus was harvested at 36 and 60 hours, combined and filtered through a 0.45-µm filter. For transduction, the cells were layered overnight with lentivirus containing 8 µg/mL polybrene. The cells were allowed to recover for 24 hours and then selected using puromycin selection marker (1ug/ml) for 2-3 days until simultaneously treated non-transduced cells were completely dead.

Transwell assay:
Plates and preparation: Corning costar plates (#3422 and #3428) with polycarbonate membrane and 8 micron pore size were purchased from Thermo Fisher Scientific. Basement membrane matrigel (BD #354234) was purchased from Thermo Fisher Scientific and 26µl was diluted in 300µl PBS and used to coat transwells at 5.8µg/cm^2 based on surface area of the well. Wells were allowed to dry overnight at room temperature and used within 2 weeks.
Transwell assay: Plates were allowed to normalize to room temperature. 3×10^5 cells/cm^2 were collected and resuspended in serum free media. To establish serum concentration gradient, 30% serum containing media was added to the outer well and the setup was allowed to incubate for either 8 or 24 hours at 37°C at 5% CO_2. Whole wells were imaged with incucyte (Sartorius) as needed per transwell assay. For continuous scans, to allow for automated focusing on empty wells, 4.0 µm beads (Invitrogen, T7283) were added at the bottom of the well. Quantification of the cell count was performed using Fiji (ImageJ) macro. The macro processes all samples automatically and is responsible for cropping images to include only the area under the transwell, thresholding images and generating a count of thresholded objects using a size cutoff of 100 pixels, which allows us to exclude beads from being counted. The results are provided for each well in a csv format. The code for the macro can be found here -
https://www.dropbox.com/s/ipmvjy85sc3tkdn/Macro_Brightfield_thresholdCellsandCount.ijm?dl=0

SEMA3C staining and flow cytometry:
SEMA3C antibody was purchased from R&D Systems (#MAB1728). Cells were collected after trypsinization, washed with FACS buffer (PBS + 1mM EDTA) and stained with SEMA3C (1×10^6 cells, 1:50 dilution, 2% BSA in PBS+1mM EDTA) for 20 minutes at room temperature. Cells were allowed to mix by placing the tubes in a tube revolver/rotator (Thermo Fisher Scientific). Anti-rat secondary antibody was purchased from Invitrogen (#A11006, Alexa 488) and used at 1:500 dilution for 20 minutes at room temperature with gentle rotation to mix the cells. Cells were washed with FACS buffer and resuspended in FACS buffer. Cells were sorted on FACSJazz Sorter or MoFlo Astrios Sorter at the Flow Cytometry Core Laboratory of The Children's Hospital of Philadelphia, Pennsylvania. Sorted cells were collected in 15ml conical tubes, pelleted, resuspended in fresh media and plated on tissue culture treated 24 or 6 well plates (Falcon brand, Thermo Fisher Scientific). Cells were allowed to recover overnight and used for various assays.

Single molecule RNA FISH:
Cells were seeded on glass bottom well plates (#12-565-470, Nunc Lab-tek, Thermo Fisher Scientific) and allowed to settle overnight. Cells were fixed in 4% formaldehyde followed by 2 washes in PBS and stored in 70% ethanol at 4°C for a minimum of 4 hours before use. Wells were washed with Wash Buffer for 5 minutes and staining buffer was added. Samples were covered with a coverslip and incubated overnight at 37°C in a humidified chamber. Next morning,coverslip was removed and wells were washed twice with Wash Buffer at 37°C for 30 minutes. DAPI (1:10000 final dilution) was added to the Wash Buffer for the second wash. After the Wash Buffer was removed, samples were incubated in 2X SSC (Ambion #AM9765) and imaged at 60X on Nikon TI-E inverted fluorescence microscope equipped with a SOLA SE U-nIR light engine (Lumencor), a Hamamatsu ORCA-Flash 4.0 V3 sCMOS camera, and 4X Plan-Fluor DL 4XF (Nikon MRH20041/MRH20045), 10X Plan-Fluor 10X/0.30 (Nikon MRH10101) and 60X Plan-Apo λ (MRD01605) objectives. We used the following filter sets to acquire different fluorescence channels: 31000v2 (Chroma) for DAPI, 41028 (Chroma) for Atto 488, SP102v1 (Chroma) for Cy3, 17 SP104v2 (Chroma) for Atto 647N, and a custom filter set for Alexa 594. We tuned the exposure times depending on the dyes used. For large tiled scans, we used a Nikon Perfect Focus system to maintain focus across the imaging area.

RNA sequencing and analysis:
RNA collection and library prep: Each treatment/sample was tested in 3 separate biological replicates. Total RNA isolation was performed using the phenol-chloroform extraction followed by RNA cleanup using RNAeasy Micro (Qiagen 74004) kit. For transwell assays, library preparation was performed using Nebnext single-cell/low input RNA library prep kit (E6420L, NEB). For *NKX2.2* CRISPR experiments, library preparation was done using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490L) integrated with NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB E7770L). NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) oligos (NEB E7600S) was used for assigning indices to all samples prior to pooling. Pooled samples were run using an Illumina NextSeq 550 75 cycle high-output kit (Illumina 20024906). Samples were processed as single end reads (75:8:8), as previously described (Mellis et al., 2021).
Data analysis: RNA sequencing reads were aligned to the human genome (hg38) with STAR v2.5.5a and uniquely mapped reads were counted using HTSeq v0.6.1 to generate a count matrix. From the counts matrix, tpm and other normalized values were calculated for each gene using publicly available scripts at: github.com/arjunrajlaboratory/RajLabSeqTools
Differential expression analysis was performed using DESeq2 in R4.1.0. Overrepresentation analysis was performed using Web-based gene set analysis toolkit (WebGestalt) using the “biological process” gene ontology functional database against the “genome protein-coding” reference set.

ATAC sequencing:
ATAC library preparation and sequencing: We used Omni-ATAC protocol described in Corces et al., 2017 (Corces et al., 2017). For transwell assays, early, late and non invading cells were collected using trypsinization and 50,000 live were collected in DNA lo-bind tubes (#13698791, Thermo Fisher Scientific). Cells were washed in PBS and lysed to collect the nuclei. Next, Illumina Tagment DNA Enzyme TDE1 (20034197) was used at the tagmentation step, followed by cleanup to remove the excess enzyme. Libraries were prepared using custom primer indices followed by double-sided bead purification using Agencourt AMPure XP magnetic beads (A63880). Libraries were tested for average fragment length using Agilent High Sensitivity DNA Kit (5067-4626) and concentration was tested using Qubit™ dsDNA HS kit (#Q32851, Invitrogen). DNA was stored at -20°C until used for sequencing. Samples were run as paired-end using 150-cycle NextSeq 500/550 High Output Kit v2.5.
ATAC-sequencing analysis: We created a paired-end read analysis pipeline using the ENCODE ATAC-seq v1 pipeline specifications (https://www.encodeproject.org/documents/c008d7bd-5d60-4a23-a833-67c5dfab006a/@@download/attachment/ATACSeqPipeline.pdf). Briefly, we aligned our ATAC-seq reads to the hg38 assembly using bowtie2 v2.3.4.1, filtered out low-quality alignments with samtools v1.1, removed

duplicate read pairs with picard 1.96, and generated artificial single-ended text-based alignment files containing inferred Tn5 insertion points with custom python scripts and bedtools v2.25.0. To call peaks, we used MACS2 2.1.1.20160309 with the command, “macs2 callpeak --nomodel --nolambda --keep-dup all --call-summits -B --SPMR --format BED -q 0.05 --shift 75 --extsize 150”. The pipeline is publicly available at github.com/arjunrajlaboratory/atac-seq_pipeline_paired-end.

Mouse tumor implantation and growth:
All mouse experiments were conducted in collaboration with Dr. Meenhard Herlyn at The Wistar Institute, Philadelphia, PA. NSG mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ) were bred in-house at The Wistar Institute Animal Facility. All experiments were performed under approval from the Wistar Institute Care and Use Committee (protocol 201174). 50,000 melanoma cells were suspended in DMEM with 10% FBS and injected subcutaneously in the left flank of the mouse. Tumors were allowed to develop and measure twice weekly using vernier calipers and tumor volumes were calculated as 0.5×L×W×W, where L is the longest side and W is a line perpendicular to L. Mice sacrifice timepoints were predetermined as follows 1) palpable tumor 2) 1 week post palpable tumor 3) tumor volume greater than 1000mm^3. It was expected that mice will be variable in reaching the 1000mm^3 mark, hence this group was expanded to 10 mice and 2 sacrificial timepoints. Mice were euthanized using CO_2 inhalation and primary tumor and lungs were harvested and covered in Optimal Cutting Temperature Compound (#23-730-571, Thermo Fisher Scientific) and immediately frozen in liquid nitrogen and stored in -80°C.

Tissue sectioning and RNA FISH:
Mouse tissue samples were cryosectioned using Leica CM1950 cryostat within the Center for Musculoskeletal Disorders (PCMD) Histology Core. 6 microns sections were placed on positively charged SuperFrost Plus Slides (Thermo Fisher Scientific) and immediately fixed in 4% formaldehyde for 10 min at room temperature. Slides were washed twice in PBS and stored in 70% ethanol in LockMailer microscope slide jars at 4°C. For staining, slides were removed from 70% ethanol and transferred to a LockMailer jar containing Wash Buffer (2X SSC, 10% formamide) for 1-2 minutes and treated with 1000µl 8%SDS for 1 minute. Slides were tipped on the side to remove the 8%SDS and transferred to Wash Buffer for 2 minutes. 50µl staining solution was prepared for each slide by adding 1µl of the probe in the hybridization buffer (10% dextran sulfate, 2X SSC, 10% formamide). After the addition of staining solution, a cover slip was added on the slide and slides were stored in a humidified chamber at 37°C overnight. Next day, slides were transferred to a LockMailer jar containing Wash Buffer for 2 washes of 30 minutes each at 37°C. During the second wash, DAPI (1:10000) was added. Slides were removed from Wash Buffer and transferred to a LockMailer jar containing 2X SSC for 1 minute. Next, 50µl 2X SSC was added to the slide, secured with a coverslip and sealed using rubber cement (Elmer) and allowed to dry before imaging.

Proliferation assay:
24 well tissue culture plates (#08-772-1, Thermo Fisher Scientific) were used for all proliferation assays. Cells were collected and counted using a manual hemocytometer (#02-671-54, Hausser Scientific, Thermo Fisher Scientific). 1000 cells were plated in each well and allowed to incubate at 37°C for 10-14. Wells were imaged using IncuCyte S3 Live Cell Imaging Analysis System (Sartorius) with a 4X objective and stitched to generate a complete well.
Quantification of the cell count was performed using Fiji (ImageJ) macro. The macro processes all samples automatically and is responsible for thresholding images and generating a count of thresholded objects using a size cutoff of 100 pixels. The results are provided for each well in a csv format. The code for the macro can be found at - https://www.dropbox.com/s/nvqjyyah3gyb942/Macro_Brightfield_thresholdCellsandCount.ijm?dl=0
Growth rates for each well were calculated using R package ‘growthcurver’ available for download from CRAN (Sprouffske and Wagner, 2016). The logistic equation describes the population size N at time t as

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}}$$

N_0 denotes population size at the beginning of the growth curve. K refers to the carrying capacity of the environment. r is the intrinsic growth rate of the population, which would occur if there were no restrictions on the total population. Growthcurver finds the best values of K and r based on the data. The values for r (growth rate) are given in hours.

Lineage tracing:
Cells were plated in 24 well plates and images were obtained using Incucyte (Sartorius) using whole well imaging at 4X. Cells were imaged for 10 days at an interval of 1 or 1.5 hours. FIJI plugin Manual Tracking was used for lineage tracing. The starting cells were picked based on the ability to form colonies (>20 cells) at the end of tracking. At every cell division, one of the sister cells were tracked, while the other sister cell was given a separate track. Simultaneously, the lineage tree was generated manually and each cell's lineage was tracked in relation to the starting cell. Several sister cells were not tracked to limit the number of cell tracks per starting cell to less than 30 and to allow inclusion of colonies from at least 5 starting cells. The tracking data was merged with lineage data using computational analysis in R.
For calculating cell parameters (growth rate and speed), we considered the following parameters for the length of tracking

$$Lifespan\ of\ the\ cell\ (hours\ or\ minutes\ ) = time_{cell\ divides\ into\ 2\ daughter\ cells} - time_{parent\ divides\ to\ give\ rise\ to\ 2\ cells}$$

$$speed = \frac{average\ distance\ traveled\ during\ the\ lifespan\ of\ the\ cell\ in\ \mu m}{lifespan\ of\ the\ cell\ in\ minutes}$$

$$growth\ rate = \frac{1}{lifespan\ of\ the\ cell\ in\ hours}$$

Related cells (sisters and cousins) were marked manually during tracking and were used to classify related cells. The cells were grouped based on the axis of division. Sister cells that were present on the left and top were clustered into group 1 and those on right and bottom were clustered into group 2. Unrelated cells were selected across lineages and any cells that were determined to be sisters or cousins were excluded.

Supplemental Figure 1
A. FS4 cells were plated as single cells in 96 well plates and allowed to grow. When clones were established, cells were collected and allowed to invade through the transwell and compared to the non-clonal cells. Error bars represent standard error across 3 technical replicates. Clones marked with '*' were selected for the analysis shown in the manuscript.
B. 1205Lu cells were plated as single cells in 96 well plates and allowed to grow. Once clones were established, cells were collected and allowed to invade through the transwell and compared to the non-clonal cells. Error bars represent standard error across 3 technical replicates. Clones marked with '*' were selected for analysis shown in the manuscript.
C.1205Lu melanoma cells were added to the transwells and allowed to invade. The invading cells were imaged at the bottom of the tissue culture plate and the cumulative number of cells were quantified. Error bars represent standard error across 3 replicates.

D. FS4 melanoma cells were transduced with lentivirus encoding *SEMA3C* ORF or empty vector. Single cell clones were isolated and tested for expression levels of *SEMA3C* by single molecule RNA FISH. Stars represent clones selected for further evaluation based on *SEMA3C* mRNA levels. Error bars represent standard error across 3 technical replicates.

Supplemental Figure 2

A. GFP-tagged early invading and mCherry tagged late invading 1205Lu melanoma cells were imaged for GFP and mCherry expression 2 hours prior to injection in NSG mice. Cells were labeled as positive for GFP/mCherry expression if mean fluorescence intensity of the cells was 1.2 times background intensity in the same channel.

B. 1205Lu early and late invading cells were tagged and mixed in equal proportion and injected in the left flank of NSG mice. The tumor growth was monitored twice weekly. Graph shows tumor volumes measured as 0.5×length×width×width.

C. 1205Lu early and late invading cells were tagged and mixed in equal proportion and injected in the left flank of NSG mice. The tumor growth was monitored twice weekly. Graph shows tumor volumes measured as 0.5×length×width×width. Tumor volumes are shown on log scale.

Supplemental Figure 3

A. 1205Lu melanoma cells transduced with *AAVS* or guide RNA targeting *NKX2.2* were tested for expression of *NKX2.2* mRNA by smFISH.

B. Matrix of distance (measured by 1 - Pearson's *r*) between samples' expression for the union of differentially expressed genes between both cell lines when comparing early invaders to non-invaders.

C. Principal Component Analysis on RNA sequencing data showing variance between 1205Lu *AAVS* and *NKX2.2* knockout cells and early- and late-invading subpopulations.

D. Principal Component Analysis of ATAC sequencing data showing variance between 1205Lu *AAVS* and *NKX2.2* knockout cells.

Supplemental Figure 4

A. Melanoma cells were sorted based on SEMA3C expression. Cells were live-imaged for ~10 days every hour and single cells were tracked manually for cell position, cell division and lineage. Lineages were traced manually from single cells. Cell speed was calculated for each cell using average distance traveled over time.

B. Growth rate was calculated for each cell from the time taken by the cell to divide.

C. For each cell, speed and growth rate was calculated and compared.

D. SEMA3C-high and SEMA3C-low cells were live-imaged for ~10 days every hour and single cells were tracked manually for cell position, cell division and lineage. Sister, cousin and unrelated cells were identified across 10 generations. Graph shows correlation between speed between sister cells.

E. Graph shows correlation of speed between cousin cells.

F. Graph shows correlation of speed between unrelated cells.

G. SEMA3C-high and SEMA3C-low cells were live-imaged for ~10 days every hour and single cells were tracked manually for cell division and lineage. Sister, cousin and unrelated cells were identified across 10 generations. Graph shows correlation between growth rate between sister cells.

H. Graph shows correlation of growth rates between cousin cells.

F. Graph shows correlation of growth rates between unrelated cells.

References:

Alexaki, V.-I., Javelaud, D., Van Kempen, L.C.L., Mohammad, K.S., Dennler, S., Luciani, F., Hoek, K.S., Juàrez, P., Goydos, J.S., Fournier, P.J., et al. (2010). GLI2-mediated melanoma invasion and metastasis. J. Natl. Cancer Inst. *102*, 1148–1159.

Arozarena, I., and Wellbrock, C. (2019). Phenotype plasticity as enabler of melanoma progression and therapy resistance. Nat. Rev. Cancer *19*, 377–391.

Chang, C.A., Jen, J., Jiang, S., Sayad, A., Mer, A.S., Brown, K.R., Nixon, A.M.L., Dhabaria, A., Tang, K.H., Venet, D., et al. (2021). Ontogeny and Vulnerabilities of Drug-Tolerant Persisters in HER2+ Breast Cancer. Cancer Discov.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Dratkiewicz, E., Simiczyjew, A., Pietraszek-Gremplewicz, K., Mazurkiewicz, J., and Nowak, D. (2019). Characterization of Melanoma Cell Lines Resistant to Vemurafenib and Evaluation of Their Responsiveness to EGFR- and MET-Inhibitor Treatment. Int. J. Mol. Sci. *21*.

Emert, B.L., Cote, C.J., Torre, E.A., Dardani, I.P., Jiang, C.L., Jain, N., Shaffer, S.M., and Raj, A. (2021). Variability within rare cell states enables multiple paths toward drug resistance. Nat. Biotechnol.

Francí, C., Takkunen, M., Dave, N., Alameda, F., Gómez, S., Rodríguez, R., Escrivà, M., Montserrat-Sentís, B., Baró, T., Garrido, M., et al. (2006). Expression of Snail protein in tumor-stroma interface. Oncogene *25*, 5134–5144.

Goyal, Y., Dardani, I.P., Busch, G.T., Emert, B., Fingerman, D., Kaur, A., Jain, N., Mellis, I.A., Li, J., Kiani, K., et al. (2021). Pre-determined diversity in resistant fates emerges from homogenous cells after anti-cancer drug treatment.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Haimovich, G., Ecker, C.M., Dunagin, M.C., Eggan, E., Raj, A., Gerst, J.E., and Singer, R.H. (2017). Intercellular mRNA trafficking via membrane nanotube-like extensions in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *114*, E9873–E9882.

Hapach, L.A., Carey, S.P., Schwager, S.C., Taufalele, P.V., Wang, W., Mosier, J.A., Ortiz-Otero, N., McArdle, T.J., Goldblatt, Z.E., Lampi, M.C., et al. (2021). Phenotypic Heterogeneity and Metastasis of Breast Cancer Cells. Cancer Res. *81*, 3649–3663.

Hoek, K.S., Schlegel, N.C., Brafford, P., Sucker, A., Ugurel, S., Kumar, R., Weber, B.L., Nathanson, K.L., Phillips, D.J., Herlyn, M., et al. (2006). Metastatic potential of melanomas defined by specific gene expression profiles with no BRAF signature. Pigment Cell Res. *19*, 290–302.

Hoek, K.S., Eichhoff, O.M., Schlegel, N.C., Döbbeling, U., Kobert, N., Schaerer, L., Hemmi, S., and Dummer, R. (2008). In vivo switching of human melanoma cells between proliferative and invasive states. Cancer Res. *68*, 650–656.

Kaur, A., Webster, M.R., Marchbank, K., Behera, R., Ndoye, A., Kugel, C.H., Dang, V.M., Appleton, J.,

O'Connell, M.P., Cheng, P., et al. (2016). sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. Nature.

Kaur, A., Ecker, B.L., Douglass, S.M., Kugel, C.H., 3rd, Webster, M.R., Almeida, F.V., Somasundaram, R., Hayden, J., Ban, E., Ahmadzadeh, H., et al. (2019). Remodeling of the Collagen Matrix in Aging Skin Promotes Melanoma Metastasis and Affects Immune Cell Motility. Cancer Discov. *9*, 64–81.

Kiani, K., Sanford, E.M., Goyal, Y., and Raj, A. (2022). Changes in chromatin accessibility are not concordant with transcriptional changes for single-factor perturbations.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Mittal, V. (2018). Epithelial Mesenchymal Transition in Tumor Metastasis. Annu. Rev. Pathol. *13*, 395–412.

Nataraj, N.B., Marrocco, I., and Yarden, Y. (2021). Roles for growth factors and mutations in metastatic dissemination. Biochem. Soc. Trans. *49*, 1409–1423.

Nguyen, B., Fong, C., Luthra, A., Smith, S.A., DiNatale, R.G., Nandakumar, S., Walch, H., Chatila, W.K., Madupuri, R., Kundra, R., et al. (2022). Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. Cell *185*, 563–575.e11.

Olmeda, D., Cerezo-Wallis, D., Riveiro-Falkenbach, E., Pennacchi, P.C., Contreras-Alcalde, M., Ibarz, N., Cifdaloz, M., Catena, X., Calvo, T.G., Cañón, E., et al. (2017). Whole-body imaging of lymphovascular niches identifies pre-metastatic roles of midkine. Nature *546*, 676–680.

Polyak, K., and Weinberg, R.A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat. Rev. Cancer *9*, 265–273.

Quinn, J.J., Jones, M.G., Okimoto, R.A., Nanjo, S., Chan, M.M., Yosef, N., Bivona, T.G., and Weissman, J.S. (2021). Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science *371*.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435.

Shaffer, S.M., Emert, B.L., Reyes Hueros, R.A., Cote, C., Harmange, G., Schaff, D.L., Sizemore, A.E., Gupte, R., Torre, E., Singh, A., et al. (2020). Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. Cell *182*, 947–959.e17.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69–80.

Spencer, S., Gaudet, S., Albeck, J., Burke, J., and Sorger, P. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature.

Sprouffske, K., and Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics *17*, 172.

Symmons, O., and Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. Mol. Cell *62*, 788–802.

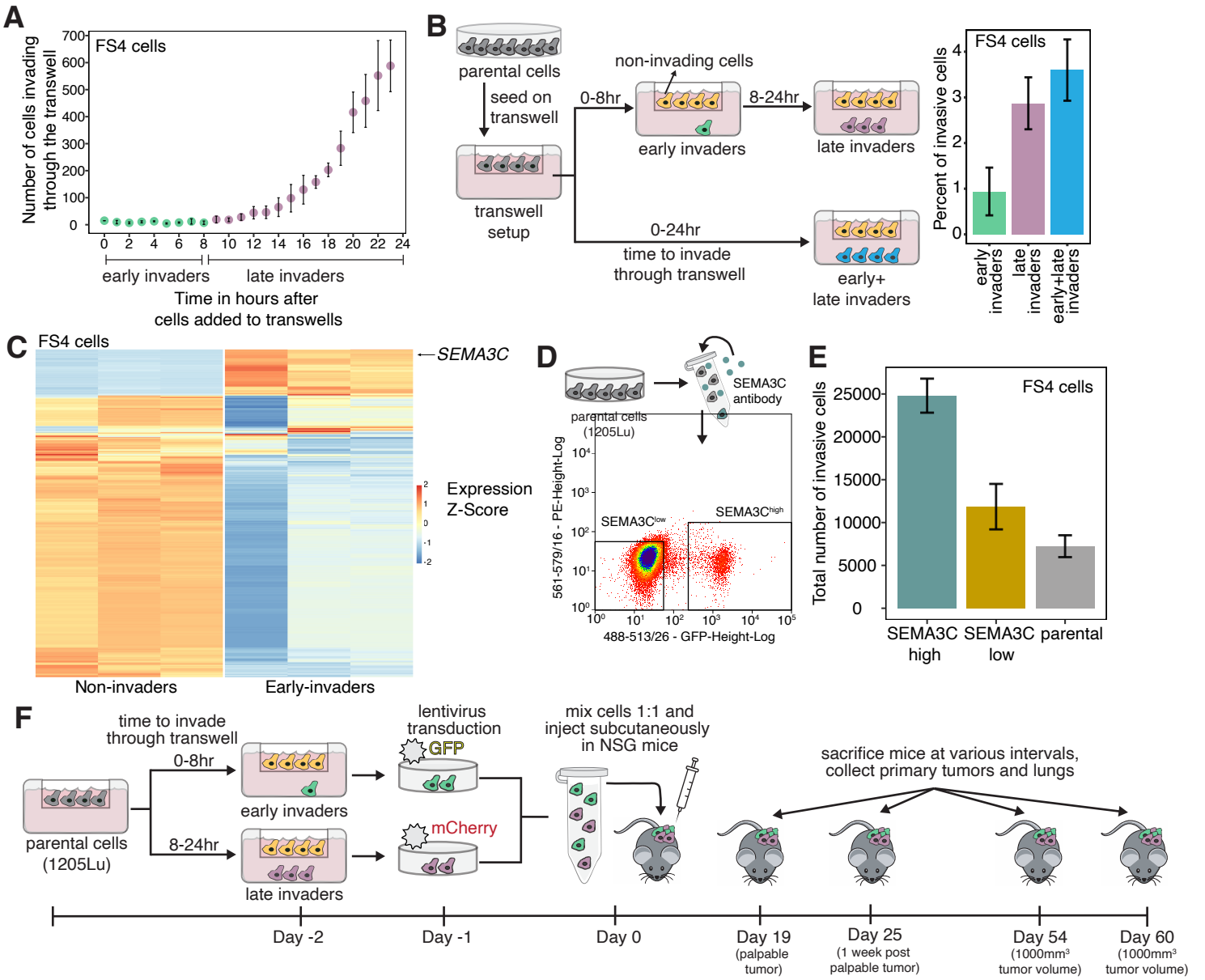
Torre, E.A., Arai, E., Bayatpour, S., Jiang, C.L., Beck, L.E., Emert, B.L., Shaffer, S.M., Mellis, I.A., Fane, M.E., Alicea, G.M., et al. (2021). Genetic screening for single-cell variability modulators driving therapy resistance. Nat. Genet. *53*, 76–85.

Travnickova, J., Wojciechowska, S., Khamseh, A., Gautier, P., Brown, D.V., Lefevre, T., Brombin, A., Ewing, A., Capper, A., Spitzer, M., et al. (2019). Zebrafish MITF-Low Melanoma Subtype Models Reveal Transcriptional Subclusters and MITF-Independent Residual Disease. Cancer Res. *79*, 5769–5784.

Wu, Y., Zanotelli, M.R., Zhang, J., and Reinhart-King, C.A. (2021). Matrix-driven changes in metabolism support cytoskeletal activity to promote cell migration. Biophys. J. *120*, 1705–1717.

Yang, C., Tian, C., Hoffman, T.E., Jacobsen, N.K., and Spencer, S.L. (2021). Melanoma subpopulations that rapidly escape MAPK pathway inhibition incur DNA damage and rely on stress signalling. Nat. Commun. *12*, 1747.

Figure 1


A
FS4 cells
Number of cells invading through the transwell
early invaders
late invaders
Time in hours after cells added to transwells
B
parental cells
seed on transwell
transwell setup
non-invading cells
0-8hr
early invaders
8-24hr
late invaders
0-24hr
time to invade through transwell
early+ late invaders
FS4 cells
Percent of invasive cells
early invaders
late invaders
early+late invaders
C
FS4 cells
SEMA3C
Expression Z-Score
Non-invaders
Early-invaders
D
parental cells (1205Lu)
SEMA3C antibody
SEMA3C^low
SEMA3C^high
561-579/16 - PE-Height-Log
488-513/26 - GFP-Height-Log
E
FS4 cells
Total number of invasive cells
SEMA3C high
SEMA3C low
parental
F
time to invade through transwell
0-8hr
8-24hr
parental cells (1205Lu)
early invaders
late invaders
lentivirus transduction
GFP
mCherry
mix cells 1:1 and inject subcutaneously in NSG mice
sacrifice mice at various intervals, collect primary tumors and lungs
Day -2
Day -1
Day 0
Day 19 (palpable tumor)
Day 25 (1 week post palpable tumor)
Day 54 (1000mm³ tumor volume)
Day 60 (1000mm³ tumor volume)

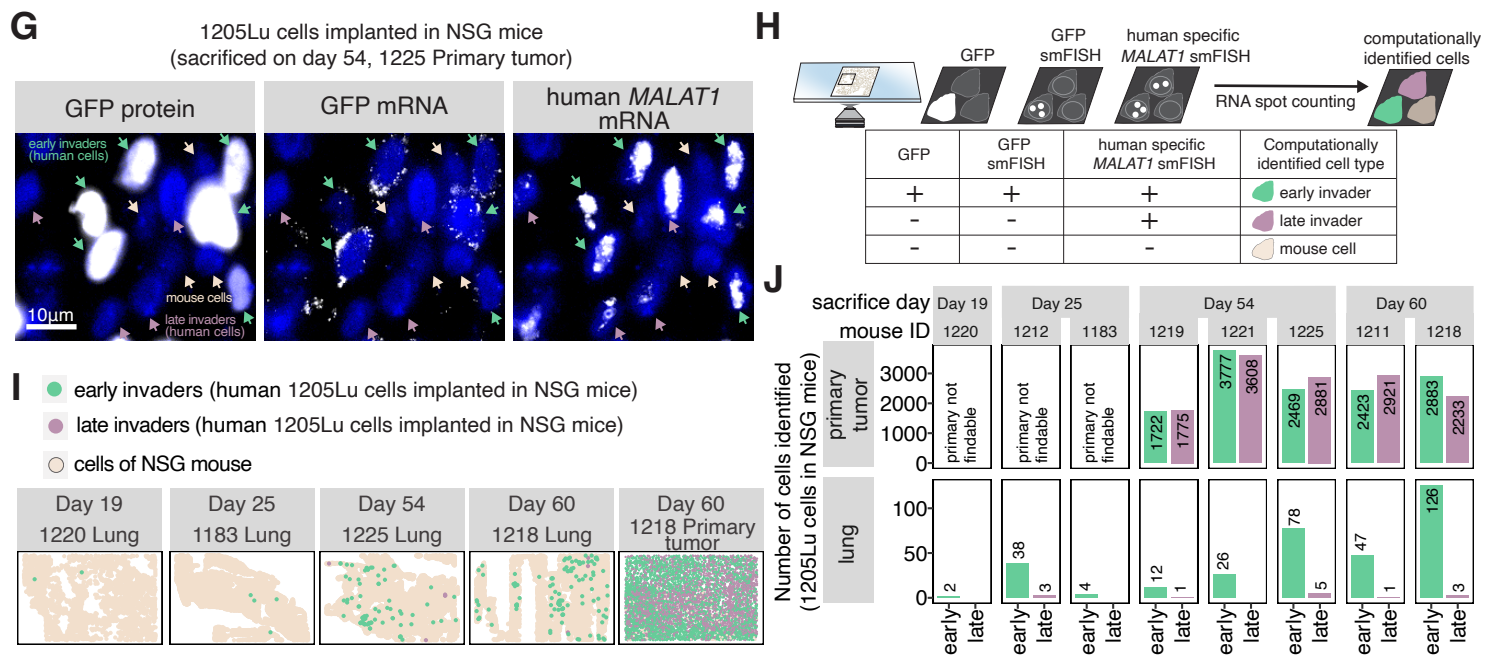

G
1205Lu cells implanted in NSG mice (sacrificed on day 54, 1225 Primary tumor)
GFP protein
GFP mRNA
human MALAT1 mRNA
early invaders (human cells)
mouse cells
late invaders (human cells)
10µm
H
GFP
GFP smFISH
human specific MALAT1 smFISH
computationally identified cells
RNA spot counting
GFP | GFP smFISH | human specific MALAT1 smFISH | Computationally identified cell type
+ | + | + | early invader
- | - | + | late invader
- | - | - | mouse cell
I
early invaders (human 1205Lu cells implanted in NSG mice)
late invaders (human 1205Lu cells implanted in NSG mice)
cells of NSG mouse
Day 19 1220 Lung
Day 25 1183 Lung
Day 54 1225 Lung
Day 60 1218 Lung
Day 60 1218 Primary tumor
J
sacrifice day
mouse ID
Day 19
Day 25
Day 54
Day 60
1220
1212
1183
1219
1221
1225
1211
1218
primary tumor
lung
Number of cells identified (1205Lu cells in NSG mice)
primary not findable
1722
1775
3777
3608
2469
2881
2423
2921
2883
2233
2
38
3
4
12
1
26
78
5
47
1
126
3
early
late

Figure 2

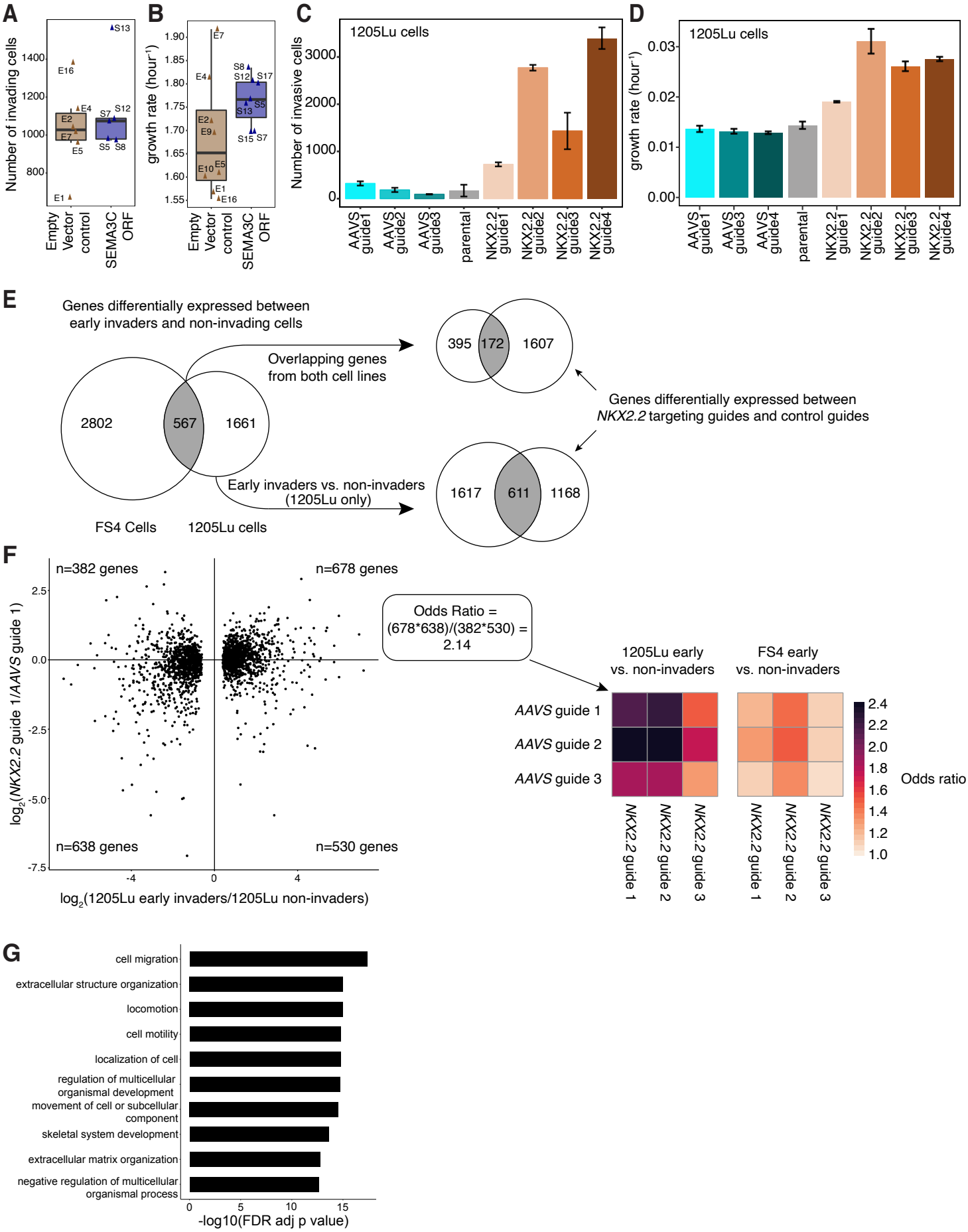

A
Number of invading cells
Empty Vector control
SEMA3C ORF
B
growth rate (hour⁻¹)
C
1205Lu cells
Number of invasive cells
AAVS guide1
AAVS guide2
AAVS guide3
parental
NKX2.2 guide1
NKX2.2 guide2
NKX2.2 guide3
NKX2.2 guide4
D
1205Lu cells
growth rate (hour⁻¹)
E
Genes differentially expressed between early invaders and non-invading cells
Overlapping genes from both cell lines
2802
567
1661
FS4 Cells
1205Lu cells
395
172
1607
Genes differentially expressed between NKX2.2 targeting guides and control guides
Early invaders vs. non-invaders (1205Lu only)
1617
611
1168
F
n=382 genes
n=678 genes
n=638 genes
n=530 genes
log2(NKX2.2 guide 1/AAVS guide 1)
log2(1205Lu early invaders/1205Lu non-invaders)
Odds Ratio = (678*638)/(382*530) = 2.14
1205Lu early vs. non-invaders
FS4 early vs. non-invaders
AAVS guide 1
AAVS guide 2
AAVS guide 3
NKX2.2 guide 1
NKX2.2 guide 2
NKX2.2 guide 3
Odds ratio
G
cell migration
extracellular structure organization
locomotion
cell motility
localization of cell
regulation of multicellular organismal development
movement of cell or subcellular component
skeletal system development
extracellular matrix organization
negative regulation of multicellular organismal process
-log10(FDR adj p value)

Figure 3

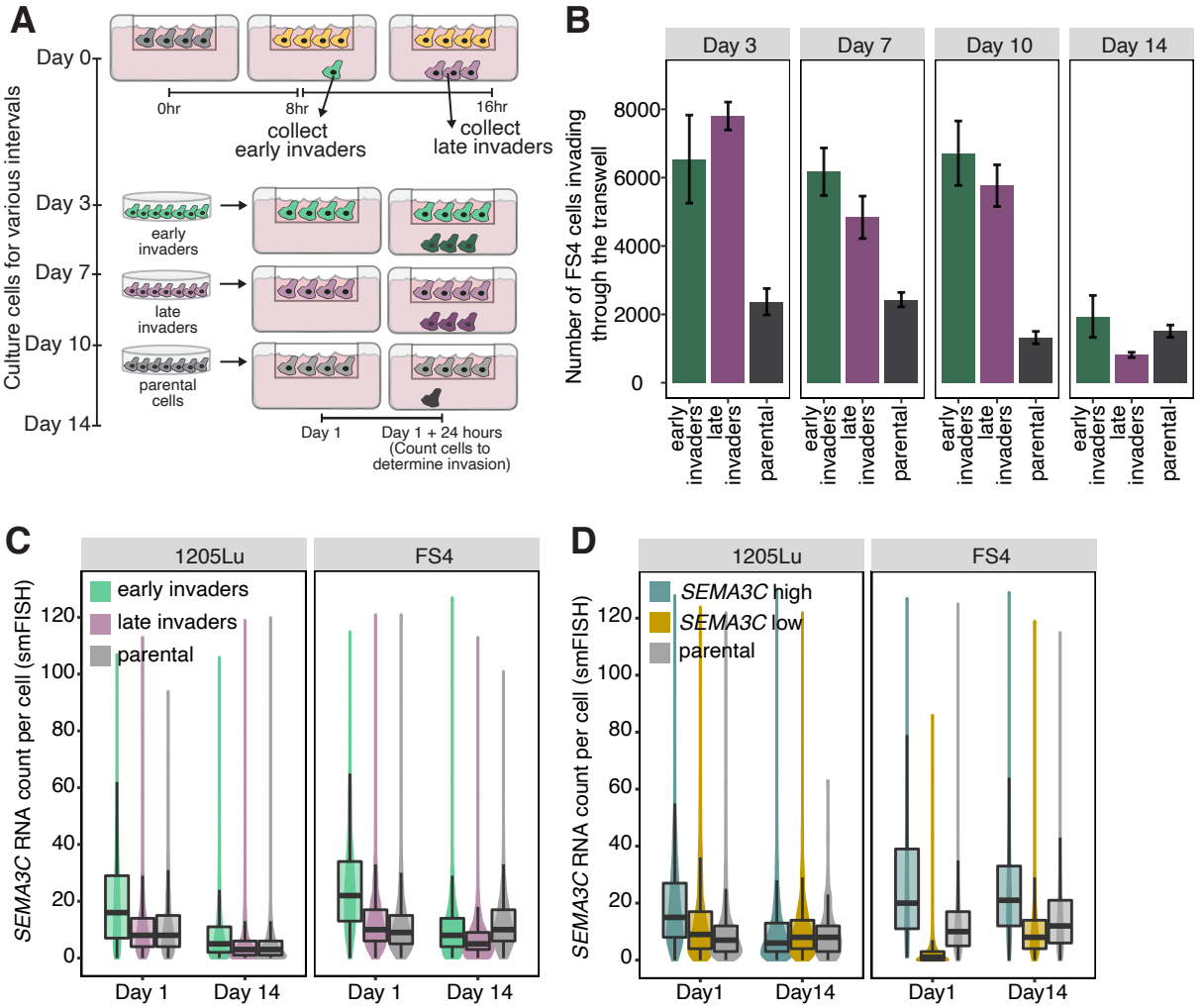
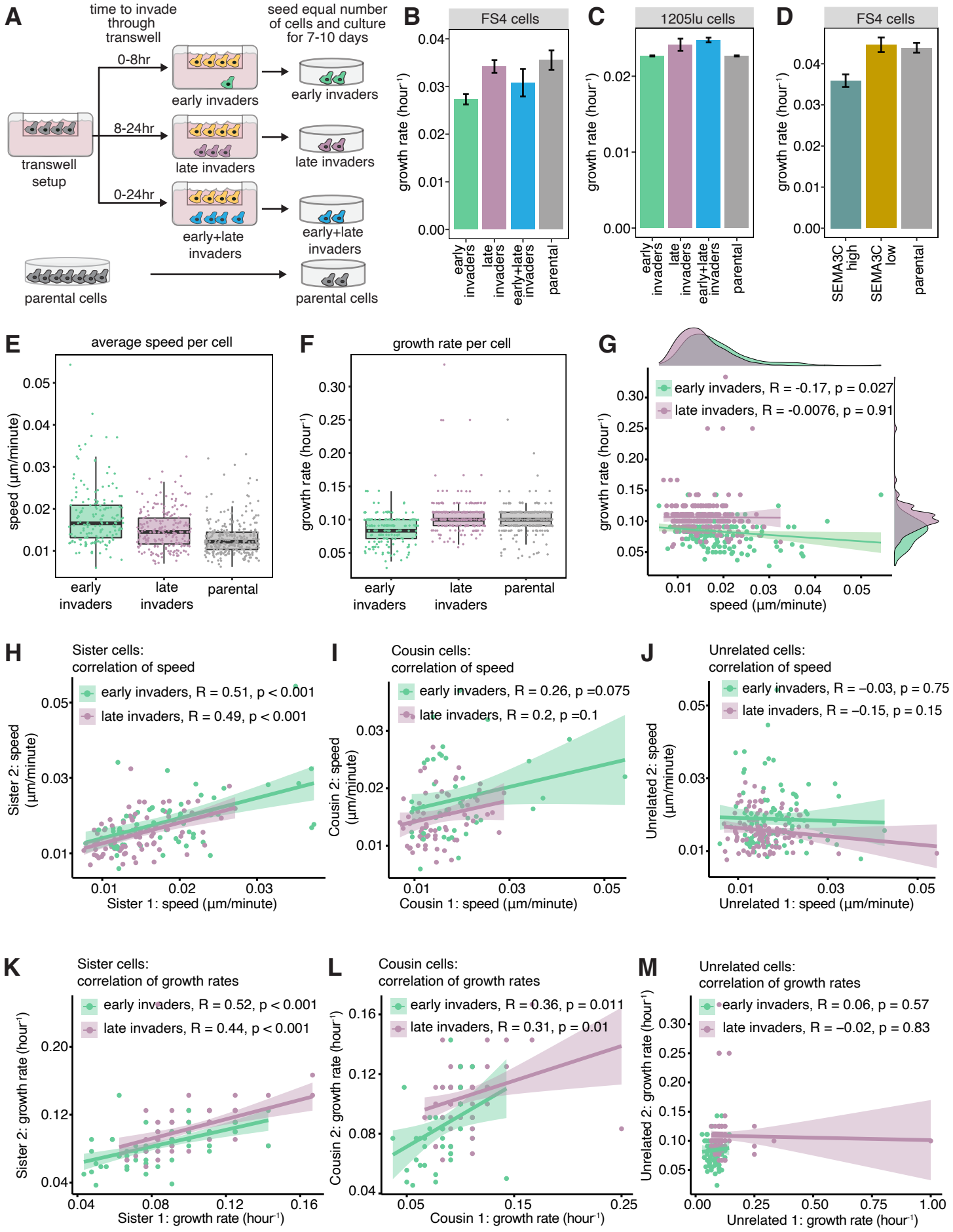

A
Day 0
0hr
8hr
16hr
collect
early invaders
collect
late invaders
Culture cells for various intervals
Day 3
early
invaders
Day 7
late
invaders
Day 10
parental
cells
Day 14
Day 1
Day 1 + 24 hours
(Count cells to
determine invasion)
B
Day 3
Day 7
Day 10
Day 14
Number of FS4 cells invading
through the transwell
8000
6000
4000
2000
0
early
invaders
late
invaders
parental
C
1205Lu
FS4
early invaders
late invaders
parental
SEMA3C RNA count per cell (smFISH)
120
100
80
60
40
20
0
Day 1
Day 14
D
1205Lu
FS4
SEMA3C high
SEMA3C low
parental
Day1
Day14

Figure 4


A
time to invade through transwell
seed equal number of cells and culture for 7-10 days
0-8hr
8-24hr
0-24hr
transwell setup
early invaders
late invaders
early+late invaders
parental cells
B
FS4 cells
growth rate (hour⁻¹)
C
1205lu cells
D
FS4 cells
SEMA3C high
SEMA3C low
parental
E
average speed per cell
speed (µm/minute)
F
growth rate per cell
G
early invaders, R = -0.17, p = 0.027
late invaders, R = -0.0076, p = 0.91
H
Sister cells: correlation of speed
early invaders, R = 0.51, p < 0.001
late invaders, R = 0.49, p < 0.001
Sister 2: speed (µm/minute)
Sister 1: speed (µm/minute)
I
Cousin cells: correlation of speed
early invaders, R = 0.26, p =0.075
late invaders, R = 0.2, p =0.1
Cousin 2: speed (µm/minute)
Cousin 1: speed (µm/minute)
J
Unrelated cells: correlation of speed
early invaders, R = −0.03, p = 0.75
late invaders, R = −0.15, p = 0.15
Unrelated 2: speed (µm/minute)
Unrelated 1: speed (µm/minute)
K
Sister cells: correlation of growth rates
early invaders, R = 0.52, p < 0.001
late invaders, R = 0.44, p < 0.001
Sister 2: growth rate (hour⁻¹)
Sister 1: growth rate (hour⁻¹)
L
Cousin cells: correlation of growth rates
early invaders, R = 0.36, p = 0.011
late invaders, R = 0.31, p = 0.01
Cousin 2: growth rate (hour⁻¹)
Cousin 1: growth rate (hour⁻¹)
M
Unrelated cells: correlation of growth rates
early invaders, R = 0.06, p = 0.57
late invaders, R = −0.02, p = 0.83
Unrelated 2: growth rate (hour⁻¹)
Unrelated 1: growth rate (hour⁻¹)