

Genetic heterogeneity within collective invasion packs drives leader and follower cell phenotypes

Elizabeth L. Zoeller^{*1}, Brian Pedro^{*1}, Jessica Konen^{1,^}, Bhakti Dwivedi², Manali Rupji², Niveda Sundararaman³, Lei Wang⁴, John R. Horton⁵, Chaojie Zhong⁹, Benjamin G. Barwick^{8,9}, Xiaodong Cheng⁵, Elisabeth D. Martinez^{4,6}, Matthew P. Torres³, Jeanne Kowalski^{2,7}, Adam I. Marcus^{2,8,#}, Paula M. Vertino^{2,9,#,^^}

¹Graduate Program in Cancer Biology, Emory University, Atlanta, GA 30322, ²Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA, ³School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, ⁴Hamon Center for Therapeutic Oncology Research, UT Southwestern Medical Center, Dallas, TX 75390, ⁵Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, ⁶Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX 75390, ⁷Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, ⁸Department of Hematology and Medical Oncology, Emory University, Atlanta, GA 30322, ⁹Department of Radiation Oncology, Emory University, Atlanta, GA 30322

*Contributed equally to this work

^Present address: Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

^^Present address: Department of Biomedical Genetics, University of Rochester Medical Center, Rochester, NY, 14642

#To whom correspondence should be addressed:

Paula M. Vertino, Paula_Vertino@URMC.Rochester.edu; Adam I. Marcus, aimarcu@emory.edu

Keywords: metastasis, lung cancer, RNA-seq, mutations, epigenetics, plasticity

Abbreviations

ARP3 actin related protein 3
FACS fluorescence-activated cell sorting
FPS Function Probability Score
H3K4me3 histone 3 lysine 4 trimethylation
H3K4me2 histone 3 lysine 4 dimethylation
H3K4me1 histone 3 lysine 4 monomethylation
K240R lysine to arginine substitution at residue 240
KDM5B lysine demethylase 5B
L685W leucine to tryptophan substitution at residue 685
NSCLC non-small cell lung cancer
PTM post-translational modification
SaGA Spatiotemporal Genomic and Cellular Analysis
SASA solvent accessible surface area
VAF variant allele frequencies

Summary statement

Image-guided capture and sequencing of invasive cancer cells reveals novel leader- and follower-specific gene mutations, highlighting an important relationship between genomic and phenotypic heterogeneity in collective invasion.

Abstract

Collective invasion, the coordinated movement of cohesive packs of cells, has become recognized as a major mode of metastasis for solid tumors. These packs are phenotypically heterogeneous and include specialized cells that lead the invasive pack and others that follow behind. To better understand how these unique cell types cooperate to facilitate collective invasion, we analyzed transcriptomic sequence variation between leader and follower populations isolated from the H1299 non-small cell lung cancer cell line using an image-guided selection technique. We now identify fourteen expressed mutations that are selectively enriched in leader or follower cells, suggesting a novel link between genomic and phenotypic heterogeneity within a collectively invading tumor cell population. Functional characterization of two phenotype-specific candidate mutations shows that ARP3 enhances collective invasion by promoting the leader cell phenotype and that wild-type KDM5B suppresses chain-like cooperative behavior. These results demonstrate an important role for distinct genetic variants in establishing leader and follower phenotypes and highlight the necessity of maintaining a capacity for phenotypic plasticity during collective cancer invasion.

Introduction

Metastatic disease is the cause of 90% of deaths among cancer patients (Mehlen and Puisieux, 2006). In non-small cell lung cancer (NSCLC), which comprises 80-85% of all lung cancer diagnoses, metastases are commonly observed in the bones, lungs, brain, liver, and adrenal glands. Patients presenting with metastatic disease have significantly worse prognoses than those with early-stage disease; for example, the 5-year survival rate for stage I NSCLC is 55%, while stage IV disease (in which distant metastases are present) carries a mere 4% five-year survival rate (American Cancer, 2016). Successful colonization of distant sites requires that cells from the primary tumor gain the capacity to invade through the surrounding basement membrane, travel through the bloodstream or lymphatic system, and ultimately expand to establish colonies at the metastatic site (Efferth and Leber, 2009; Fidler, 2003).

Cells migrate through the microenvironment via multiple mechanisms. A classic example is the epithelial-to-mesenchymal transition, where cells lose expression of epithelial features such as E-cadherin and gain expression of mesenchymal proteins including vimentin and N-cadherin. This shift is thought to promote cellular detachment and enable cancer cells to undergo single-cell invasion. In contrast, collective cell migration refers to the coordinated movement of a group of cohesive cells (Friedl and Gilmour, 2009). This phenomenon is well-described in embryonic development and wound healing, and histological evidence from primary patient tumor samples (Westcott et al., 2015; Cheung et al., 2013; Kabla, 2012; Friedl et al., 2012; Ewald et al., 2012), mouse models of metastasis (Gilbert-Ross et al., 2017; Richardson et al., 2018), and 3-D cultures (Friedl et al., 1995; Hegerfeldt et al., 2002; Nabeshima et al., 1999; Cheung et al., 2013; Konen et al., 2017; Leighton et al., 1960) suggest that cells from solid tumors often migrate and invade in cohesive packs as well. These collective invasion packs and streams vary in width, shape, and cell number, as well as in the mechanisms guiding their movement (Friedl, 2004; Alexander et al., 2008, 2013; Friedl and Wolf, 2010; Haeger et al., 2014; van Zijl et al., 2011; Pandya et al., 2017).

Understanding the mechanisms that underlie the outgrowth of metastatic clones is further complicated by the heterogeneous mix of cell populations within each tumor. This intratumor heterogeneity arises from cell-to-cell variation in the genetic background each expanding to create a unique subclonal population (McGranahan and Swanton, 2017). Superimposed upon this subclonal genomic heterogeneity is the potential for epigenetic heterogeneity reflected in variations in gene expression even among genetically identical cells. Recently, multiregional sequencing of primary lung tumors characterized the intratumor heterogeneity of NSCLC and showed that upwards of 24% to 30% of mutations went undetected in at least one sampling region, demonstrating that almost a third of mutations are occurring in spatially distinct subclonal populations and may have been missed in broad scale

data based on single sampling, such as those analyzed as part of the TCGA project (de Bruin et al., 2014; Zhang et al., 2014). These unique subpopulations may be endowed with properties that enhance attributes beneficial to tumor cells, such as resistance to drug therapy or the ability to invade and metastasize (Waclaw et al., 2015; Lipinski et al., 2016; McPherson et al., 2016; Campbell et al., 2010). For instance, the clonal profile of metastatic disease often does not reflect the profile of the primary tumor, but instead includes one or just a few subpopulations from the primary site (McPherson et al., 2016; Rinner et al., 2012; Gerlinger et al., 2012; Cheung et al., 2016).

One example of phenotypic heterogeneity associated with invasive behavior includes rare, specialized leader cells that lead collective invasive packs, and follower cells that adhere to and follow behind the leaders, both of which cooperate to achieve collective invasion (Konen et al., 2017; Cheung et al., 2013; Westcott et al., 2015; Haeger et al., 2015; Pandya et al., 2017). We developed a novel platform (**S**patiotemporal **G**enomic and **C**ellular **A**nalysis, or **SaGA**) to isolate specific leader and follower cell populations from collectively invading NSCLC cells (Konen et al., 2017). Characterization of these cell types revealed that isolated follower cells are highly proliferative but poorly invasive, while isolated leader cells are highly invasive, but poorly proliferative (Konen et al., 2017). These cellular sub-types cooperate through an atypical angiogenic signaling pathway that is dependent upon VEGF. Previous data suggest a symbiotic relationship, in which both leader and follower cells are necessary for collective invasion to proceed successfully; however, key questions remain as to what drives the biology and emergence of leader and follower cells from a tumor cell population.

In this study, we aimed to elucidate the role of genetic heterogeneity on collective invasion. We analyzed invading leader and follower populations arising from a common H1299 parental NSCLC cell line grown as 3-D spheroids. Strikingly, this revealed mutational landscapes that differ significantly between leader and follower cells, including several expressed mutations that were found exclusively in one cell type or the other. To our knowledge this is the first identification of leader- and follower-specific gene mutations within the same collectively invading tumor cell population.

Results

Leader and follower cell populations contain distinct mutational profiles.

We previously developed the SaGA technique for isolation of leader and follower cells from collectively invading packs of human NSCLC cancer cells (Konen et al., 2017). Briefly, cells expressing Dendra2 green-to-red photoconvertible fluorescent protein are formed into multicellular spheroids, embedded in Matrigel and allowed to invade for 24 hours. Leader or follower cells are then selected based upon physical positioning within invasive chains, optically highlighted by photoconversion using a 405nm laser, and separated by fluorescence-

activated cell sorting (FACS) (Fig. 1A). Using this approach, we isolated three follower populations and two leader populations from H1299 parental cells. Following expansion of each population in 2-D culture, RNA-seq was performed in triplicate, using three separate passages of parental H1299 cells, the three separately isolated populations of follower cells, and the two separately isolated populations of leader cells (including two passages of one of the leader populations). Sequence variants were determined for each population (leader, follower, parental) independently by mapping the RNA-seq profiles to human reference genome Hg19 (GRCh37), resulting in a total of 6240 filtered variants combined in the three populations (see Methods for details). Notably, when comparing variant allele frequencies (VAF) via pairwise t-test analysis, a number of variants were disproportionately present in the leader versus follower populations. We therefore further filtered for those variants that exhibited >20% VAF in either leaders or the followers and <1% in the other (VAF student's t-test p-value <0.01 between leaders and followers). For the purposes of this study, we further excluded those located in 5' or 3' UTRs and known SNPs. Application of these criteria identified fourteen missense mutations – six leader-specific and eight follower-specific (Table 1). This represents the first identification of leader- and follower-specific mutations within the collective invasion pack.

Given the cell type specificity, we hypothesized that these mutations could be key contributors to the emergence of leader vs. follower phenotypes from the parental population. To test this, we chose two candidate mutations for further study – one leader-enriched, *ACTR3* chr2:114699797 A to G, which results in a mutation in ARP3 at lysine 240 (ARP3 K240R), and one follower-enriched, *KDM5B* chr1:202715414 A to C, resulting in a mutation at L685 (KDM5B L685W) (Fig. 1B). We first confirmed these mutations by Sanger sequencing of genomic DNA and cDNA from the parental H1299 population as well as the isolated leader and follower populations (Fig. S1). Both variants were detectable at subclonal levels in genomic DNA, indicating that they were unlikely to have arisen *de novo* during transcription, but represented a subpopulation of genomic alleles in the parental population. Moreover, the selectivity for the leader or follower population observed at the RNA levels was preserved at the genomic DNA level (Fig. S1; Fig. 1C). Analysis of ARP3 and KDM5B expression in H1299 parental, leader and follower cells showed that ARP3 mRNA and protein levels were comparable between the populations (Fig. 1D,E). KDM5B mRNA and protein were significantly reduced in follower cells relative to leaders (Fig. 1D,E). Despite the reduced overall levels, the KDM5B mRNA in follower cells retained the ~2:1 ratio of wild-type vs. mutant observed at the genomic DNA level (Fig. 1C) suggesting that the follower-enriched KDM5B L685W mutation is expressed. Similarly, while there was some variation in the frequency of both mutations in the parental population between methods and DNA/RNA samples isolated at different times, there was little variation in the allelic balance in leader and follower

populations, which maintained a consistent frequency of their respective mutations at both DNA and RNA level, suggesting that there is no allelic bias in the expression of the mutant version in either case. Thus, our selection of leader and follower cells based on phenotypic criteria also selected for subpopulations with distinct allelic balance of expressed mutations.

Predicted functional impact of the leader-enriched ARP3 K240R mutation.

We next sought to characterize the potential impact of the observed mutations. We started with the leader-enriched *ACTR3* mutation, which results in a K to R shift in ARP3 (K240R). ARP3 is a key component of the Arp2/3 complex that helps facilitate cellular migration by promoting lamellipodia protrusion (Goley and Welch, 2006). Overexpression of ARP3 has been correlated with invasion, metastasis, and poor survival in multiple cancer types, including gastric, colorectal, liver, and gallbladder (Zheng et al., 2008; Iwaya et al., 2007; Lv et al., 2018; Yang et al., 2013). Furthermore, multiple mass spectrometry studies indicate ARP3 K240 as a post-translational modification (PTM) site, with evidence of both ubiquitylation and acetylation (Mertins et al., 2013; Wagner et al., 2011, 2012) (Fig. 2A, inset). To evaluate the functional impact of the K240R mutation on the ARP3 protein, we used SAPH-ire (Structural Analysis of PTM Hotspots) (Torres et al., 2016), which predicts the functional potential of PTMs in protein families that have existing experimental and 3D structure data. SAPH-ire calculates a Function Probability Score (FPS) using a neural network model trained with an array of protein sequence and PTM-specific features extracted from PTMs with established functional impact. Consistent with these data, K240 had among the highest FPS values of all known modified residues within the ARP3 protein family and was among the top 90% of PTMs with well-established functional significance (i.e. 4 or more publications) across all protein families (Fig. 2A). SAPH-ire also revealed six experimental ubiquitylation sites in the ARP3 protein family between residues alignment positions 1298 – 1315, four of which correspond to ubiquitylation of ARP3 specifically (K240, K244, K251, and K254). Of these, K240 had the highest mean solvent accessible surface area (SASA) and was also proximal to a protein interaction interface (Fig. 2B). The high solvent accessibility, proximity to a protein-protein interface, and predicted functional impact of the ARP3 K240 site suggest that this mutation could indeed be altering the activity of ARP3 in leader cells.

ARP3 K240R promotes invasion and leader cell behavior.

We sought to determine the function of the leader-specific *ACTR3* mutation by introducing this mutation into a follower cell background and determining the impact on collective invasion and the outgrowth of cells with leader behavior. Given the relatively low VAF (23.4%) of mutant *ACTR3* in leader cell DNA/RNA, we sought to replicate the leader cell *ACTR3* allelic balance by employing a rescue approach. ARP3 levels were first stably knocked

down using two separate short hairpin RNAs (Fig. S2A,B), including one (shACTR3 #2) targeted to the 5' UTR of endogenous *ACTR3*. Knockdown of ARP3 significantly reduced 3-D invasion in H1299 parental, leader and follower cells compared to pLKO.1 controls (Fig. S2C,D). Using a sulforhodamine B (SRB) assay to measure cell growth, we found little difference upon ARP3 knockdown until 96 hours in the leader and follower populations, and 120 hours in the parental population, when proliferation was decreased (Fig. S2E).

To test the functional consequences of ARP3 mutation, follower cells expressing shACTR3 #2 were then 'rescued' by stably expressing empty vector, mCherry-tagged wild-type ARP3, or mCherry-tagged ARP3 K240R, under the control of the moderate activity UBC promoter (Qin et al., 2010). Higher expression was achieved for ARP3 K240R compared to wild-type ARP3 (Fig. 2C, upper bands), suggesting that ARP3 K240R may be more stable than the wild-type protein. When grown as a spheroid, embedded in a Matrigel matrix, and allowed to invade for 24 hours (Fig. 2D), ARP3 knockdown follower cells reconstituted with ARP3 K240R, and to a lesser extent those reconstituted with wild-type ARP3, exhibited significantly higher invasive area compared with unmodified followers or those reconstituted with empty vector (Fig. 2E). This indicates that ARP3 expression, and especially ARP3 K240R expression, can increase invasive capacity even in normally poorly-invasive follower cells.

Next, we sought to examine whether ARP3 K240R could specifically promote leader cell behavior (i.e. facilitate collective invasion and travel at the front of invasive chains) in a heterogeneous population. We created 3-D spheroids comprised of 90% unmodified H1299 followers and 10% ARP3 depleted followers rescued with either empty vector, wild-type ARP3, or ARP3 K240R (Fig. 3A). After 24 hours embedded in Matrigel, we observed a significant increase in invasive area and average number of chains per spheroid, and decreased circularity (indicating more chain-like and less sheet-like invasion) in the mixed spheroids containing 10% ARP3 K240R-rescued followers, as compared with the other three conditions (Fig. 3B,C). To determine whether the ARP3 K240R-rescued followers were in fact leading these invasive chains, we used confocal fluorescence imaging to quantify the fraction of chains that exhibited mCherry-ARP3 K240R-rescued followers in the leader position. In spheroids containing 10% wild-type ARP3-rescued followers, we found rescued cells in the leader position in 53.8% of those chains (95% confidence interval 29.1% to 76.8%; Fig. 3D). By contrast, in spheroids containing 10% ARP3 K240R-rescued followers, rescue cells were found in the leader position in 87.2% of chains (95% confidence interval 78.0% to 92.9%; Fig. 3D). Thus, while both wild-type ARP3-rescued and ARP3 K240R-rescued cells led chains at a higher frequency than expected by random chance, the ARP3 K240R-rescued cells were more efficient in this regard. Together these data indicate that ARP3 K240R confers key leader-like behaviors onto follower cells, including increased invasive capacity, increased numbers of invasive chains, and a greater ability to lead those chains.

As noted above, the K240R-reconstituted ARP3 knockdown cells express higher levels of ARP3 than those reconstituted with the wildtype ARP3. To distinguish the impact of ARP3 dosage from that of ARP3 K240R mutation, we re-created the same rescue cell lines using a CMV promoter (Fig. 3E) and repeated the above experiments creating mixed spheroids with 10% rescued cells (Fig. 3F). At this higher protein expression level, there was no significant difference in invasive area between the groups (Fig. 3G). Additionally, while chain number increased and circularity decreased in spheroids containing either wild-type ARP3-rescued or ARP3 K240R-rescued followers (Fig. 3G), there was no significant difference between the two, suggesting that the functional difference between ARP3 K240R and wild-type ARP3 is mitigated at higher expression levels. Furthermore, both wild-type ARP3-rescued followers (81.5% of chains led; 95% confidence interval 63.3% to 91.8%) and ARP3 K240R-rescued followers (84.2% of chains led; 95% confidence interval 62.4% to 94.5%) promoted leader cell behavior when mixed with 90% unmodified followers (Fig. 3H). The specificity of the effect was further confirmed by mixing experiments with mCherry-empty vector-rescued follower cells, which led only 17.7% of chains (95% confidence interval 6.19% to 41.0%; Fig. 3H), suggesting that the enhanced leader ability of the ARP3 K240R cells was not simply due to the different cell types segregating within the spheroid. Based upon these findings, it appears that increased dosage of ARP3 is sufficient to promote leader-like behavior, and that the selective ability of ARP3 K240R to lead invasive chains and drive collective invasion at lower expression levels may arise, in part, from increased effective dose of ARP3 protein.

Functional impact of the follower-enriched KDM5B L685W mutation.

Subsequently, we sought to assess the effects of a follower-specific mutation on collective invasion. As noted above, the KDM5B L685W mutation was selectively expressed in the parental and follower populations but was absent from the leader population (Fig.1B). KDM5B is a lysine demethylase that catalyzes the removal of di- and trimethylation from histone H3 methylated at lysine 4 (H3K4me₂, H3K4me₃). As an epigenetic regulator, KDM5B is uniquely situated among the identified mutations to influence multiple pathways directly involved in guiding invasive behavior (Table 1) (Han et al., 2017; Zheng et al., 2019; Tam and Weinberg, 2013). Indeed, KDM5B has been implicated in the regulation of invasive and migratory behavior across many cancer types, including hepatocellular, gastric, breast, lung, and prostate (Tang et al., 2015; Wang et al., 2015; Bamodu et al., 2016; Viré et al., 2014; Kuo et al., 2018; Shen et al., 2015; Lu et al., 2015). The point mutation at amino acid 685 observed in followers results in a leucine to tryptophan (L685W) substitution and lies in close proximity to the zinc finger domain, a region necessary for demethylase activity *in vitro* and in cells (Horton et al., 2016a; Yamane et al., 2007) (Fig. 4A). Furthermore, structural modelling of the L685W substitution showed that residue 685 sits in a hydrophobic pocket within KDM5B, near

the binding site of the N-terminal tail of histone H3 (Fig. 4B). As tryptophan has a much bulkier sidechain than leucine, the introduction of this large, aromatic sidechain has the potential to alter or disrupt the interaction of KDM5B with histone H3 (Fig. 4B).

Overall, the follower cells expressed less KDM5B protein than leader cells or the parental H1299 population (Figs. 1E, 4C), and, as noted above, at least a fraction of that (~30%) would be expected to harbor the L685W mutation. Comparison of the global levels of the KDM5 substrates showed that while H3K4me₂ and me₃ were similar across populations, follower cells had reduced global levels of H3K4 monomethylation (H3K4me₁), the end-product of KDM5B catalytic activity, suggesting reduced cellular H3K4 demethylase activity in follower cells versus leader cells (Fig. 4C). To test this directly, we generated parental H1299 cells stably overexpressing HA-tagged wild-type KDM5B or KDM5B L685W and performed *in vitro* demethylation assays on cell lysates (Fig. 4D). Whereas overexpression of wild-type KDM5B trended toward increased cellular H3K4 demethylation activity by approximately 2-fold, overexpression of KDM5B L685W had no impact on endogenous H3K4 demethylation activity despite similar levels of wild-type and mutant expression (Fig. 4D). Taken together these data indicate that the KDM5B L685W mutation seen in the followers is associated with reduced cellular KDM5 activity.

KDM5B L685W modulates collective invasion by tipping the balance toward follower cell behavior.

The above data suggests that KDM5B L685W may confer a partial loss of function/decreased activity on the follower cells, and further that this mutation is selected against in the leader population (addressed in more detail below). We therefore sought to determine how the expression of the L685W might behave in the context of an otherwise wild-type KDM5B background (to mimic the mixed genotype observed in followers) by expressing KDM5B L685W in leader cells and determining the impact on collective invasion in 3-D spheroid culture (Fig. 5A,B). Interestingly, overexpression of wild-type KDM5B in leader cells suppressed their propensity for chain-like invasion, as indicated by an increase in circularity and a decrease in the number of chains relative to cells expressing empty vector (Fig. 5B,C). However, invasive area was similar between leaders expressing empty-vector and wild-type KDM5B, suggesting that the wild-type KDM5B expressing leaders retain invasive capabilities but lose some of the collective (cooperative) behavior manifested as chain-like invasion. By contrast, overexpression of the KDM5B L685W mutant enhanced slightly the chain-like invasion, with spheroid behavior resembling the high chain number and low circularity observed in empty vector expressing leader cells (Fig. 5B,C).

To reconcile these findings, we considered the possibility that heterogeneity in KDM5B expression/activity may be required for cells in a population to take on the cooperative roles

necessary to form an invasive chain. Even in spheroids derived from a purified leader population, such as those shown above, there must be at least some cells that take the trailing position, or evolve follower-like behavior, even in the absence of new genetic events. One interpretation of the above data is that overexpression of wild-type KDM5B may prevent the emergence of cells with follower behavior, thus suppressing chain-like co-operative behavior; by pushing the balance even further towards high KDM5B activity, the reduced population plasticity might decrease the probability that cells can take on the follower role. By contrast, overexpression of KDM5B L685W might allow for greater variation in KDM5B activity, promoting heterogeneity and the emergence of cells with follower behavior, further enhancing collective behavior. If this is the case, then one might predict that ectopic expression of wild-type KDM5B or KDM5B L685W would have relatively little impact on follower cells, which on their own do not exhibit significant cooperative behavior. Indeed, when expressed in follower cells, the further expression of KDM5B or KDM5B L685W had a negligible impact on circularity or the chain number of follower-only spheroids (Fig. 5B,C). These data also indicate that increasing the concentration of wild-type KDM5B is in itself not sufficient to impose leader behavior on follower cells.

The benchmark test for leader cell activity is to determine how such cells behave when mixed with followers in 3-D invasion assays. As noted above for ARP3, leader cells can have a strong effect on chain formation, even when they make up a very small proportion of the total cells in a spheroid. We therefore tested directly how expression of KDM5B L685W affects the ability to lead in a heterogeneous population by creating mixed spheroids comprised of predominantly (75%) unmodified follower cells into which leader cells overexpressing wild-type KDM5B or KDM5B L685W (25%) had been added and monitoring chain-like invasion (Fig. 5D,E). We found again that even in the context of unmodified followers, overexpression of wild-type KDM5B suppressed chain-like behavior relative to control (vector only) leaders. In contrast, expression of KDM5B L685W not only retained but even somewhat enhanced the ability to lead invasive chains compared to those expressing either empty-vector or wild-type KDM5B (Fig. 5D,E). These data suggest that overexpression of KDM5B L685W mutant sustains chain-like cooperative behavior. Taken together, these data support the idea that wild-type KDM5B supports the leader phenotype in part by suppressing emergence of follower-like behavior, while expression of the L685W mutant supports, and even somewhat promotes chain-like invasion.

To further test the impact of KDM5B L6895W population heterogeneity on collective invasion, we mixed leader lines expressing empty vector, wild-type KDM5B, and KDM5B L685W with unmodified leaders expressing mCherry at different proportions and monitored chain-like invasion in 3-D spheroids (Fig. S3). As both lines already express Dendra2, we quantified the fraction of chains led by red/magenta positive cells (mCherry, unmodified

leaders) or green-only cells (KDM5B modified lines) (Fig. S3B). We found that overexpression of wild-type KDM5B again suppressed chain formation in spheroids (Fig. S3E, 100% KDM5B WT) relative to empty vector, which was recovered proportionally as the fraction of unmodified leaders in the spheroid increased (Fig. S3E). Despite this suppression in total chains, the fraction that were led by a green-only, wild-type KDM5B overexpressing leader was roughly proportional to its abundance in the mixed spheroid (compare Fig. S3 panels D and E, 90% vs. 10% wild-type KDM5B). By contrast, leaders expressing L685W showed no suppression of chain-like behavior relative to vector-only cells (Fig S3E, 100% KDM5B L685W), and exhibited a similar frequency of green-only vs. magenta (unmodified leaders) in the lead position regardless of their relative proportions. These data support the idea that overexpression of wild-type KDM5B may suppress the emergence of cooperative behavior necessary to form a chain, such cells nevertheless retain the capacity to lead in a mixed population.

SNP analysis suggests distinct leader and follower cell lineages.

If high expression of wild-type KDM5B enforces the leader phenotype and/or suppresses the emergence of an alternate phenotype (e.g. followers), and this has an impact on collective behavior, then one might predict that there would be a selection against the expression of KDM5B L685W in leader cells and selection for KDM5B L685W in follower cells independently captured from H1299 spheroids relative to the parental population. Fortuitously, a common SNP (rs1141108, chr1: 202715284, G>A) is located within the same exon as the KDM5B mutation, and the H1299 cell line was determined to be heterozygous for this SNP in our initial sequence analysis. Targeted deep resequencing of an amplicon including the region containing both the mutation (chr1:202715414) and the SNP (chr1: 202715284) in genomic DNA thus enabled us to trace the relationship between allelic balance and the frequency of the KDM5B L685W mutation in each cell population (average depth= 348,327 reads).

We first found that approximately one third of alleles in each population carried the A variant and two thirds carry the G variant at rs1141108, leading us to conclude that there are 3 copies of *KDM5B* and/or chromosome 1 in our strain of H1299 cells and that this ploidy is maintained across the three populations. The *ACTR3* mutation that gives rise to ARP3 K240R on chromosome 2 also occurs in approximately one third of reads from genomic DNA in leader cells (Fig. 1B). Focal copy number alterations in the genomic regions surrounding *ACTR3* and *KDM5B* have not been detected in H1299 cells by SNP copy number analyses (COSMIC Cell Lines Project: https://cancer.sanger.ac.uk/cell_lines). These data suggest that our strain of H1299 cells are functionally triploid. We further ascertained that the *KDM5B* variant (chr1:202715414: A>C) giving rise to the L685W mutation resides exclusively in *cis* with the SNP rs1141108 G allele (99.7% concordance across 267,414 total reads containing the

mutation). The analysis further confirmed the proportions of *KDM5B* wild-type vs. mutant alleles shown in Fig. 1, including the more variable mutation frequency observed among different isolates of the parental population, the complete absence of the *KDM5B* L685W mutation ($<0.00001\%$) in the leader cells and the very consistent $\sim 30\%$ in the follower cells (Fig. 6A). Additionally, the relative proportions of the wild-type and variant expressed at the mRNA level were largely reflective of the proportion at the genomic DNA level across all three populations, again suggesting that there was no allelic bias in expression of the wild-type versus mutant forms of *KDM5B* (Fig. 6A).

The finding that the *KDM5B* L685W mutation was exclusive to the G allele while the A allele was exclusively wild-type allowed us to trace the relative proportions of the three genotypes (WT/A, WT/G, mutant/G) across the parental, leader, and follower populations. These data showed that whereas 4-12% of the parental population carries the *KDM5B* mutant/G allele, this allele was excluded from the leader population, which exhibited essentially none of the *KDM5B* mutant/G alleles (Fig. 6B), but was nearly 2-fold enriched in the isolated followers (30%) ($p=0.008$; ANOVA plus Tukey's post-hoc correction) and approached the expected frequency (if each allele is independently sorted). Taken together these data suggest that whereas the parental population varies in the fraction of cells containing the mutant/SNP G allele, growth in 3-D culture, SaGA-based capture and subsequent expansion of leader cells selectively enriches for cells expressing only wild-type *KDM5B*, whereas that of followers selects for a population in which nearly every cell contains (and expresses) one copy of the mutant *KDM5B* L685W allele.

Discussion

The greatest threat to cancer patient mortality is the metastatic spread of tumor cells from the primary site (Mehlen and Puisieux, 2006). Collective migration and invasion are major contributors to the dispersion of metastatic cancer cells (Friedl and Gilmour, 2009; Bronsert et al., 2014; Haeger et al., 2015; Mayor and Etienne-Manneville, 2016). Collective invasion is typified by the coordinated movement of a group of cohesive cells, often including multiple heterogeneous cell populations with specialized functions. One well-studied example of collective invasion is that of chain-like invasion, in which specialized leader cells lead groups of cells termed follower cells, out of the tumor (Konen et al., 2017; Mayor and Etienne-Manneville, 2016; Haeger et al., 2015), with both populations playing important roles in the process of invasion. Until now, studies of the distinct populations within invasive chains have been limited by the inability to separate these populations. Development of the SaGA technique (Konen et al., 2017), enabled us to independently analyze leader and follower cells with different phenotypes to gain insight into population dynamics and the emergence of populations that differ in cell behavior. Our results identify a set of expressed mutations that

define leader and follower cells, representing, to our knowledge, the first known instance of distinct mutations as contributors to the leader/follower phenotypes within collectively invading packs.

We confirmed the importance of the leader cell-enriched mutation ARP3 K240R to the invasive leader cell phenotype by introducing it into a population of non-invasive H1299 follower cells. Both in pure spheroids and when mixed with 90% unmodified followers, ARP3 K240R-expressing followers displayed increased ability to invade and lead collective chains at both lower and higher protein expression levels. Rescue with wild-type ARP3 also conferred leader cell behavior, but only when expressed at supra-physiologic levels. One potential explanation is that ARP3 K240R increases the effective dosage of ARP3 protein, essentially recapitulating ARP3 overexpression even at low expression. Indeed, ARP3 K240R accumulated to higher levels than wild-type ARP3 when exogenously expressed from the same vector. The K240R mutation might interfere with ubiquitylation at K240, resulting in either decreased ARP3 turnover, or enhanced ARP3 activity. Ubiquitylation at K240 has previously been observed by mass spectroscopy in multiple human cell lines as well as mouse tissue, and K240R was predicted by SAPH-ire to have a high likelihood of functional consequence. Further experiments will be necessary to determine whether ARP3 K240R is indeed resistant to ubiquitylation, and whether this impacts the activity of ARP3 in leader cells. ARP3 is a key subunit of the Arp2/3 complex that regulates intracellular actin dynamics in a number of processes, including lamellipodia protrusion during cell motility (Goley and Welch, 2006). Indeed, we observed significantly reduced invasion in parental, leader, and follower cells upon ARP3 knockdown, supporting its importance for cell migration and invasion. Overexpression of Arp2/3 complex subunits including ARP2, ARP3, ARPC2, and ARPC5 has been shown to promote invasion in multiple cancer types including lung, colorectal, glioblastoma, and others (Liu et al., 2013; Zheng et al., 2008; Iwaya et al., 2007; Semba et al., 2006; Kinoshita et al., 2012; Lv et al., 2018; Yang et al., 2013). Our results now further indicate a role for ARP3 as contributing to tumor collective invasion by promoting the leader cell phenotype.

We also investigated the role of the follower cell enriched mutation, KDM5B L685W. Surprisingly, overexpression of wild-type KDM5B in leader cells suppressed invasive chains, as well as the ability to lead invasive chains when mixed with followers. By contrast, overexpression of KDM5B L685W supports chain-like invasion. Interestingly, these data are seemingly at odds with published data showing that overexpression of wild-type KDM5B in cancer cells increases their migration and invasion in 2-D assays (Tang et al., 2015; Wang et al., 2015; Bamodu et al., 2016; Viré et al., 2014; Kuo et al., 2018; Lu et al., 2015), including in H1299 cells (Shen et al., 2015), and may be a reflection of the differences in the assessment of invasion in 2-D, which focuses on the behavior of single cells, and that of cells grown as spheroids and undergoing collective invasion. Collective invasion requires cooperation

between cell phenotypes, and the ability of cells in the population to take on distinct roles upon suspension in 3-D. One interpretation of our data is that overexpression of wild-type KDM5B suppresses chain-like invasion by suppressing phenotypic plasticity, tipping the balance away from cooperative behavior, and that KDM5B L685W is somehow deficient in this regard. KDM5B acts as a transcriptional repressor, and its sustained expression in leader cells may act to promote phenotypic stability by repressing the follower 'program', thus preventing the emergence of the alternate (follower) phenotype. Consistent with this idea, the isolation of cells that take the lead position in 3-D invasive streams from the parental favors those that have retained parental levels of KDM5B and lack the mutation, while cells that take the follower position are selectively enriched for mutant KDM5B and exhibit significantly reduced overall KDM5B expression and a skewed intracellular H3K4me2/3 to H3K4me1 ratio relative to leaders, consistent with reduced/compromised intracellular KDM5 activity. It is also possible that whereas other "dominant-acting" mutations (such as ARP3 and potentially others) directly promote leader behavior, wild-type KDM5B may be a passive participant in the selective outgrowth of leaders. Indeed, wild-type KDM5B expression was not sufficient to drive leader activity in follower cells. Rather, we favor the hypothesis that reduced KDM5B activity/ expression facilitates population plasticity and cooperative behavior. Interestingly, KDM5B was recently identified as a key suppressor of transcriptomic heterogeneity. Whereas KDM5B overexpression enforces phenotypic stability of luminal breast cancer cells, its depletion/inhibition was shown to promote transcriptional plasticity and to facilitate breast cancer cells to overcome therapeutic resistance (Hinohara et al., 2018). Reduced KDM5B expression/activity and the ensuing transcriptomic heterogeneity may also be a necessary step for the emergence of the specialized leader and follower phenotypes that define collectively invading chains. In this regard, it is interesting to note that leader cell cultures tend to be phenotypically stable over time and retain their highly invasive phenotype, whereas follower cultures exhibit a greater degree of phenotypic plasticity, with a tendency to spin off new invasive chains (and hence take on leader characteristics) over time (Konen et al., 2017), suggesting that at least some components of collective behavior are epigenetically driven. At least in some cases, knockdown of KDM5B in H1299 parental cells (but not isolated leaders or followers) promotes an increase in chain number and invasive area (Fig. S4).

Previous studies have suggested that cooperation between phenotypically and genetically distinct subpopulations derived from a common tumor cell population are necessary to promote invasion and/or metastasis *in vivo*. For example, Calbo *et al.* (Calbo et al. 2011) found that neuroendocrine and non-neuroendocrine-like clones derived from a common primary murine SCLC tumor were incapable of completing metastasis on their own, but once re-mixed became highly metastatic. Shared copy number variations confirmed a common clonal origin, but some genetic alterations (e.g. amplification of *Myc/1*) were only

observed in the neuroendocrine-like population. Likewise, Mateo *et al.* (Mateo *et al.*, 2014) found that co-culture of phenotypically distinct subpopulations isolated from the PC-3 prostate cancer cell line promoted invasion and metastasis over either population in isolation. Other studies have shown that the post-hoc admixture of cells engineered to undergo EMT or to express activated HRAS with the non-engineered control population was necessary to fully complete metastasis (Tsuji *et al.*, 2008) or to promote collective motility of breast epithelial cells (Liu *et al.*, 2012). The necessity for heterogeneity in the transcriptional program, even when genetically driven, is an emerging theme, and is consistent with our observations here.

Emerging research regarding how followers cooperate with leaders to promote their invasive capabilities center on concepts such as contact inhibition of locomotion (CIL) and movement along a chemical gradient (Mayor and Etienne-Manneville, 2016; Jolly *et al.*, 2018). CIL occurs when a migrating cell contacts another cell and begins forming protrusions opposite the site of contact to move in the opposite direction (Abercrombie, 1970; Abercrombie and Heaysman, 1953; Mayor and Carmona-Fontaine, 2010). In this case, follower cells contact leader cells, forcing them to polarize and move in a forward direction. Follower cells also regulate signaling to leader cells, often through creation of chemical gradients that motivate leader cells to move in a particular direction (Mayor and Etienne-Manneville, 2016; Theveneau *et al.*, 2010; Malet-Engra *et al.*, 2015; Valentin *et al.*, 2007; Dona *et al.*, 2013; Muinonen-Martin *et al.*, 2014). Indeed, we previously discovered that follower and leader cells engage in a symbiotic relationship, in which follower cells promote the survival and proliferation of leader cells, which in turn secrete VEGFA to promote the motility of follower cells (Konen *et al.*, 2017).

Our sequencing data and the ability to discern the clonality of the distinct leader and follower subpopulations relative to that of the parental population from which they were derived allows us to construct a model of the population origins (Fig. 6C). In a population of tumor cells which are already functionally triploid, *ACTR3* and *KDM5B* undergo mutation, each in separate cells. These mutational events are the beginning of the divergent paths that will develop two separate but cooperative subpopulations. Cells containing mutant *ACTR3* go on to form the highly invasive, slow proliferating leader cells while cells containing the *KDM5B* mutation ultimately become the follower population of invasion-deficient, rapidly proliferating cells. Expressing the leader specific *ACTR3* mutation in follower cells increases chain-like invasion. Likewise, expressing the follower specific *KDM5B* mutation in leader cells increases chain-like invasion. Our phenotypic data adds to this model the necessity for cells exhibiting both the leader phenotype and a follower phenotype to the cooperative behavior demonstrated in collective invasion.

The approach used herein, while innovative, does have certain limitations. First, the work focuses on the characterization of leader and followers isolated from a single NSCLC-

derived cell line. How universal the identified mutations/genes might be in contributing to leader/follower behavior in other cell lines or cancer types is currently unknown. In addition, while our functional characterization supports a potential role for ACTR3 and KDM5B in modulating collective invasion in this setting, these were only two of the phenotype-selective mutations identified. It is possible that no single alteration in isolation is sufficient to fully drive either phenotype, but rather the combinatorial effects of multiple genetic and epigenetic alterations contribute to collective behavior. Studies aimed at similar analyses of multiple cell lines, or collective invasion models and across cancer types may converge on critical “drivers” or pathways. Ultimately the identification of such key alterations may help in clinical decision-making by identifying predictive biomarkers or new therapeutic targets. Indeed, several molecular inhibitors of the KDM5 family are currently under development, particularly in the setting of combination therapy and therapy resistance (Horton et al., 2016a,b; Vazquez-Rodriguez et al., 2018; Nie et al., 2018; Vinogradova et al., 2016; Johansson et al., 2016; Tumber et al., 2017; Bayo et al., 2018). Chemical inhibition of KDM5B as an anti-metastatic modality has not yet been tested. Moreover, a recently discovered Arp2/3 inhibitor CK-666 (Nolen et al., 2009) has been shown to inhibit cell motility *in vivo* (Ilatovskaya et al., 2013). An alternative approach would be targeting other components of the Arp2/3 pathway. One such target, PLK4, has been implicated as a driver of cancer invasion and metastasis in part through its interaction with Arp2/3 subunits (Kazazian et al., 2017), and PLK4 inhibitors are currently under clinical investigation for patients with advanced solid tumors. Our identification of a panel of mutations delineating leader and follower cell phenotypes in a non-small cell lung cancer tumor population is an initial step toward elucidation of how heterogeneous genetic mutations contribute to cancer metastasis and how these vulnerabilities can be exploited to circumvent the development of metastasis.

Materials and Methods

Cell lines and transfections. Leader and follower cells isolated via the SaGA technique from the H1299 human NSCLC cell line (Konen et al., 2017), as well as the parental H1299 cell line (ATCC, Manassas, VA), were cultured in RPMI-1640 media supplemented with 10% fetal bovine serum and 100 units mL⁻¹ of penicillin/streptomycin and maintained at 37°C and 5% CO₂. H1299 cells were mycoplasma tested and authenticated using single nucleotide polymorphism analysis through the Emory Genomics Core as previously described (Konen et al. 2017). H1299 cells had been transfected with the Dendra2 as previously described (Konen et al., 2017). 293T cells were maintained in Dulbecco’s Modified Eagle’s Media (DMEM) supplemented with 10% fetal bovine serum and 100 units mL⁻¹ of penicillin/streptomycin at 37°C and 5% CO₂.

For ARP3 lines, lentivirus was prepared by seeding 2×10^6 HEK293T cells in a 100 cm dish and co-transfecting with 5 μ g transfer vector, 0.5 μ g pMD2.G (Plasmid #12259, Addgene), 5 μ g psPAX2 (Plasmid #12260, Addgene), and 1 μ g of lentiviral vector. After 24 hours, media was replaced with 5 mL fresh complete media, then virus-containing media was collected after 24 hours. Media was centrifuged for 3 minutes at 1000 rpm, 4°C, and then supernatant was filtered through a 0.45 μ m low protein-binding filter.

Target cells were seeded at 70% confluence in a 6-well plate one day prior to lentivirus collection. After virus collection, target cell media was replaced with 1 mL complete media plus 340 μ L virus stock and 1.34 μ L polybrene (10 mg mL⁻¹ stock), added dropwise. After 24 hours, media was replaced with 2 mL complete media. Selection antibiotics (shRNA: puromycin; ARP3 expression vectors: hygromycin) were added 48 hours after viral infection.

shACTR3 constructs were obtained from Millipore Sigma TRCN0000029383 and TRCN0000380403. mCherry-ARP3 lentiviral constructs were created by cloning ARP3-pmCherryC1 (a gift from Christien Merrifield; Addgene plasmid # 27682 (Taylor et al., 2011)) into the pCDH-UBC-MCS-EF1 Hygro backbone. The UBC promoter was subsequently exchanged for a CMV promoter. The ARP3 K240R mutation was created using site-directed mutagenesis.

The HA-KDM5B, pLKO-scrambled, pLKO-KDM5Bsh1, and pLKO-KDM5Bsh2 constructs were a kind gift from Dr. Qin Yan (Sayegh et al., 2013). The puromycin selection marker was exchanged for a hygromycin cassette, and the KDM5B L685W mutation was introduced by site directed mutagenesis. H1299 cells were transfected with 2 μ g plasmid vector to 1×10^5 cells per well of 6-well plate with Lipofectamine 3000 (ThermoFisher Scientific Inc., Carlsbad, CA, USA). After 48 hours, transduced cells were selected with 500 μ g mL⁻¹ hygromycin.

Lentivirus was prepared by seeding 9×10^5 HEK293T cells in a 6cm dish and co-transfecting with 0.1 μ g VSV.G (viral envelope plasmid), 0.9 μ g dR8.2 (viral packaging plasmid), and 1 μ g of lentiviral vector 24 hours post-seeding. After additional 24 hours, media was replaced with 6 mL fresh complete media, then virus-containing media was collected after a further 24 hours. Media was centrifuged for 3 min at 1000 rpm, 4°C, and then supernatant was filtered through a 0.45 μ m low protein-binding filter. Target cells (H1299 parental, leader, and follower cell lines) were seeded at 1×10^5 cells per well in a 6-well plate. Target cells were infected the next day by overlay with 1 mL complete media plus 300 μ L viral stock and 10 μ g mL⁻¹ polybrene. After 24 hours, media was replaced with 2 mL complete media. Cells were selected with 3 μ g mL⁻¹ puromycin at 48 hours after viral infection.

RNA-sequencing and variant calling. RNA-sequencing was performed in triplicate on H1299 parental, leader and follower cells. For parental cells, three different passages were used. For follower cells, three separately-isolated populations were used. For leader cells, two separately-isolated populations were used: one passage of one population, and two passages of the other. RNA library preparation and sequencing were carried out by the Emory Integrated Genomics core and Omega Bio-Tek, Inc. using the TruSeq Stranded mRNA kit, followed by quantification using a Quantus Fluoremeter (Promega, Madison, WI, USA) and integrity assessment using an Agilent 2200 TapeStation instrument. Sequencing was performed using a HiSeq2500 instrument (Illumina, Inc., San Diego, CA, USA), with 50M total sequencing reads generated per sample using the PE100 run format.

Data processing and statistical analyses were performed by the Emory Biostatistics and Bioinformatics Shared Resource. Raw paired-end fastq reads were assessed for quality and contamination using FastQC (Andrews, 2010) and trimmed with Trimmomatic v0.32 (Bolger et al., 2014). Quality filtered reads were mapped against human reference genome hg19 using STAR aligner v2.3.0 (Dobin et al., 2012). Picard tools v1.111 (<http://broadinstitute.github.io/picard>) was used to assess post-alignment QC and to remove PCR duplicates. With an average yield of 30M post-filtered reads, 88% of the reads mapped uniquely with 78% of reads covered in the coding and UTR region. Genomic variants from RNA-seq were called using SamTools v0.1.19 mpileup (Li et al., 2009) with Varscan v2.3.6 (Koboldt et al., 2012) and functionally annotated using ANNOVAR (Wang et al., 2010). A filtering criterion was applied requiring that reported variant had $\geq 6X$ read depth coverage, $\geq 2X$ supporting alternate reads in all samples. Variants associated with intronic, intergenic or synonymous changes, pseudo genes, non-coding RNAs, or sex chromosomes were excluded. Variants were filtered if they were known in dbSNP but not present in COSMIC database. This resulted in a total of 6240 variants. A pairwise two-sided independent t-test was done to compare cell populations mean variant allele frequencies between the groups.

Western blotting. Total cellular protein expression was assessed via Western blotting as previously described (Konen et al., 2016). For histone detection, lifted cells were washed twice with PBS and lysed with Triton extract buffer (PBS containing 0.5% Triton X 100, 2 mM phenylmethylsulfonyl fluoride (PMSF), and 0.02% sodium azide) 10 minutes on ice. Samples were collected by centrifugation for at 4°C for 10 minutes at 2000 rpm. Histones were then acid extracted in 0.2 M HCl by rotation overnight at 4°C. Samples were then electrophoresed on a 6.5% SDS-PAGE gel and subject to western blotting.

Reagents and antibodies. Primary antibodies for Western blot: ARP3 antibody (Santa Cruz, cat. no. sc-48344 – immunogen: residues 1-110 of human ARP3) was used at 1:1000. KDM5B antibody (Sigma, cat. no. HPA 027179 – immunogen: residues 1401-1481 of human KDM5B) was used at 1:500. GAPDH antibody (Cell Signaling, cat. no. 2118) was used at 1:30,000. Beta-tubulin antibody was used (Sigma, cat. no. T4026) at 1:5000. HA antibody (Biolegend cat. no. 901501) was used at 1:1000. H3K4me3 and H3K4me2 antibodies (Millipore cat. nos. 07-473 and 07-030 respectively) were used at 1:10,000. H3K4me1 antibody (Cell Signaling cat. no. 5326T) was used at 1:5000. H4 antibody (Millipore cat. no. 05-858R) was used at 1:20,000. Horseradish peroxidase-conjugated secondary antibodies (Jackson ImmunoResearch) were used at 1:10,000 for Western blot.

3-D invasion assays, spheroid microscopy and image analysis. Spheroids were generated as previously described (Konen et al., 2016) and embedded in 2 mg mL⁻¹ Matrigel (BD Biosciences) diluted in complete media. Images were taken at 0, 24, and in some cases 48 hours post-embedding at 4x using either an Olympus IX51 or CKX41 microscope.

For mixed spheroid experiments, cells were plated together in low-adhesion wells in the indicated ratios with 3000 total cells per spheroid and embedded as previously described. After 24 hours, spheroids were imaged using a Leica SP8 inverted confocal microscope. Invasive area and spheroid circularity were measured using ImageJ as previously described (Konen et al., 2017).

Target validation. DNA and RNA were isolated from H1299 parental, leader and follower cells using DNeasy Blood & Tissue Kit and the RNeasy Mini Kit (Qiagen Sciences, Germantown, MD, USA), respectively. Isolation of samples occurred in two independent biological replicates. RNA was reverse transcribed with M-MLV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) to generate cDNA. Primers were designed and used to amplify regions surrounding each locus of interest subject to Sanger sequencing at GENEWIZ (South Plainfield, NJ, USA). Primer sequences are listed in Table S1. PCR products were cloned into the pCR4-TOPO TA vector by TOPO-TA cloning (ThermoFisher Scientific 450071, Carlsbad, CA, USA) and transformed into bacteria. Fifty individual colonies for each gene in each cell type were and re-streaked on a new ampicillin plate, and twenty colonies each sent to GENEWIZ for Sanger sequencing.

Cell proliferation assay. H1299 cells expressing constructs for either overexpression or knockdown of KDM5B or ARP3 were seeded at 1x10³ cells per well in six wells each of a 96 well plate. After 24-120hr growth, cells were fixed with 100µL per well cold 10% trichloroacetic acid for 1 hr at 4°C, washed three times in water, and air dried. Cells were then stained with

100 μ L 0.4% sulforhodamine B in 1% acetic acid for 30 minutes, washed three times in 1% acetic acid, and dried. Dye was reconstituted in 200 μ L 10 mM Tris base (pH 10.5) for 30 min. Optical density (OD) at 510 nm was measured, and the fold-change in OD determined as: (OD_{X hours}) / (OD_{24 hours}).

Targeted deep resequencing and analysis. Targeted region of KDM5B surrounding exon 15 was PCR amplified using primer sequences listed in Table S1. Amplified products were purified by SPRI beads (KAPA Biosystems, Cape Town, South Africa). Libraries were then created with custom TruSeq compatible adapters and KAPA Hyper Prep Kit (KAPA Biosystems, Cape Town, South Africa). Quality for each library was checked using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequence on an Illumina HiSeq 4000 (Illumina, Inc., San Diego, CA, USA) using 150 bp paired-end sequencing at NYU Genome Technology Center. Quality trimmed reads were mapped to the human genome (GRCh37) using Bowtie 2 (Langmead and Salzberg, 2012). Variant allele frequencies for the KDM5B SNP (rs1141108) and the linked KDM5B L685W mutation were quantified as the fraction of reads (average depth= 348,327 reads per line) exhibiting the A or G allele at position chr1: 202715284 or the A or C variant at position chr1:202715414 using R packages Rsamtools, ShortRead, GenomicAlignments, and BSgenome.Hsapiens.UCSC.hg19 (Lawrence et al., 2013; Morgan et al., 2009, 2017; Team, 2017; Team, 2014). Differences in KDM5B variant allele frequencies were based on analysis of variance with Tukey's post-hoc correction using the R functions 'aov' and 'TukeyHSD', respectively.

Histone demethylase activity assay. Direct quantification of H3K4me3 demethylase activity, cells were lysed by sonication in PBS (3x 4 sec) in the presence of protease inhibitors. Equal amounts of protein were incubated with a histone H3K4me3 substrate for 2 hours at 37°C before specific immune-detection of the H3K4me2 product using Epigenase™ JARID Demethylase Activity/Inhibition Assay Kit (Epigentek P-3083, Farmingdale, NY, USA). Detailed reaction conditions were as previously described (Wang et al., 2013; Dalvi et al., 2017).

Statistical analysis. Two-tailed, unpaired Student's t-test was used to assess statistical significance between any two conditions. Ordinary one-way ANOVA with Tukey's multiple comparisons test was used for experiments in which three or more conditions were being compared. Confidence intervals of proportions in the mixed spheroid experiments were calculated via the Wilson/Brown method. Fisher's exact test was used to test the association of the identified mutations and the phenotypes (e.g. mutant vs. wild-type, leader versus follower) in TOPO-TA cloning experiments.

Data Deposition:

RNAseq data generated as part of this project have been deposited in the NCBI Sequence Read Archive (SRA) under accession # PRJNA542374.

Acknowledgements

The authors wish to thank Dr. Qin Yan (Yale University) for the KDM5B expression constructs, and Dr. Zhengjia (Nelson) Chen (Emory Biostatistics and Bioinformatics Shared Resource) for his assistance with statistical analysis.

Author contributions E.L.Z., B.P., A.I.M., and P.M.V. conceptualized the project and designed the experiments. E.L.Z and B.P. performed the functional validation and biological characterization of the ARP3 and KDM5B mutations and prepared the original manuscript. J. Konen developed the leader and follower cell lines and designed the RNA-seq experiments. B.D., M.R., and J. Kowalski analyzed RNAseq data and performed statistical analyses. B.G.B. and M.Z. performed targeted deep resequencing of KDM5B and analyzed sequencing data. N.S. and M.P.T. conducted the SAPH-ire and PTM analyses. L.W. and E.D.M performed the H3K4me3 demethylase activity assays. J.R.H. and X.C. performed KDM5B protein structure modelling. All authors contributed to the writing, reviewing, and editing of the manuscript.

Competing interests. No competing interests declared.

Funding. This project has been funded in part by National Institute of Health grants 2R01CA077337 (to P.M.V.), R01CA201340-01 and 1U54CA209992 (to A.I.M.) and R01GM117400 (to M.P.T.). Additional support was provided to M.P.T. by the Georgia Clinical & Translational Science Alliance of the National Institutes of Health under Award Number UL1TR002378. E.L.Z. was supported by NIH Predoctoral Fellowship 1F31CA186511. B.P. was support by NIH Predoctoral Fellowship 1F31CA225049. BGB was supported by postdoctoral fellowship #PF-17-109-1-TBG from the American Cancer Society. Support of the Emory Integrated Genomics Core Shared Resource, the Emory University Integrated Cellular Imaging Microscopy Core, and the Biostatistics and Bioinformatics Shared Resource is provided by the Winship Cancer Institute of Emory University core grant under award number 2P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Abercrombie, M.** 1970. Contact inhibition in tissue culture. *In Vitro*, 6, 128-42.
- Abercrombie, M. & Heaysman, J. E.** 1953. Observations on the social behaviour of cells in tissue culture. I. Speed of movement of chick heart fibroblasts in relation to their mutual contacts. *Exp Cell Res*, 5, 111-31.
- Alexander, S., Koehl, G. E., Hirschberg, M., Geissler, E. K. & Friedl, P.** 2008. Dynamic imaging of cancer growth and invasion: a modified skin-fold chamber model. *Histochem Cell Biol*, 130, 1147-54.
- Alexander, S., Weigelin, B., Winkler, F. & Friedl, P.** 2013. Preclinical intravital microscopy of the tumour-stroma interface: invasion, metastasis, and therapy response. *Curr Opin Cell Biol*, 25, 659-71.
- American Cancer, S.** 2016. Cancer Facts & Figures.
- Andrews, S.** 2010. *FastQC: A Quality Control tool for High Throughput Sequence Data*.
- Bamodu, O. A., Huang, W. C., Lee, W. H., Wu, A., Wang, L. S., Hsiao, M., Yeh, C. T. & Chao, T. Y.** 2016. Aberrant KDM5B expression promotes aggressive breast cancer through MALAT1 overexpression and downregulation of hsa-miR-448. *BMC Cancer*.
- Bayo, J., Tran, T. A., Wang, L., Pena-Llopis, S., Das, A. K. & Martinez, E. D.** 2018. Jumonji Inhibitors Overcome Radioresistance in Cancer through Changes in H3K4 Methylation at Double-Strand Breaks. *Cell Reports*, 25, 1040-1050.e5.
- Bolger, A. M., Usadel, B. & Lohse, M.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.
- Bronsert, P., Enderle-Ammour, K., Bader, M., Timme, S., Kuehs, M., Csanadi, A., Kayser, G., Kohler, I., Bausch, D., Hoeppner, J., et al.** 2014. Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer-host interface. *J Pathol*, 234, 410-22.
- Calbo, J., Van Montfort, E., Proost, N., Van Drunen, E., Beverloo, H. B., Meuwissen, R. & Berns, A.** 2011. A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. *Cancer Cell*, 19, 244-256.
- Campbell, P. J., Yachida, S., Mudie, L. J., Stephens, P. J., Pleasance, E. D., Stebbings, L. A., Morsberger, L. A., Latimer, C., McLaren, S., Lin, M. L., et al.** 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467, 1109-13.
- Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J.** 2013. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell*, 155, 1639-1651.

- Cheung, K. J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J. D., Fairchild, A. N., Gorin, M. A., Verdone, J. E., Pienta, K. J., Bader, J. S., et al.** 2016. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci U S A*, 113, E854-63.
- Dalvi, M. P., Wang, L., Zhong, R., Kollipara, R. K., Park, H., Bayo, J., Yenerall, P., Zhou, Y., Timmons, B. C., Rodriguez-Canales, J., et al.** 2017. Taxane-Platin-Resistant Lung Cancers Co-develop Hypersensitivity to JumonjiC Demethylase Inhibitors. *Cell Rep*, 19, 1669-1684.
- De Bruin, E. C., Mcgranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A. J., et al.** 2014. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346, 251-6.
- Dobin, A., Davis, C. A., Zaleski, C., Schlesinger, F., Drenkow, J., Chaisson, M., Batut, P., Jha, S. & Gingeras, T. R.** 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
- Dona, E., Barry, J. D., Valentin, G., Quirin, C., Khmelinskii, A., Kunze, A., Durdu, S., Newton, L. R., Fernandez-Minan, A., Huber, W., et al.** 2013. Directional tissue migration through a self-generated chemokine gradient. *Nature*, 503, 285-9.
- Efferth, T. & Leber, M.** 2009. Molecular principles of cancer invasion and metastasis (Review). *International Journal of Oncology*, 34.
- Ewald, A. J., Huebner, R. J., Palsdottir, H., Lee, J. K., Perez, M. J., Jorgens, D. M., Tauscher, A. N., Cheung, K. J., Werb, Z. & Auer, M.** 2012. Mammary collective cell migration involves transient loss of epithelial features and individual cell migration within the epithelium. *J Cell Sci*, 125, 2638-54.
- Fidler, I.** 2003. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Reviews Cancer*, 3, 453-458.
- Friedl, P.** 2004. Prespecification and plasticity: shifting mechanisms of cell migration. *Curr Opin Cell Biol*, 16, 14-23.
- Friedl, P. & Gilmour, D.** 2009. Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews. Molecular cell biology*, 10, 445-457.
- Friedl, P., Locker, J., Sahai, E. & Segall, J. E.** 2012. Classifying collective cancer cell invasion. *Nat Cell Biol*, 14, 777-783.
- Friedl, P., Noble, P. B., Walton, P. A., Laird, D. W., Chauvin, P. J., Tabah, R. J., Black, M. & Zanker, K. S.** 1995. Migration of coordinated cell clusters in mesenchymal and epithelial cancer explants in vitro. *Cancer Res*, 55, 4557-60.
- Friedl, P. & Wolf, K.** 2010. Plasticity of cell migration: a multiscale tuning model. *J Cell Biol*, 188, 11-9.

- Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., et al.** 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, 366, 883-892.
- Gilbert-Ross, M., Konen, J., Koo, J., Shupe, J., Robinson, B. S., Wiles, W. G. T., Huang, C., Martin, W. D., Behera, M., Smith, G. H., et al.** 2017. Targeting adhesion signaling in KRAS, LKB1 mutant lung adenocarcinoma. *JCI Insight*, 2, e90487.
- Goley, E. D. & Welch, M. D.** 2006. The ARP2/3 complex: an actin nucleator comes of age. *Nature reviews. Molecular cell biology*, 7, 713-26.
- Haeger, A., Krause, M., Wolf, K. & Friedl, P.** 2014. Cell jamming: collective invasion of mesenchymal tumor cells imposed by tissue confinement. *Biochim Biophys Acta*, 1840, 2386-95.
- Haeger, A., Wolf, K., Zegers, M. M. & Friedl, P.** 2015. Collective cell migration: guidance principles and hierarchies. *Trends Cell Biol*, 25, 556-66.
- Han, M., Xu, W., Cheng, P., Jin, H. & Wang, X.** 2017. Histone demethylase lysine demethylase 5B in development and cancer. *Oncotarget*.
- Hegerfeldt, Y., Tusch, M., Bro, E.-B. & Friedl, P.** 2002. Collective Cell Movement in Primary Melanoma Explants. *Cancer Res*, 62, 2125-2130.
- Hinohara, K., Wu, H. J., Vigneau, S., Mcdonald, T. O., Igarashi, K. J., Yamamoto, K. N., Madsen, T., Fassl, A., Egri, S. B., Papanastasiou, M., et al.** 2018. KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell*, 34, 939-953.e9.
- Horton, J. R., Engstrom, A., Zoeller, E. L., Liu, X., Shanks, J. R., Zhang, X., Johns, M. A., Vertino, P. M., Fu, H. & Cheng, X.** 2016a. Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases. *J Biol Chem*, 291, 2631-46.
- Horton, J. R., Liu, X., Gale, M., Wu, L., Shanks, J. R., Zhang, X., Webber, P. J., Bell, J. S. K., Kales, S. C., Mott, B. T., et al.** 2016b. Structural Basis for KDM5A Histone Lysine Demethylase Inhibition by Diverse Compounds. *Cell Chem Biol*, 23, 769-781.
- Ilatovskaya, D. V., Chubinskiy-Nadezhdin, V., Pavlov, T. S., Shuyskiy, L. S., Tomilin, V., Palygin, O., Staruschenko, A. & Negulyaev, Y. A.** 2013. Arp2/3 complex inhibitors adversely affect actin cytoskeleton remodeling in the cultured murine kidney collecting duct M-1 cells. *Cell Tissue Res*, 354, 783-92.
- Iwaya, K., Oikawa, K., Semba, S., Tsuchiya, B., Mukai, Y., Otsubo, T., Nagao, T., Izumi, M., Kuroda, M., Domoto, H., et al.** 2007. Correlation between liver metastasis of the colocalization of actin-related protein 2 and 3 complex and WAVE2 in colorectal carcinoma. *Cancer Sci*, 98, 992-9.

- Johansson, C., Velupillai, S., Tumber, A., Szykowska, A., Hookway, E. S., Nowak, R. P., Strain-Damerell, C., Gileadi, C., Philpott, M., Burgess-Brown, N., et al.** 2016. Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat Chem Biol*, 12, 539-45.
- Jolly, M. K., Somarelli, J. A., Sheth, M., Biddle, A., Tripathi, S. C., Armstrong, A. J., Hanash, S. M., Bapat, S. A., Rangarajan, A. & Levine, H.** 2018. Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. *Pharmacol Ther*.
- Kabla, A. J.** 2012. Collective cell migration: leadership, invasion and segregation. *J R Soc Interface*, 9, 3268-78.
- Kazazian, K., Go, C., Wu, H., Brashavitskaya, O., Xu, R., Dennis, J. W., Gingras, A.-C. & Swallow, C. J.** 2017. Plk4 Promotes Cancer Invasion and Metastasis through Arp2/3 Complex Regulation of the Actin Cytoskeleton. *Cancer Res*, 77, 1-14.
- Kinoshita, T., Nohata, N., Watanabe-Takano, H., Yoshino, H., Hidaka, H., Fujimura, L., Fuse, M., Yamasaki, T., Enokida, H., Nakagawa, M., et al.** 2012. Actin-related protein 2/3 complex subunit 5 (ARPC5) contributes to cell migration and invasion and is directly regulated by tumor-suppressive microRNA-133a in head and neck squamous cell carcinoma. *Int J Oncol*, 40, 1770-8.
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., Mclellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K.** 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22, 568-576.
- Konen, J., Summerbell, E., Dwivedi, B., Galior, K., Hou, Y., Rusnak, L., Chen, A., Saltz, J., Zhou, W., Boise, L., et al.** 2017. Image-guided genomics of phenotypically heterogeneous populations reveals vascular signaling during symbiotic collective cancer invasion. *Nature Communications*, 8, 15078.
- Konen, J., Wilkinson, S., Lee, B., Fu, H., Zhou, W., Jiang, Y. & Marcus, A. I.** 2016. LKB1 kinase-dependent and -independent defects disrupt polarity and adhesion signaling to drive collagen remodeling during invasion. *Molecular biology of the cell*, 27, mbc.E15-08-0569-1084.
- Kuo, K. T., Huang, W. C., Bamodu, O. A., Lee, W. H., Wang, C. H., Hsiao, M., Wang, L. S. & Yeh, C. T.** 2018. Histone demethylase JARID1B/KDM5B promotes aggressiveness of non-small cell lung cancer and serves as a good prognostic predictor. *Clin Epigenetics*.
- Langmead, B. & Salzberg, S. L.** 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9, 357-9.

- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. & Carey, V. J.** 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol*, 9, e1003118.
- Leighton, J., Kalla, R., Turner, J. & Fennell, R.** 1960. Pathogenesis of Tumor Invasion II. Aggregate Replication. *Cancer Research*, 20, 586.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S.** 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- Lipinski, K. A., Barber, L. J., Davies, M. N., Ashenden, M., Sottoriva, A. & Gerlinger, M.** 2016. Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer*, 2, 49-63.
- Liu, J. S., Farlow, J. T., Paulson, A. K., Labarge, M. A. & Gartner, Z. J.** 2012. Programmed cell-to-cell variability in Ras activity triggers emergent behaviors during mammary epithelial morphogenesis. *Cell Rep*, 2, 1461-70.
- Liu, Z., Yang, X., Chen, C., Liu, B., Ren, B., Wang, L., Zhao, K., Yu, S. & Ming, H.** 2013. Expression of the Arp2/3 complex in human gliomas and its role in the migration and invasion of glioma cells. *Oncol Rep*, 30, 2127-2136.
- Lu, W., Liu, S., Li, B., Xie, Y., Adhiambo, C., Yang, Q., Ballard, B. R., Nakayama, K. I., Matusik, R. J. & Chen, Z.** 2015. SKP2 inactivation suppresses prostate tumorigenesis by mediating JARID1B ubiquitination. *Oncotarget*, 6, 771-88.
- Lv, J., Liu, J., Xiao, M., Xu, H., Xu, C., Zhang, X., Tang, L., Jiang, F., Zhou, Y., Zhang, Z., et al.** 2018. ARP3 promotes tumor metastasis and predicts a poor prognosis in hepatocellular carcinoma. *Pathol Res Pract*, 214, 1356-1361.
- Malet-Engra, G., Yu, W., Oldani, A., Rey-Barroso, J., Gov, N. S., Scita, G. & Dupre, L.** 2015. Collective cell motility promotes chemotactic prowess and resistance to chemorepulsion. *Curr Biol*, 25, 242-250.
- Mateo, F., Meca-Cortes, O., Celia-Terrassa, T., Fernandez, Y., Abasolo, I., Sanchez-Cid, L., Bermudo, R., Sagasta, A., Rodriguez-Carunchio, L., Pons, M., et al.** 2014. SPARC mediates metastatic cooperation between CSC and non-CSC prostate cancer cell subpopulations. *Mol Cancer*, 13, 237.
- Mayor, R. & Carmona-Fontaine, C.** 2010. Keeping in touch with contact inhibition of locomotion. *Trends Cell Biol*, 20, 319-28.
- Mayor, R. & Etienne-Manneville, S.** 2016. The front and rear of collective cell migration. *Nat Rev Mol Cell Biol*, 17, 97-109.
- Mcgranahan, N. & Swanton, C.** 2017. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168, 613-628.

- Mcpherson, A., Roth, A., Laks, E., Masud, T., Bashashati, A., Zhang, A. W., Ha, G., Biele, J., Yap, D., Wan, A., et al.** 2016. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet*, 48, 758-67.
- Mehlen, P. & Puisieux, A.** 2006. Metastasis: a question of life or death. *Nat Rev Cancer*, 6, 449-58.
- Mertins, P., Qiao, J. W., Patel, J., Udeshi, N. D., Clauser, K. R., Mani, D. R., Burgess, M. W., Gillette, M. A., Jaffe, J. D. & Carr, S. A.** 2013. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods*, 10, 634-7.
- Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H. & Gentleman, R.** 2009. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25, 2607-8.
- Morgan, M., Pagès, H., Obenchain, V. & Hayden, N.** 2017. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import.
- Muinonen-Martin, A. J., Susanto, O., Zhang, Q., Smethurst, E., Faller, W. J., Veltman, D. M., Kalna, G., Lindsay, C., Bennett, D. C., Sansom, O. J., et al.** 2014. Melanoma cells break down LPA to establish local gradients that drive chemotactic dispersal. *PLoS Biol*, 12, e1001966.
- Nabeshima, K., Inoue, T., Shima, Y., Kataoka, H. & Kono, M.** 1999. Cohort migration of carcinoma cells: differentiated colorectal carcinoma cells move as coherent cell clusters or sheets. *Histol Histopathol.*, 14, 1183-1197.
- Nie, Z., Shi, L., Lai, C., O'Connell, S. M., Xu, J., Stansfield, R. K., Hosfield, D. J., Veal, J. M. & Stafford, J. A.** 2018. Structure-based design and discovery of potent and selective KDM5 inhibitors. *Bioorg Med Chem Lett*, 28, 1490-1494.
- Nolen, B. J., Tomasevic, N., Russell, A., Pierce, D. W., Jia, Z., McCormick, C. D., Hartman, J., Sakowicz, R. & Pollard, T. D.** 2009. Characterization of two classes of small molecule inhibitors of Arp2/3 complex. *Nature*, 460, 1031-4.
- Pandya, P., Orgaz, J. L. & Sanz-Moreno, V.** 2017. Actomyosin contractility and collective migration: may the force be with you. *Curr Opin Cell Biol*, 48, 87-96.
- Qin, J. Y., Zhang, L., Clift, K. L., Hurler, I., Xiang, A. P., Ren, B. Z. & Lahn, B. T.** 2010. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One*, 5, e10611.
- Richardson, A. M., Havel, L. S., Koyen, A. E., Konen, J. M., Shupe, J., Wiles, W. G., Martin, W. D., Grossniklaus, H. E., Sica, G., Gilbert-Ross, M., et al.** 2018. Vimentin Is Required for Lung Adenocarcinoma Metastasis via Heterotypic Tumor Cell–Cancer-Associated Fibroblast Interactions during Collective Invasion. *Clinical Cancer Research*, 24, 420-432.

Rinner, B., Galle, B., Trajanoski, S., Fischer, C., Hatz, M., Maierhofer, T., Michelitsch, G., Moinfar, F., Stelzer, I., Pfragner, R., et al. 2012. Molecular evidence for the bi-clonal origin of neuroendocrine tumor derived metastases. *BMC Genomics*, 13, 594.

Sayegh, J., Cao, J., Zou, M. R., Morales, A., Blair, L. P., Norcia, M., Hoyer, D., Tackett, A. J., Merkel, J. S. & Yan, Q. 2013. Identification of small molecule inhibitors of Jumonji AT-rich interactive domain 1B (JARID1B) histone demethylase by a sensitive high throughput screen. *J Biol Chem*, 288, 9408-17.

Semba, S., Iwaya, K., Matsubayashi, J., Serizawa, H., Kataba, H., Hirano, T., Kato, H., Matsuoka, T. & Mukai, K. 2006. Coexpression of actin-related protein 2 and Wiskott-Aldrich syndrome family verproline-homologous protein 2 in adenocarcinoma of the lung. *Clin Cancer Res*, 12, 2449-54.

Shen, X., Zhuang, Z., Zhang, Y., Chen, Z., Shen, L., Pu, W., Chen, L. & Xu, Z. 2015. JARID1B modulates lung cancer cell proliferation and invasion by regulating p53 expression. *Tumour Biol*, 36, 7133-42.

Tam, W. L. & Weinberg, R. A. 2013. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med*, 19, 1438-49.

Tang, B., Qi, G., Tang, F., Yuan, S., Wang, Z., Liang, X., Li, B., Yu, S., Liu, J., Huang, Q., et al. 2015. JARID1B promotes metastasis and epithelial-mesenchymal transition via PTEN/AKT signaling in hepatocellular carcinoma cells. *Oncotarget*, 6, 12723-39.

Taylor, M. J., Perrais, D. & Merrifield, C. J. 2011. A high precision survey of the molecular dynamics of mammalian clathrin-mediated endocytosis. *PLoS Biol*, 9, e1000604.

Team, R. C. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Team, T. B. D. 2014. BSgenome.Hsapiens.UCSC.hg19: Full genome sequences for Homo sapiens (UCSC version hg19).

Theveneau, E., Marchant, L., Kuriyama, S., Gull, M., Moepps, B., Parsons, M. & Mayor, R. 2010. Collective chemotaxis requires contact-dependent cell polarity. *Dev Cell*, 19, 39-53.

Torres, M. P., Dewhurst, H. & Sundararaman, N. 2016. Proteome-wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease. *Molecular & Cellular Proteomics*, 15, 3513.

Tsuji, T., Ibaragi, S., Shima, K., Hu, M. G., Katsurano, M., Sasaki, A. & Hu, G. F. 2008. Epithelial-mesenchymal transition induced by growth suppressor p12CDK2-AP1 promotes tumor cell local invasion but suppresses distant colony growth. *Cancer Res*, 68, 10377-86.

Tumber, A., Nuzzi, A., Hookway, E. S., Hatch, S. B., Velupillai, S., Johansson, C., Kawamura, A., Savitsky, P., Yapp, C., Szykowska, A., et al. 2017. Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem Biol*, 24, 371-380.

- Valentin, G., Haas, P. & Gilmour, D.** 2007. The chemokine SDF1a coordinates tissue migration through the spatially restricted activation of Cxcr7 and Cxcr4b. *Curr Biol*, 17, 1026-31.
- Van Zijl, F., Krupitza, G. & Mikulits, W.** 2011. Initial steps of metastasis: cell invasion and endothelial transmigration. *Mutat Res*, 728, 23-34.
- Vazquez-Rodriguez, S., Wright, M., Rogers, C. M., Cribbs, A. P., Velupillai, S., Philpott, M., Lee, H., Dunford, J. E., Huber, K. V. M., Robers, M. B., et al.** 2018. Design, Synthesis and Characterization of Covalent KDM5 Inhibitors. *Angew Chem Int Ed Engl*.
- Vinogradova, M., Gehling, V. S., Gustafson, A., Arora, S., Tindell, C. A., Wilson, C., Williamson, K. E., Guler, G. D., Gangurde, P., Manieri, W., et al.** 2016. An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat Chem Biol*, 12, 531-8.
- Viré, E., Curtis, C., Davalos, V., Git, A., Robson, S., Villanueva, A., Vidal, A., Barbieri, I., Aparicio, S., Esteller, M., et al.** 2014. The Breast Cancer Oncogene EMSY Represses Transcription of Antimetastatic microRNA miR-31. *Mol Cell*.
- Waclaw, B., Bozic, I., Pittman, M. E., Hruban, R. H., Vogelstein, B. & Nowak, M. A.** 2015. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525, 261-4.
- Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., Cox, J., Mann, M. & Choudhary, C.** 2011. A Proteome-wide, Quantitative Survey of In Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. *Molecular & Cellular Proteomics*, 10, M111.013284-M111.013284.
- Wagner, S. A., Beli, P., Weinert, B. T., Scholz, C., Kelstrup, C. D., Young, C., Nielsen, M. L., Olsen, J. V., Brakebusch, C. & Choudhary, C.** 2012. Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues. *Molecular & Cellular Proteomics*, 1578-1585.
- Wang, K., Li, M. & Hakonarson, H.** 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38, e164.
- Wang, L., Chang, J., Varghese, D., Dellinger, M., Kumar, S., Best, A. M., Ruiz, J., Bruick, R., Pena-Llopis, S., Xu, J., et al.** 2013. A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth. *Nat Commun*, 4, 2035.
- Wang, Z., Tang, F., Qi, G., Yuan, S., Zhang, G., Tang, B. & He, S.** 2015. KDM5B is overexpressed in gastric cancer and is required for gastric cancer cell proliferation and metastasis. *Am J Cancer Res*, 5, 87-100.
- Westcott, J. M., Prechtel, A. M., Maine, E. A., Dang, T. T., Esparza, M. A., Sun, H., Zhou, Y., Xie, Y. & Pearson, G. W.** 2015. An epigenetically distinct breast cancer cell subpopulation promotes collective invasion. *J Clin Invest*, 125, 1927-43.

- Yamane, K., Tateishi, K., Klose, R. J., Fang, J., Fabrizio, L. A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P. & Zhang, Y.** 2007. PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. *Molecular Cell*, 25, 801-812.
- Yang, Z.-L., Miao, X., Xiong, L., Zou, Q., Yuan, Y., Li, J., Liang, L., Chen, M. & Chen, S.** 2013. CFL1 and Arp3 are Biomarkers for Metastasis and Poor Prognosis of Squamous Cell/Adenosquamous Carcinomas and Adenocarcinomas of Gallbladder. *Cancer Investigation*, 31, 132-139.
- Zhang, J., Fujimoto, J., Zhang, J., Wedge, D. C., Song, X., Zhang, J., Seth, S., Chow, C. W., Cao, Y., Gumbs, C., et al.** 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*, 346, 256-9.
- Zheng, H. C., Zheng, Y. S., Li, X. H., Takahashi, H., Hara, T., Masuda, S., Yang, X. H., Guan, Y. F. & Takano, Y.** 2008. Arp2/3 overexpression contributed to pathogenesis, growth and invasion of gastric carcinoma. *Anticancer Research*, 28, 2225-2232.
- Zheng, Y. C., Chang, J., Wang, L. C., Ren, H. M., Pang, J. R. & Liu, H. M.** 2019. Lysine demethylase 5B (KDM5B): A potential anti-cancer drug target. *Eur J Med Chem*, 161, 131-140.

Table 1 RNA-seq reveals leader- and follower-specific gene mutations.

	Gene symbol	Full name	Protein Function	Variant locus (GRCh37)	VAF (%)		
					Parental	Leaders	Followers
Leader-enriched	ACTR3	Actin-related protein 3 (ARP3)	Major component of Arp2/3 complex	chr2:114699797; A:G	8.76	23.4	0.08
	MCM5	Minichromosome maintenance complex component 5	Pre-replication complex during DNA replication	chr22:35809920; G:A	7.59	26.9	0.14
	MIPEP	Mitochondrial intermediate peptidase	Oxidative phosphorylation protein maturation	chr13:24413837; A:C	7.75	37.2	0.98
	NAE1	NEDD8 activating enzyme E1 subunit 1	Activation of neddylation pathway	chr16:66852492; T:C	26.2	58.9	0.13
	NUP93	Nucleoporin 93	Component of nuclear pore complex	chr16:56868312; G:A	25.9	57.8	0.29
	ZNF302	Zinc finger protein 302	Function has yet to be determined	chr19:35175335; G:C	4.71	34.6	0.47
Follower-enriched	CLEC11A	C-type lectin domain family 11, member A	Growth factor for hematopoietic progenitor cells	chr19:51228679; C:G	8.42	0	22.4
	KDM5B	Lysine Demethylase 5B	Demethylates lysine 4 of histone H3	chr1:202715414; A:C	17.5	0.74	28.1
	NDUFS1	NADH:Ubiquinone oxidoreductase core subunit S1	Core subunit of electron transport chain Complex I	chr2:207012514; C:A	10.4	0.28	22.8
	RERE	Arginine-glutamate dipeptide repeats	Possible role in controlling cell survival	chr1:8416225; G:C	38.7	0	65.4
	RNF115	Ring finger protein 115	E3 ubiquitin ligase	chr1:145686997; T:C	14.7	0.27	21.9
	SKA1	Spindle and kinetochore associated complex subunit 1	Chromosome segregation during mitosis	chr18:47902232; C:G	13.0	0	21.6
	TBP	TATA-Box binding protein	Involved in transcription initiation by RNA pol. II	chr6:170871308; G:T	12.4	0.36	21.7

Figures

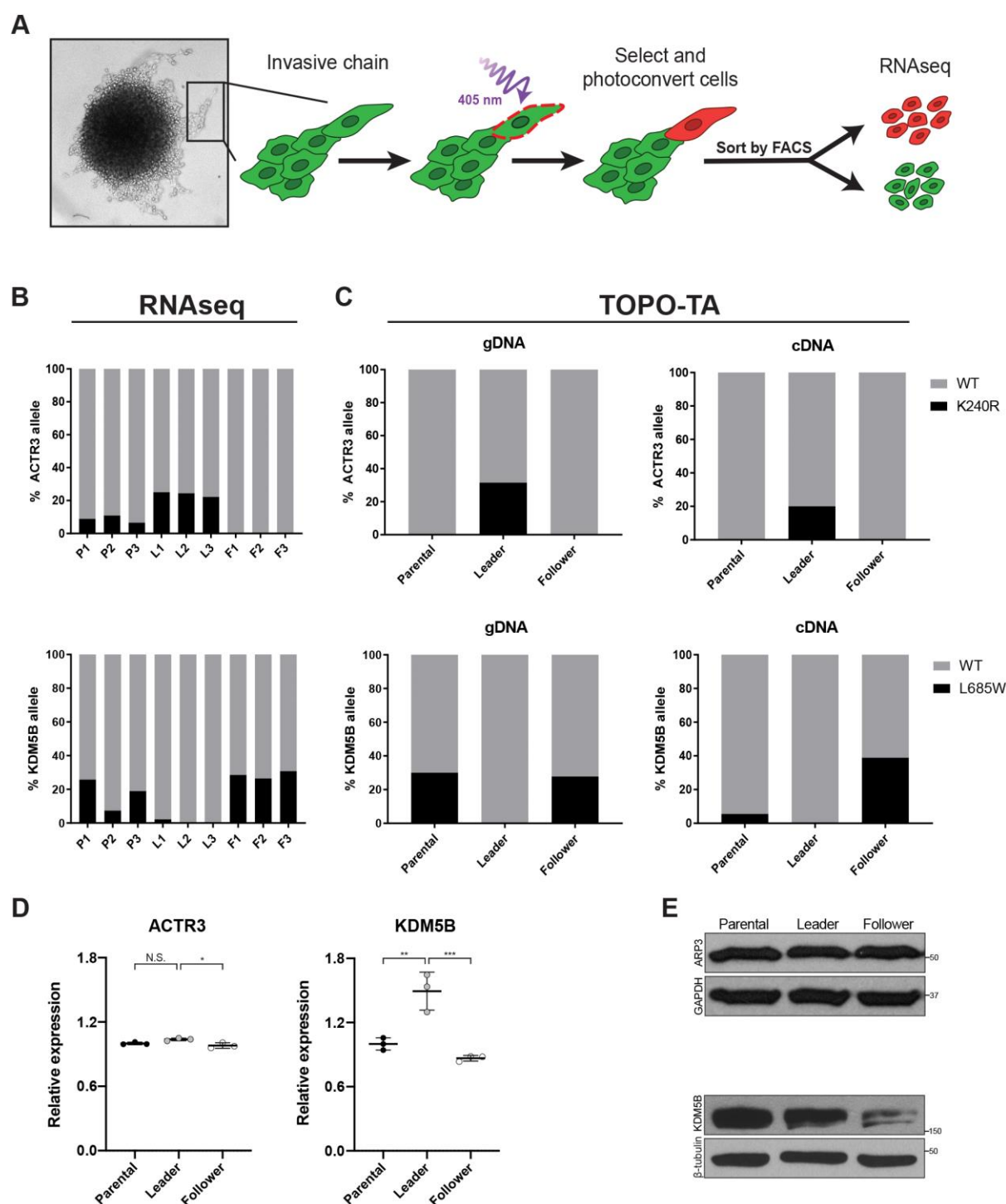


Fig. 1. ARP3 K240R and KDM5B L685W are validated mutations in H1299 leader and follower cells. (A) Schematic of the SaGA protocol used to isolate leader and follower cell populations. (B) Variant allele frequencies for ACTR3 and KDM5B from RNA-sequencing of H1299 parental (P), leader (L) and follower (F) cells. $n=3$ separate populations per group. (C) TOPO-TA cloning and subsequent Sanger sequencing confirms the presence of ACTR3

K240R and KDM5B L685W mutations in both cDNA and genomic DNA (gDNA) from parental, leader and follower cells, respectively. $n=20$ colonies (ARP3 parental, follower gDNA and parental, leader cDNA and KDM5B parental gDNA); 19 colonies (ARP3 leader gDNA and KDM5B leader gDNA); 18 colonies (ARP3 follower cDNA and KDM5B follower gDNA and parental, follower cDNA); and 17 colonies (KDM5B leader cDNA) (association between the genotype and cell phenotype was determined by Fisher's exact test as follows: mutant vs. wild-type, leader versus follower ARP3 gDNA $p=0.008$, ARP3 cDNA $p=0.11$, KDM5B gDNA $p=0.02$, and KDM5B cDNA $p=0.008$). (D) Relative mRNA expression (via RNA-seq; normalized to parental average) and (E) protein levels (via Western blot) of ACTR3 and KDM5B in H1299 parental, leader, and follower populations. $*p<0.05$, $**p<0.01$, $***p<0.001$ by one-way ANOVA with Tukey's post-test.

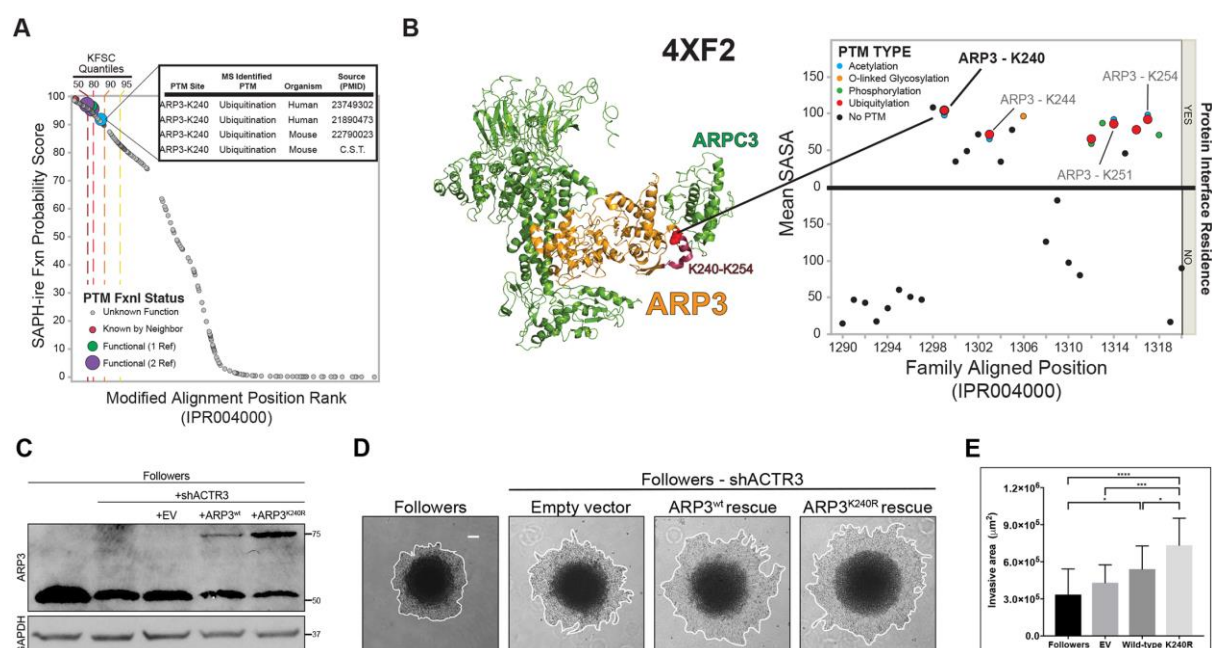


Fig. 2. PTM hotspot analysis of ARP3 K240 suggests functional impact of the K240R mutation. (A) Plot of SAPH-ire probability score by rank for all modified alignment positions in the ARP protein family IPR004000. The ARP3 K240 ubiquitylation site is highly ranked along with other MAPs that contain PTMs with well-established function (4 or more supporting references), as indicated by known function source count (KFSC) quantiles. (Inset) Table of ARP3 K240 PTMs identified by mass spectrometry of human and mouse tissues, including literature sources. (B) Local PTM topology of the ARP3 family near ARP3 K240. PTM sites plotted by solvent accessible surface area (SASA) and proximity to the interface of a protein-protein interaction. Human ARP3 PTMs are labeled, revealing multiple ubiquitylation sites between K240-K254. (Left) Structure of Arp2/3 complex (PDB 4XF2) indicating ARP3 K240 (spheres) within the K240-K254 region of ARP3 (red). (C) Western blot showing exogenous (upper band) versus endogenous (lower band) ARP3 expression in unmodified followers, and shACTR3-followers rescued with either empty vector (EV), wild-type ARP3, or ARP3 K240R. Rescue constructs were mCherry-tagged to allow for visualization within invasive chains. (D) Images of invasion in Matrigel at 24hrs of spheroids comprised of unmodified follower cells, or shACTR3 followers transfected with either empty vector, wild-type ARP3, or ARP3 K240R constructs. (E) Quantification of spheroid invasive area (mean±s.d., $n=17, 15, 16,$ and 17 spheroids per group, respectively, across $N=3$ experiments. $**p<0.01$, $***p<0.001$, $****p<0.0001$ by one-way ANOVA with Tukey's post-test). Scale bar: 100 μm.

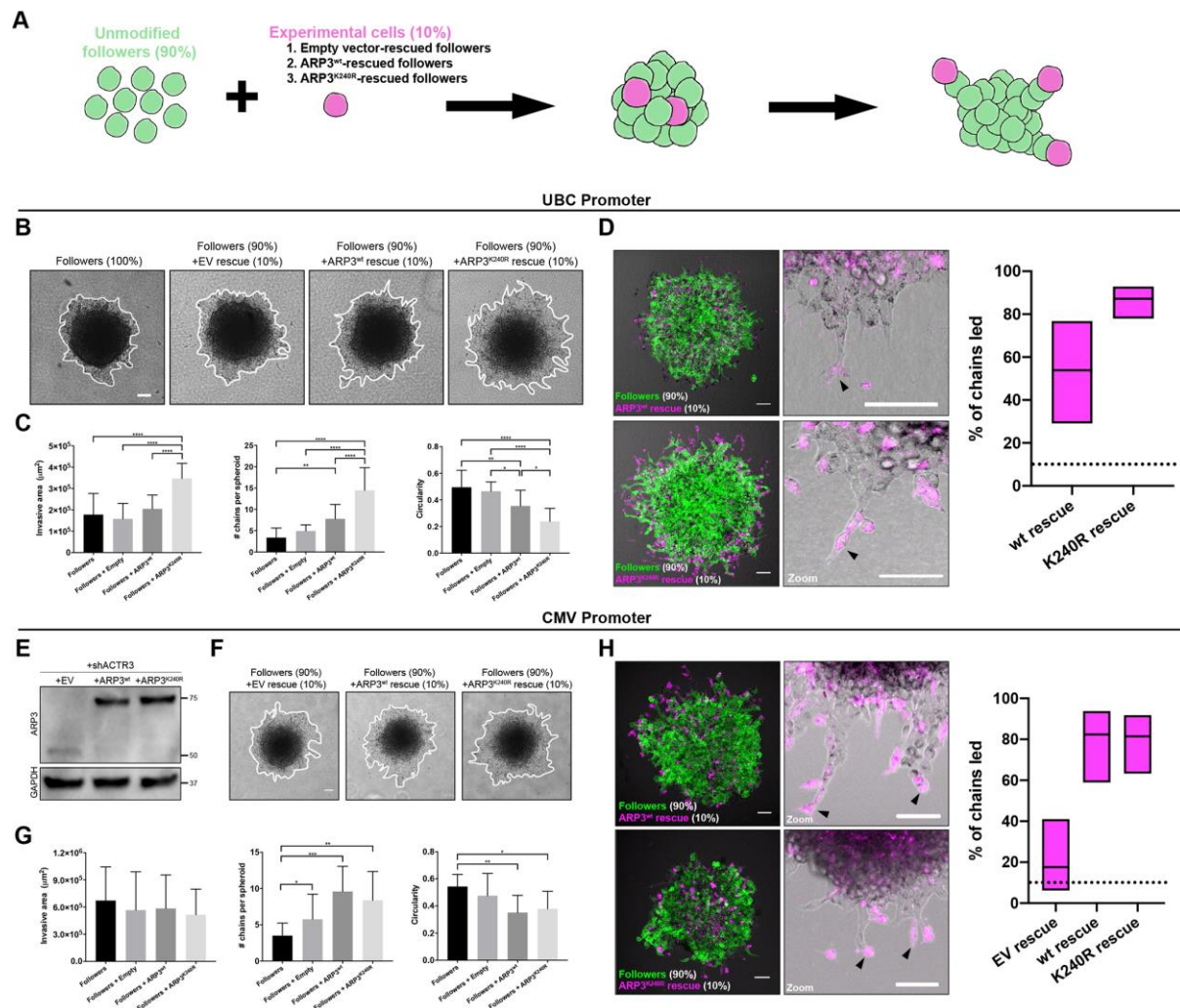


Fig. 3. ARP3 K240R confers leader-like properties when expressed in follower cells. (A) Illustration of spheroid mixing experimental setup. Spheroids were comprised of either 100% unmodified followers, or 90% unmodified followers plus 10% of shACTR3 followers rescued with either empty vector, wild-type ARP3, or ARP3 K240R. **(B)** Invasion of mixed spheroids in Matrigel after 24 hr. Rescue constructs were expressed under control of the UBC promoter. Representative images shown for each condition. **(C)** Quantification of invasive area, circularity, and average number of chains per spheroid for each condition (mean \pm s.d., $n=14$, 11, 18, and 16 spheroids for unmodified followers, EV rescue, wildtype ARP3 rescue, and ARP3 K240R rescue, respectively, across $N=3$ experiments. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ by one-way ANOVA with Tukey's post-test). **(D)** Confocal fluorescence imaging of mixed spheroids, with unmodified followers shown in green and mCherry-ARP3 K240R-rescued cells or mCherry-wild-type ARP3-rescued cells shown in magenta. Black arrows indicate invasive chains being led by experimental rescue cells. Graphs show percentage of chains (mean \pm 95% confidence intervals) led by wild-type ARP3-rescued and ARP3 K240R-rescued followers. Dotted line denotes 10% of chains led, corresponding to the proportion of

ARP3 rescued cells in the mixed spheroids. Scale bar: 100 μ m. (E) Western blot showing protein expression levels of ARP3 in followers after knockdown of endogenous ARP3 and rescue with empty vector, wild-type ARP3, or ARP3 K240R. Rescue constructs were expressed under control of the CMV promoter. (F-G) 24 hr invasion of mixed spheroids in Matrigel. Representative images (F) and quantification (G) of invasive area, circularity, and average numbers of chains per spheroid shown for each condition (mean \pm s.d., $n=12, 12, 12$, and 11 spheroids per group, respectively, across $N=2$ experiments. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, by one-way ANOVA with Tukey's post-test). Scale bars: 100 μ m. (H) Confocal fluorescence imaging of mixed spheroids, with unmodified followers shown in green and mCherry-wild-type ARP3-rescued cells or mCherry-ARP3 K240R-rescued cells shown in magenta. Black arrows indicate invasive chains being led by experimental rescue cells. Graphs show the percentage of chains (mean \pm 95% confidence intervals) led by mCherry-empty vector-rescued, wild-type ARP3-rescued and ARP3 K240R-rescued followers. Dotted line denotes the proportion of ARP3 rescued cells in the mixed spheroids thus the 10% of chains expected to be led based on random chance.

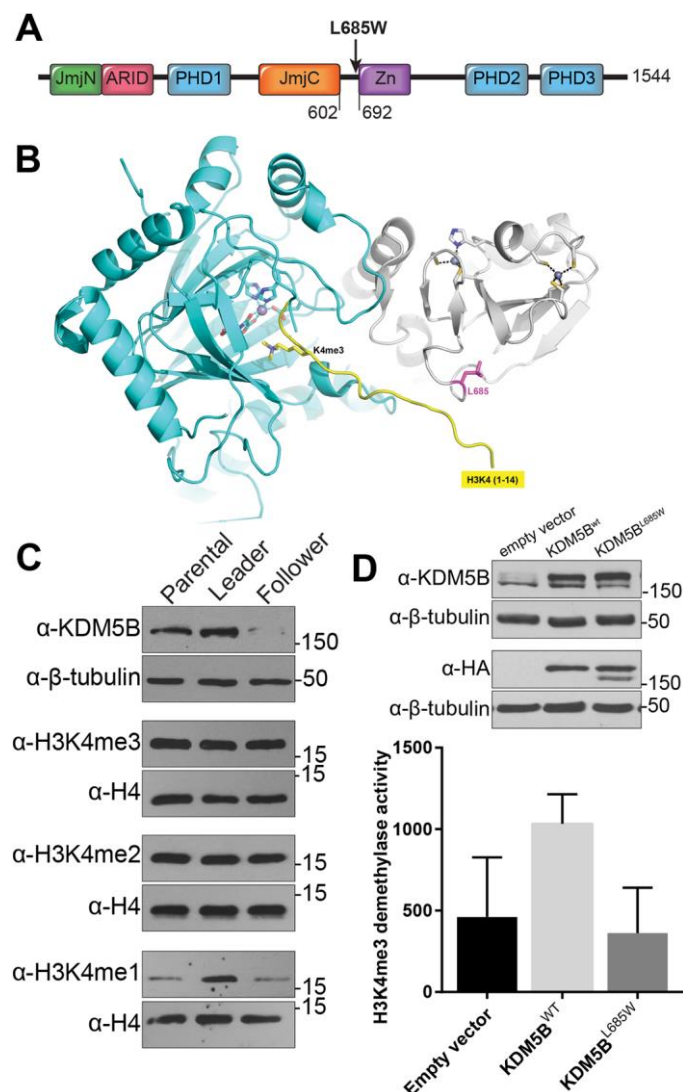


Fig. 4. Characterization of the follower-enriched KDM5B L685W mutation. (A) Diagram of KDM5B protein domains and relative position of amino acid residue 685 to zinc finger domain. (B) Partial model of KDM5B L685W with H3 N-terminal tail (residues 1-14). The model was generated by superimposing a structure of KDM5B (PDB 5A1F) with that of KDM6A in complex with histone H3 peptide (PDB 3AVR), which placed a trimethylated H3K4 in the active site of KDM5B, near the Fe(II)-binding site, and the C-terminal tail of histone H3 near Leu685. (C) Western blot analysis of KDM5B expression and mono, di- and tri- methylated H3K4 in lysates from H1299 parental, leader, and follower cell populations. Loading controls β -tubulin and H4 are probed from the same blot as preceding bands. Experiment was repeated on three separate protein collections from cells at different passages ($N=3$) (D) *In vitro* H3K4 demethylation assay performed on lysates derived from three separate, biological replicates of parental H1299 cells expressing empty vector, wild-type KDM5B, or KDM5B L685W (mean \pm s.d. specific activity; $N=3$ biological replicates).

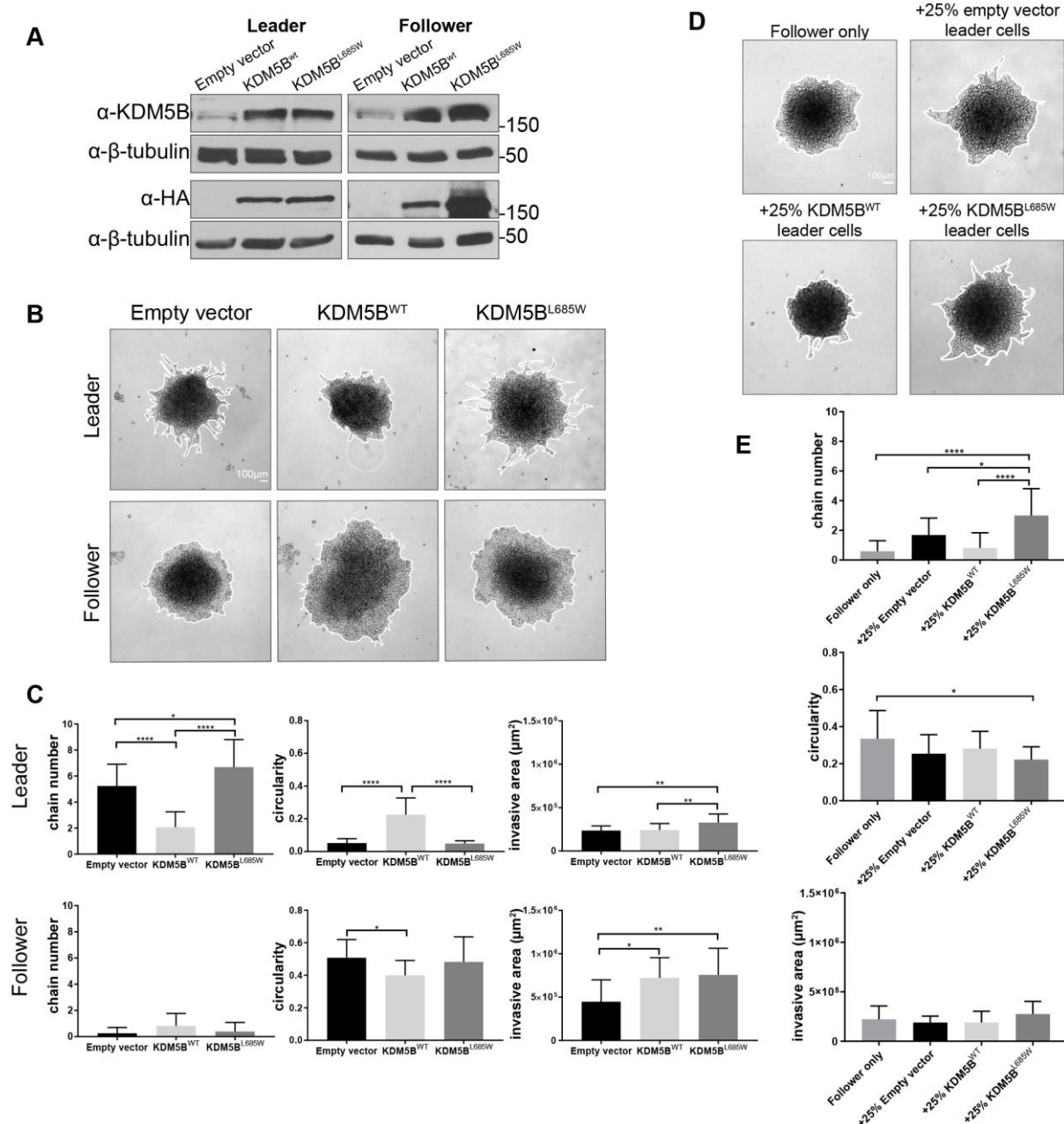


Fig. 5. Wild-type KDM5B suppress and KDM5B L685W drives chain-like invasion. (A) Western blot analysis demonstrating successful expression of HA-tagged wild-type and L685W KDM5B in leader and follower cell populations. (B) Images of invasion in Matrigel at 24 hr of leader and follower spheroids expressing empty vector, wild-type HA-KDM5B, or HA-KDM5B L685W. (C) Quantification of invasive area, circularity, and chain number from spheroids depicted in (B) (mean±s.d., $n=16$ spheroids (follower empty vector, leader KDM5B^{WT}, and leader KDM5B^{L685W}), $n=17$ spheroids (leader empty vector and follower KDM5B^{WT}), $n=18$ spheroids (follower KDM5B^{L685W}), across $N=3$ independent experiments, * $p<0.05$, ** $p<0.01$, **** $p<0.0001$ by one-way ANOVA with Tukey's post-test). (D) Images of invasion in Matrigel at 24 hr of spheroids composed of 75% follower cells and 25% leader cells

expressing either empty vector, wild-type HA-KDM5B, or HA-KDM5B-L685W. (E) Quantification of invasive area, circularity, and chain number from spheroids depicted in (D) (mean±s.d., $n=16$ spheroids (+25% leader empty vector), $n=17$ spheroids (follower only and +25% leader KDM5B^{WT}), $n=18$ spheroids (+25% leader KDM5B^{L685W}), across three independent experiments, * $p<0.05$, **** $p<0.0001$ by one-way ANOVA with Tukey's post-test).

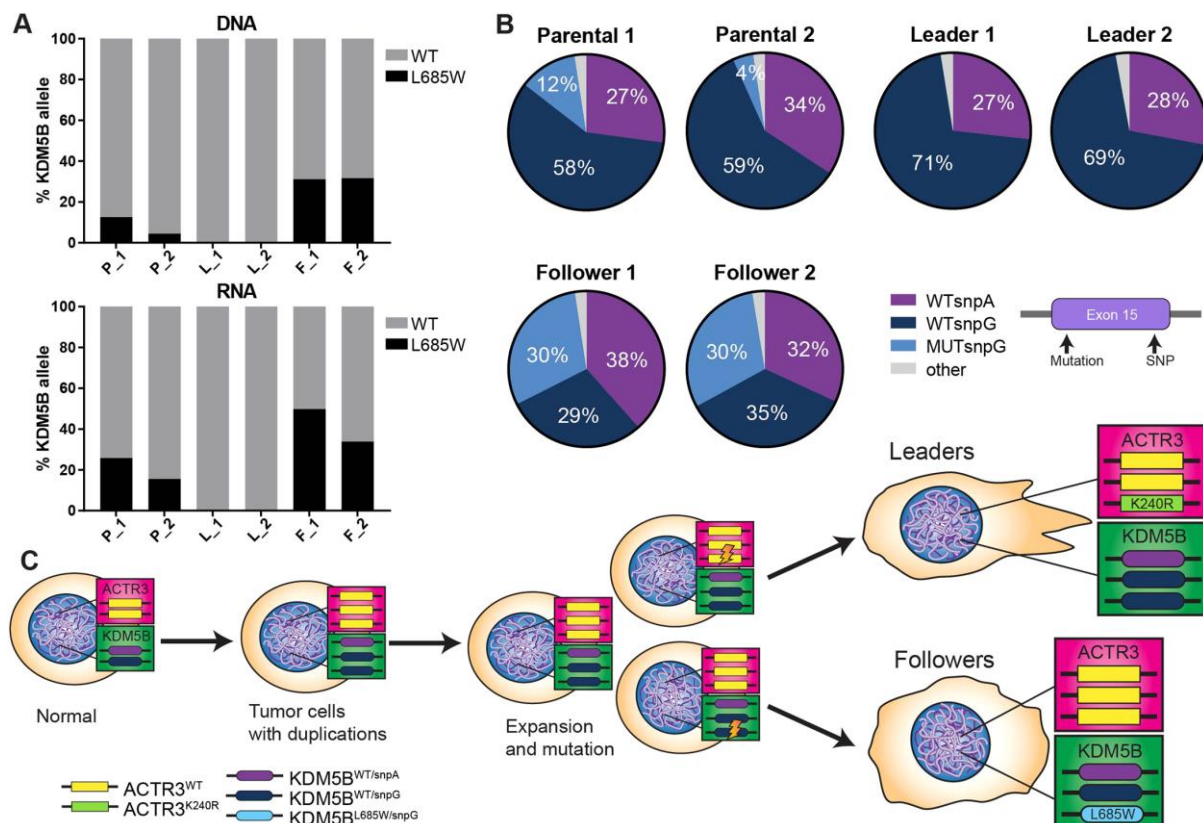


Fig. 6. Leader and follower cells are derived from two separate populations defined by mutational profile. (A) Deep targeted resequencing across KDM5B exon 15 in genomic DNA (top) and RNA (bottom) isolated from parental, leader, and follower populations ($N=2$ independent isolates of DNA/RNA at separate passages of cells derived from a single phenotypic isolation. Average depth= 348,327 reads per sample). (B) Pie charts depicting proportions of KDM5B genotypes across parental, leader, and follower populations as determined from the deep amplicon sequencing shown in panel A. (C) Model of the potential history of leader and follower populations from parental cell population as inferred from the genetic profiles of KDM5B and ARP3.

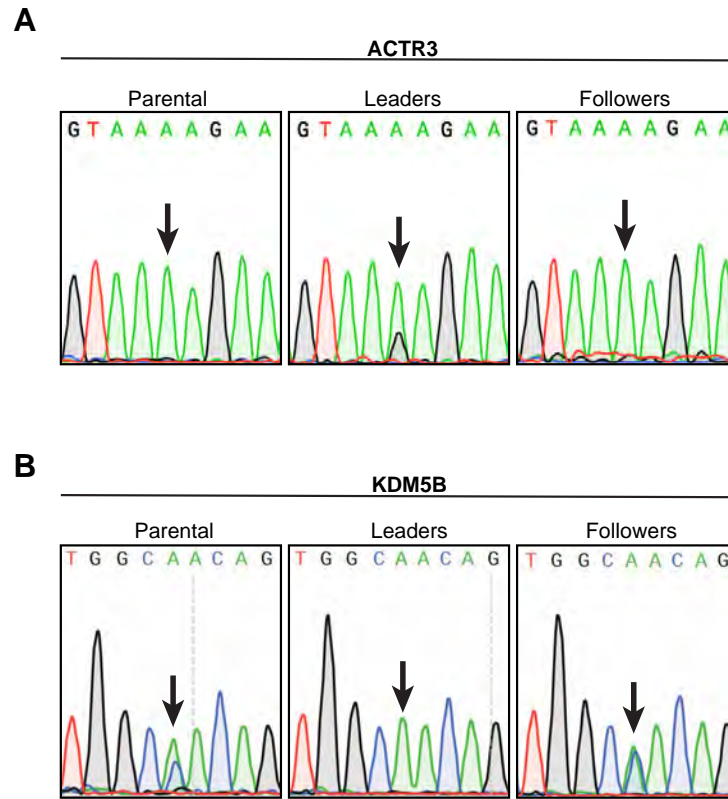


Fig. S1. Confirmation of leader- and follower-enriched mutations. Sanger sequencing confirming leader-enriched ACTR3 mutation (A) and follower-enriched KDM5B mutation (B) in cDNA (shown) and genomic DNA isolated from H1299 parental, leader and follower populations. Black arrows indicate the bases of interest. (A) Only the wild-type A peak is seen in the parental and follower populations, while the leader population contains both A and G peaks. (B) Only the wild-type A peak is seen in the leader population, while the parental and follower populations contain both A and C peaks.

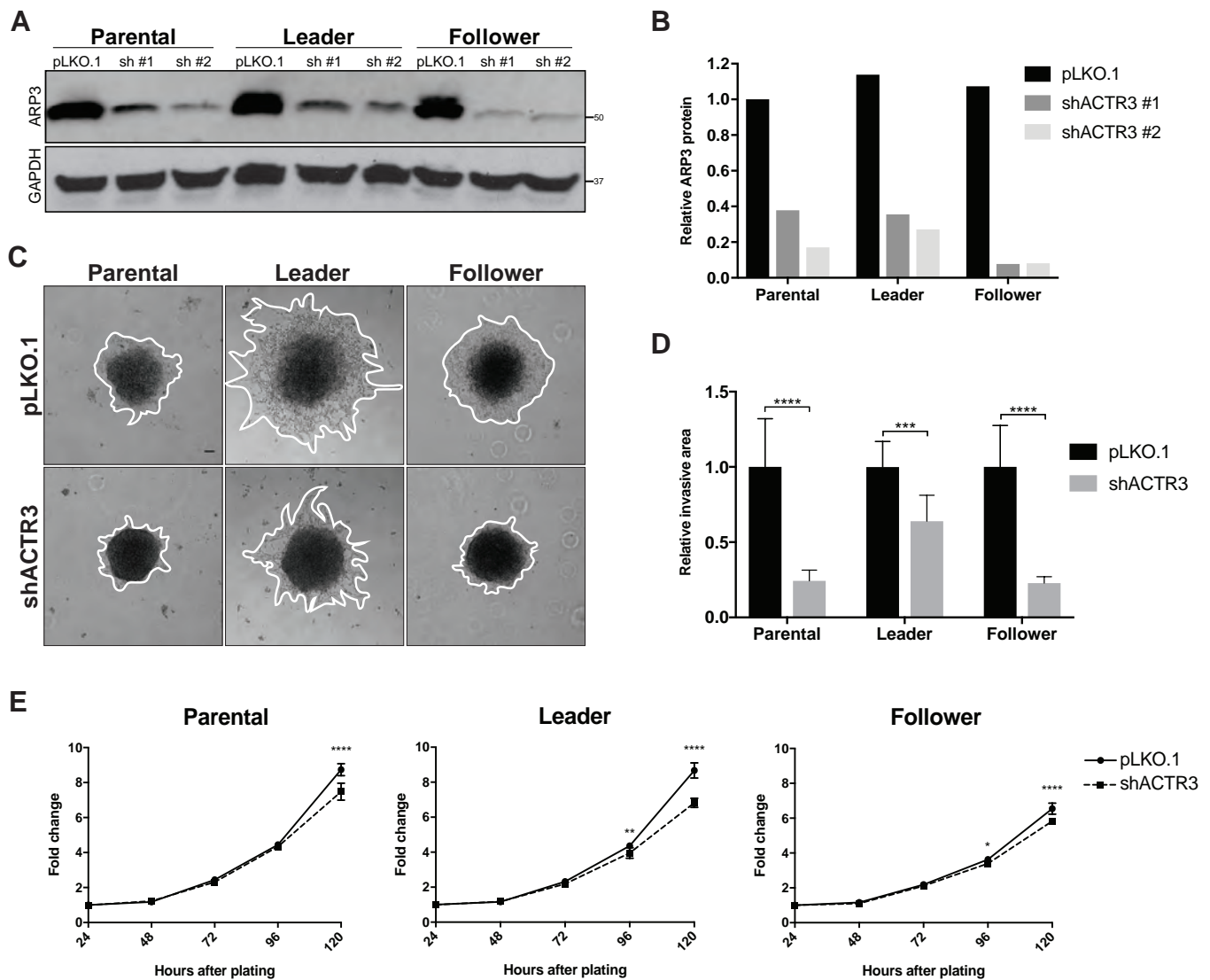


Fig. S2. ARP3 knockdown inhibits 3-D invasion. (A) Western blot showing ARP3 protein levels in H1299 parental, leader and follower cells upon expression of empty pLKO.1 vector, ARP3 shRNA #1 (Millipore Sigma TRCN0000029383), or ARP3 shRNA #2 (Millipore Sigma TRCN0000380403). (B) Western blot densitometry quantification, indicating 70-90% knockdown of ARP3 protein using either shRNA #1 or shRNA #2. (C) Representative images of 24-hour invasion of H1299 parental, leader, and follower spheroids expressing either empty pLKO.1 or shACTR3 #2. Scale bar = 100µm. (D) Quantification of relative 24-hour invasive area, normalized to pLKO.1 control for each group. (mean±s.d., n=5, 11, and 5 spheroids for parental, leader and follower lines, respectively. ***p<0.001, ****p<0.0001 by two-way ANOVA with Sidak correction). (E) Growth rate of parental, leader, and follower lines expressing either empty pLKO.1 or shACTR3 #2. (mean±s.d., n=5 replicates per time point. *p<0.05, ***p<0.001, ****p<0.0001 by two-way ANOVA with Šidák correction).

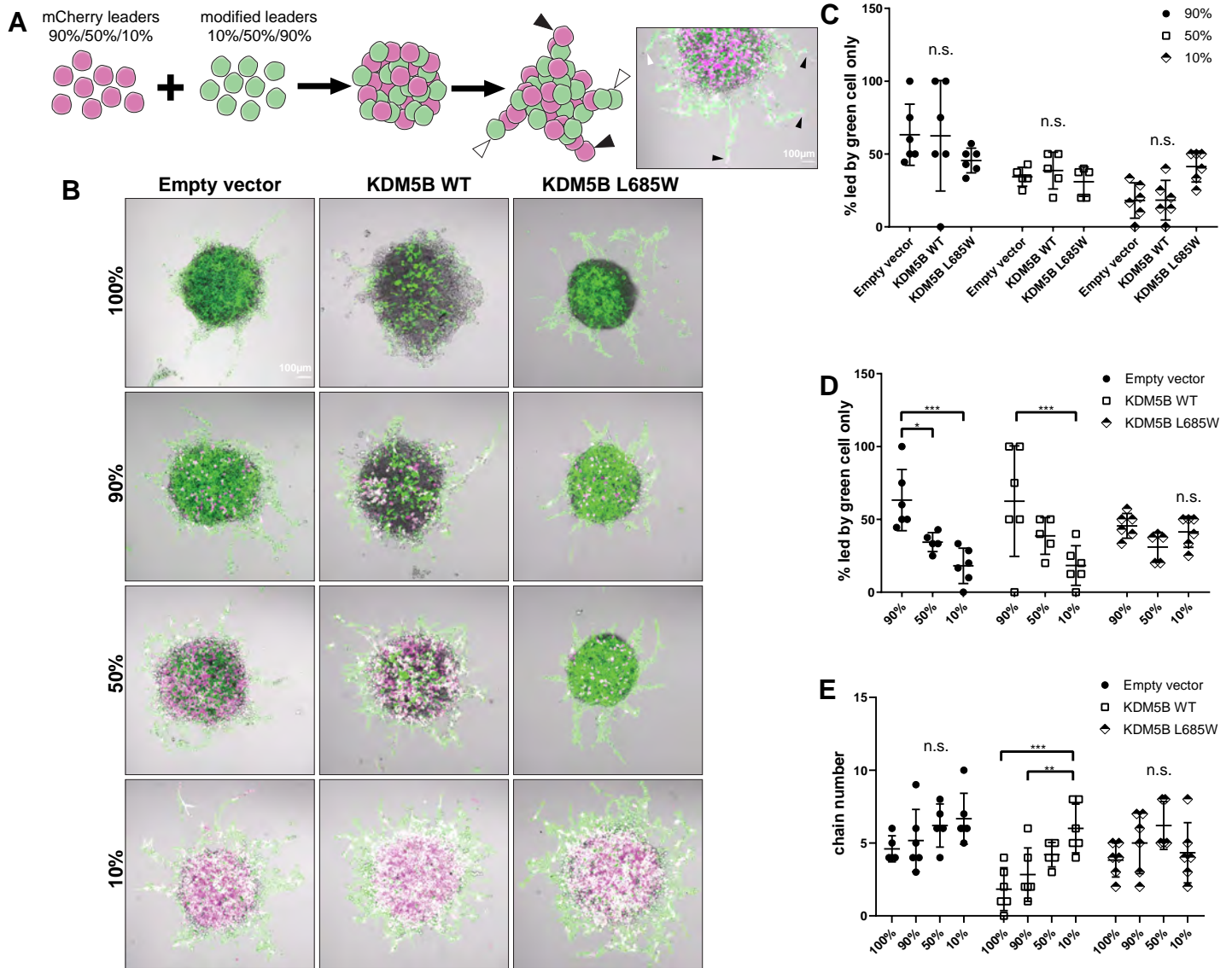


Figure S3. Leader cells overexpressing wild-type KDM5B but not KDM5B L685W exhibit diminished chain formation but retain leader activity in leader-only spheroids. (A) Schematic of spheroid mixing experiment in which unmodified mCherry leaders were mixed with empty vector, KDM5B wild-type, or KDM5B L685W leaders and representative image of a spheroid containing 50% empty vector expressing and 50% unmodified leaders. Black arrows indicate magenta cell (unmodified leader) led chains. White arrows indicate green-only led chains. Both populations express Dendra2 (green) whereas the unmodified leaders also express mCherry (magenta). (B) Representative confocal fluorescence imaging of spheroids in which, 10%, 50%, or 90% leaders stably expressing empty vector, wild-type HA-KDM5B, or HA-KDM5B L685W were mixed with unmodified leaders 24 hours after embedding in Matrigel. (C) Percent invasive chains led by green-only (vector/KDM5B/KDM5B L685W) transduced cells, grouped by KDM5B overexpression cell line (mean±s.d., $n=6$ spheroids for 90% and 10% mixes or $n=5$ spheroids for 50% mixes across $N=1$ biological replicate, n.s. not significant, $p>0.05$, by two-way ANOVA with Tukey's post-test). (D) Same data as in panel B except grouped by the fraction of green-only (vector/KDM5B/KDM5B L685W) transduced cells in the spheroid (mean±s.d., $n=6$ spheroids for 90% and 10% mixes or $n=5$ spheroids for 50% mixes across $N=1$ biological replicate, n.s. not significant, $*p<0.05$, $***p<0.001$ by two-way ANOVA with Tukey's post-test). (E) Average number of chains per spheroid grouped by percentage of the indicated transduced leader cell line (mean±s.d., $n=6$ spheroids for 90% and 10% mixes or $n=5$ spheroids for 50% mixes across $N=1$ biological replicate, n.s. not significant ($p>0.05$), $**p<0.01$, $***p<0.001$ by two-way ANOVA with Tukey's post-test).

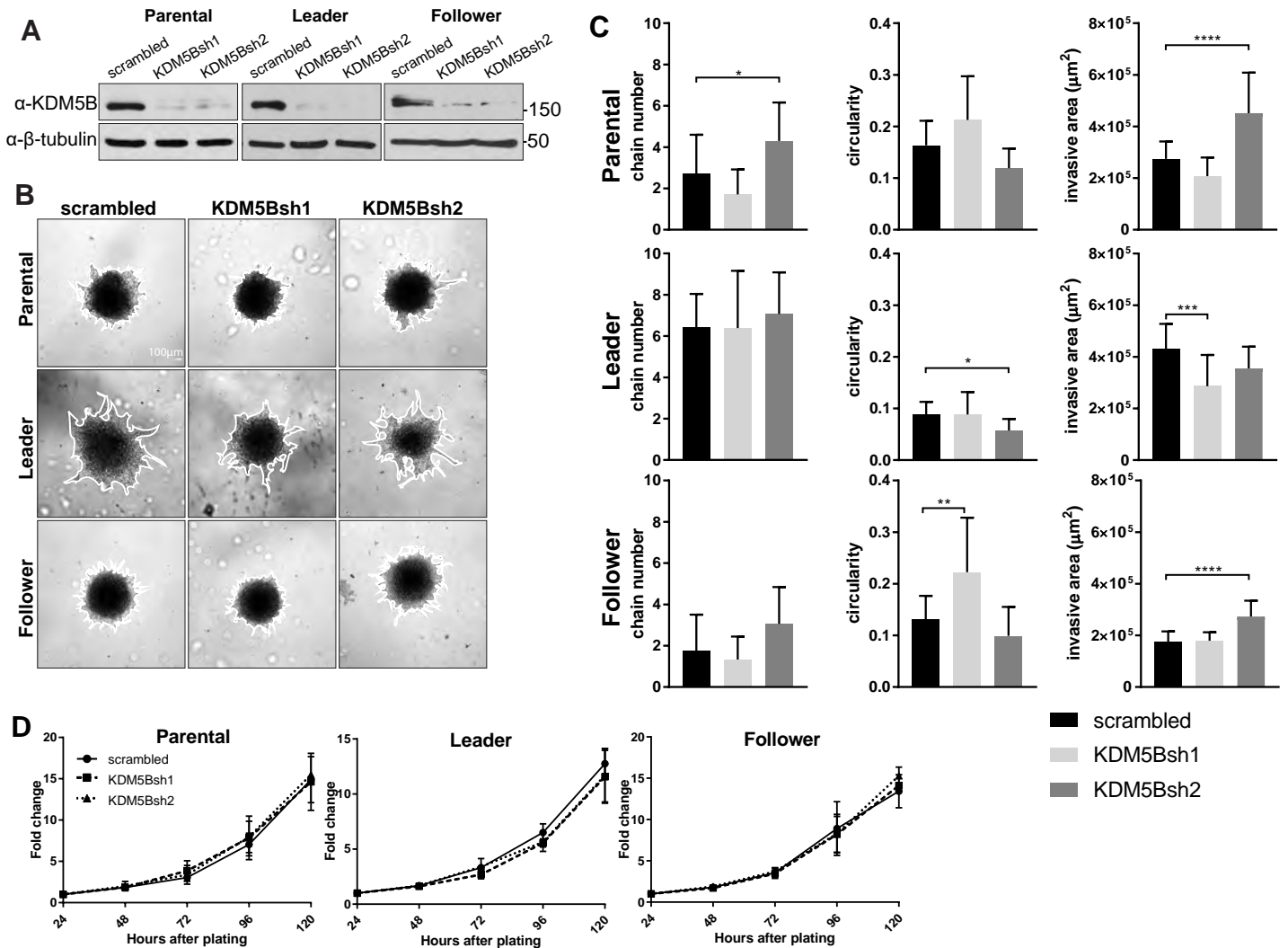


Figure S4. Impact of KDM5B knockdown on growth and chain-like invasion. (A) Western blot of KDM5B in parental, leader, and follower cell populations with two different shRNAs, shRNA1, shRNA2 against KDM5B or a scrambled control. (B) Parental, leader, and follower cells expressing scrambled shRNA, KDM5Bsh1, or KDM5Bsh2 were grown as 3-D spheroids in Matrigel at 24 hours. (C) Quantification of invasive area, circularity, and chain number from spheroids depicted in (B) (mean±s.d., $n=13$ spheroids (follower scrambled), $n=14$ spheroids (leader KDM5Bsh2), $n=15$ spheroids (leader scrambled and follower KDM5Bsh1), $n=17$ spheroids (parental scrambled, KDM5Bsh1, and KDM5Bsh2 and follower KDM5Bsh2), or $n=18$ spheroids (leader KDM5Bsh2) across $N=3$ separate experiments, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ by one-way ANOVA with Tukey's post-test). (D) Growth of parental, leader, and follower H1299 lines stably expressing scrambled shRNA, KDM5Bsh1, or KDM5Bsh2 (mean±s.d of triplicate determinations from $N=3$ independent experiments).