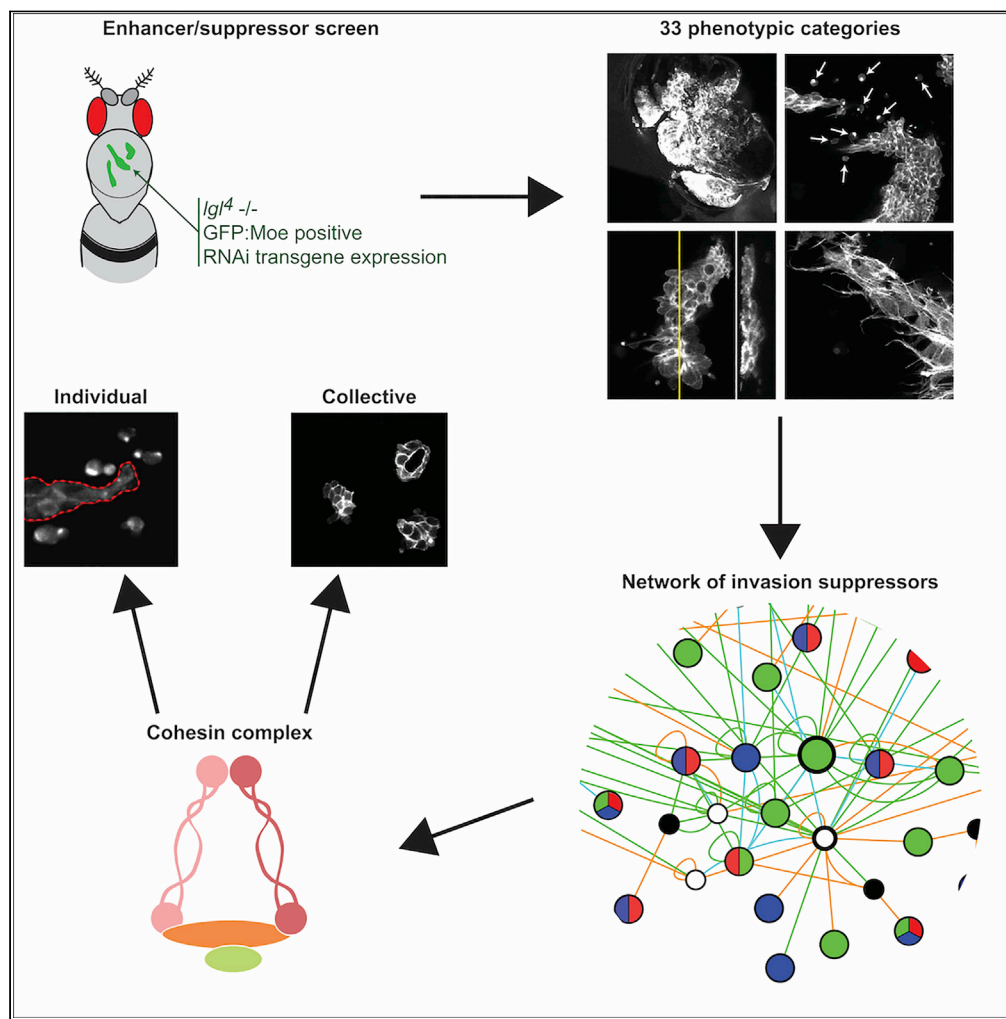


Article

A Genetic Analysis of Tumor Progression in *Drosophila* Identifies the Cohesin Complex as a Suppressor of Individual and Collective Cell Invasion



Brenda Canales
Coutiño, Zoe E.
Cornhill, Africa
Couto, ..., Ruman
Rahman, Sean T.
May, Marios
Georgiou

marios.georgiou@
nottingham.ac.uk

HIGHLIGHTS

Screen identifies genes
that affect tumor behavior
in a wide variety of ways

A functionally validated
network of invasion-
suppressor genes was
generated

Loss of cohesin complex
function can promote
individual or collective cell
invasion

The fly pupal notum is an
excellent *in vivo* system to
study tumor progression

Canales Coutiño et al.,
iScience 23, 101237
June 26, 2020 © 2020 The
Authors.
[https://doi.org/10.1016/
j.isci.2020.101237](https://doi.org/10.1016/j.isci.2020.101237)

Article

A Genetic Analysis of Tumor Progression in *Drosophila* Identifies the Cohesin Complex as a Suppressor of Individual and Collective Cell Invasion

Brenda Canales Coutiño,^{1,7} Zoe E. Cornhill,^{1,7} Africa Couto,^{1,7} Natalie A. Mack,^{1,2,7} Alexandra D. Rusu,¹ Usha Nagarajan,^{1,3} Yuen Ngan Fan,^{1,4} Marina R. Hadjicharalambous,^{1,5} Marcos Castellanos Uribe,² Amy Burrows,¹ Anbarasu Lourdasamy,⁶ Ruman Rahman,⁶ Sean T. May,² and Marios Georgiou^{1,8,*}

SUMMARY

Metastasis is the leading cause of death for patients with cancer. Consequently it is imperative that we improve our understanding of the molecular mechanisms that underlie progression of tumor growth toward malignancy. Advances in genome characterization technologies have been very successful in identifying commonly mutated or misregulated genes in a variety of human cancers. However, the difficulty in evaluating whether these candidates drive tumor progression remains a major challenge. Using the genetic amenability of *Drosophila melanogaster* we generated tumors with specific genotypes in the living animal and carried out a detailed systematic loss-of-function analysis to identify conserved genes that enhance or suppress epithelial tumor progression. This enabled the discovery of functional cooperative regulators of invasion and the establishment of a network of conserved invasion suppressors. This includes constituents of the cohesin complex, whose loss of function either promotes individual or collective cell invasion, depending on the severity of effect on cohesin complex function.

INTRODUCTION

Metastasis is the major cause of mortality in human cancers, yet we know relatively little about the biology that underlies the important transition to invasive malignancy (Sporn, 1996; Hanahan and Weinberg, 2000) and currently few genes have been identified that suppress this process (Steeg, 2003; Yan et al., 2013). Most human cancers are epithelial in origin; consequently cancer cell invasion, where individual cells or groups of cells break away from the primary tumor to invade the surrounding tissue, is a key hallmark of tumor progression. Invasion is highly complex, involving concurrent dramatic changes in cytoskeletal organization, cell polarity, cell-cell junctions, and focal contacts, as cells within the developing tumor collectively destroy the normal architecture of the host epithelium and deregulate the local microenvironment (Hurst and Welch, 2011). Understanding and dissecting the molecular mechanisms that promote tumor progression and cancer cell invasion will be important for the development of new therapeutic strategies in our battle against this disease.

Drosophila melanogaster has become an increasingly important model system in the study of cancer biology. Conservation of major signaling pathways related to tumorigenesis and metastasis, coupled with the genetic amenability of this organism, has directly led to advances in our understanding of this disease (Rudrapatna et al., 2012; Brumby and Richardson, 2005). The short lifespan and low running costs of this organism make it particularly amenable to large-scale screens, and there is now a vast array of published literature using the fly to study cancer (Gonzalez, 2013; Rudrapatna et al., 2012; Mirzoyan et al., 2019).

We have developed an *in vivo* system in *Drosophila* that allows us to study epithelial cell and tissue morphogenesis in real time (Georgiou et al., 2008; Georgiou and Baum, 2010; Cohen et al., 2010; Couto et al., 2017). This system allows the shape, dynamics, and behavior of labeled mutant epithelial cells to be followed in high resolution in the living animal. In this current study, we use this *in vivo* system to

¹School of Life Sciences, University of Nottingham, Nottingham NG7 2UH, UK

²School of Biosciences, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, UK

³Department of Biochemistry, School of Interdisciplinary and Applied Sciences, Central University of Haryana, Jant-Pali, Mahendergarh, Haryana, 123029, India

⁴Faculty of Biology, Medicine & Health, University of Manchester, Manchester M13 9PL, UK

⁵Department of Pharmacy and Pharmacology, University of Bath, Bath BA2 7AY, UK

⁶School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK

⁷These authors contributed equally

⁸Lead Contact

*Correspondence: marios.georgiou@nottingham.ac.uk
<https://doi.org/10.1016/j.isci.2020.101237>



generate tumors with specific genotypes on the dorsal thorax epithelium of the fly and to observe tumor cell morphology and behavior in high spatial and temporal resolution. Although several large-scale cancer screens have been carried out in the fly (for example, [Moberg et al., 2001](#); [Tapon et al., 2001](#); [Woodhouse et al., 2003](#); [Pagliarini and Xu, 2003](#); [Zoranovic et al., 2018](#)), our focus was to image and detail primary tumor behavior and progression in the living animal. By combining sophisticated *Drosophila* genetic techniques with transgenic RNAi technology we present here a detailed systematic loss-of-function (LOF) analysis that has identified genes that enhance or suppress tumor progression in this epithelium. We identify a number of conserved invasion suppressors that promote tumor cell invasion upon loss of expression. We further characterize components of the cohesin complex, which we find to be an important invasion suppressor and show that cohesin LOF can promote either individual or collective cell invasion, depending on the subunit that is mutated and the degree of effect on cohesin function.

RESULTS

We developed an *in vivo* genetic system in the fly that allows us to (1) generate a patch of tissue on the dorsal thorax that is homozygous mutant for a tumor suppressor, surrounded by wild-type (WT) tissue; (2) specifically label the mutant tissue with GFP:Moe (the actin-binding domain of moesin fused to GFP), thereby labeling the actin cytoskeleton of these cells; and (3) overexpress an RNAi transgene to deplete expression of a gene of interest specifically within the mutant, labeled tissue. Coupled with our ability to image this epithelium in the living animal in high temporal and spatial resolution ([Couto et al., 2017](#)), this system allowed us to conduct a large-scale genetic screen to identify genes that affect tumor behavior and tumor progression in a wide variety of ways.

Design of an *In Vivo* Assay to Identify Modulators of Epithelial Tumor Progression

We combined the Flp/FRT system ([Xu and Rubin, 1993](#)) the MARCM technique ([Lee and Luo, 2001](#)), and Pannier-Gal4 to generate positively marked homozygous mutant clones specifically within the epithelium of the fly pupal notum (the dorsal thorax). When imaging GFP:Moe-labeled WT clones within the pupal notum (at 20–24 h APF [after puparium formation]) we observed columnar epithelial cells that formed an organized monolayer on the back of the fly ([Figure 1A and A'](#)). Preparatory experiments identified *lethal* (2) *giant larvae*⁴ homozygous mutant clones (*lgl*⁴) as a suitable genetic background for our screen, as tumors lacking *lgl* were large, partially multilayered, and presented a low-level invasive phenotype, representing an ideal scenario for an enhancer/suppressor screen ([Figures 1B–1D](#)). *lgl* is highly conserved, critical for the correct maintenance of cell polarity, and has also been found to control tissue growth and differentiation ([Stephens et al., 2018](#)). *lgl* is a member of the scribble polarity complex, constituents of which (*lgl*, scribble, *dlg*) have been termed “neoplastic” tumor suppressors because mutations in these genes can generate highly disorganized multilayered tumors that are immortal, fail to differentiate, and show a high metastatic potential upon transplantation ([Bilder, 2004](#); [Gateff, 1978](#)). In addition, expression of scribble complex genes has been shown to be lost or downregulated in numerous types of human cancer ([Lo et al., 2012](#)).

Although multilayered, amorphous, and invasive overgrowth is observed in *lgl*, *scribble*, or *dlg* mutant tissue, overgrowth is not observed when small mutant clones are generated, surrounded by WT tissue; here clones are restrained from overgrowth via a process known as “cell competition.” Mutant cells, despite undergoing excessive cell proliferation, are eliminated from the epithelium by Jun N-terminal kinase (JNK) pathway-mediated apoptosis ([Brumby and Richardson, 2003](#); [Nagata and Igaki, 2018](#)). Both *scribble* and *lgl*⁴ mutants have previously been shown to cooperate with oncogenic Notch overexpression to overcome the effects of cell competition and cause neoplastic overgrowths within the proliferative epithelial primordia known as the *imaginal discs* ([Brumby and Richardson, 2003](#); [Khan et al., 2013](#)). We wanted to see whether we could observe a similar cooperative effect within the pupal notum, which at the developmental stage of our analysis (20–24 h APF), is largely post-mitotic. When generating GFP:Moe-labeled clones of cells expressing activated Notch (*N*^{intra}) in the notum, we observed relatively normal clones, with no effect either on cell shape or on tissue organization, and with no invasive characteristics ([Figures 1E and 1I–J](#)). When overexpressing *N*^{intra} specifically within *lgl*⁴ clones, however, we observed a strong cooperative effect—these clones showed strong hyperproliferation, with increased levels of cell division, loss of normal epithelial architecture, and increased invasion when compared with *lgl*⁴ alone ([Figures 1F–1J](#)). We therefore had generated an *in vivo* system that would allow us to identify mutations that work cooperatively with *lgl*⁴ to promote tumor progression.

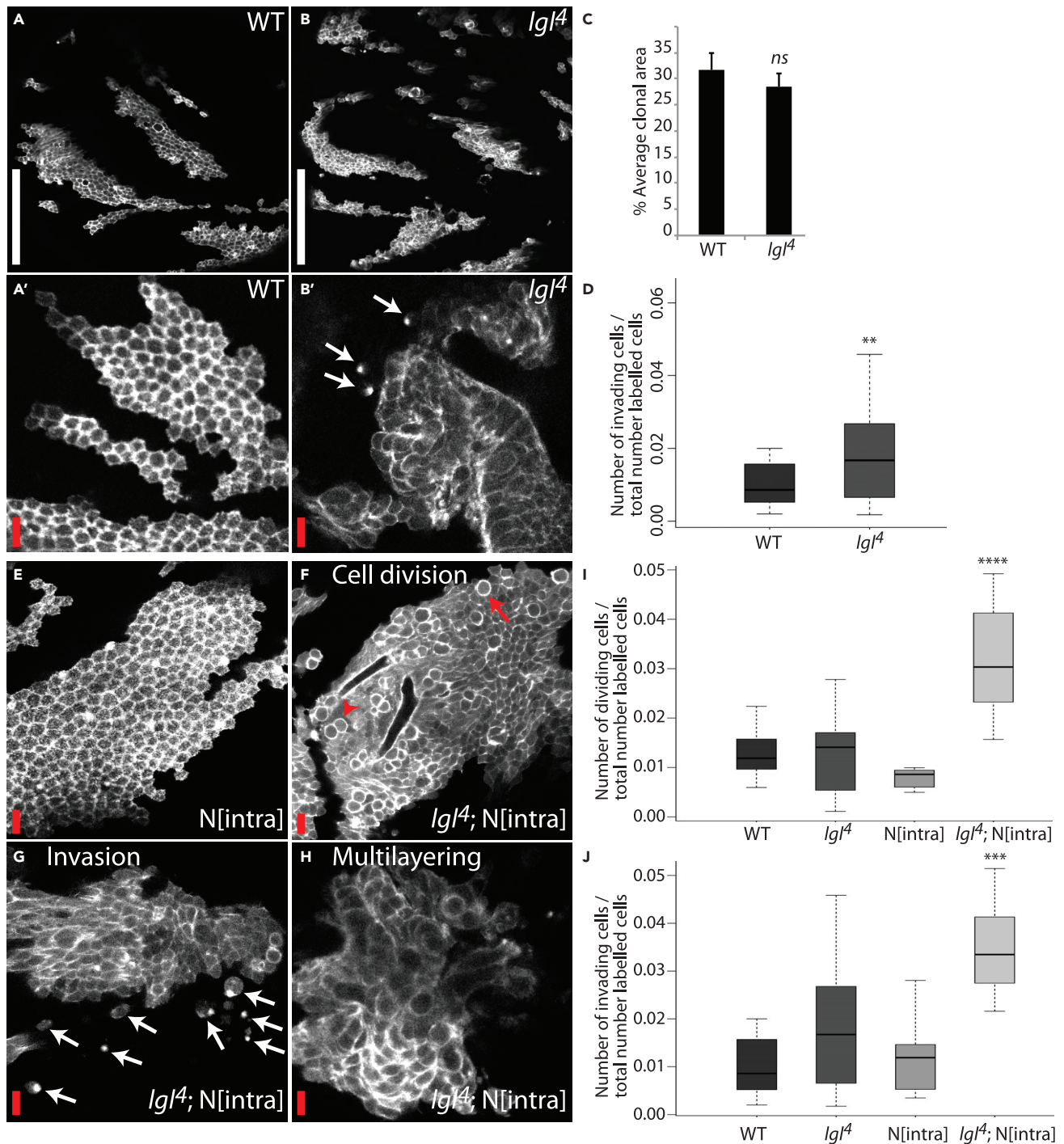


Figure 1. *Igl4* Mutant Clones Provide an Ideal Genetic Background for an Enhancer/Suppressor Screen for Tumor Progression

(A and B) GFP:Moe-labeled genetic clones in the dorsal thorax epithelium of living fly pupae. Clones shown are wild-type (A and A') or homozygous mutant for the neoplastic tumor suppressor *Igl* (B and B').

(C and D) Quantification of average clonal area (C) ($n = 10$ [WT]; 18 [*Igl4*]) and the number of invading cells/the total number of labeled cells (D) ($n = 30$ [WT]; 41 [*Igl4*]). Quantification shows *Igl4* mutant clones to be similar to WT clones in size, with a significant increase in the number of invading cells.

(E–H) GFP:Moe-labeled genetic clones in the dorsal thorax epithelium of living fly pupae. Clones shown are overexpressing activated Notch (*N^{intra}*; (E) or simultaneously homozygous mutant for *Igl4* and overexpressing *N^{intra}* (F–H). Highlighted are effects on cell division (F), invasion (G), and multilayering (H).

Figure 1. Continued

(I and J) Quantification of the number of dividing cells (I) and the number of invading cells (J) over the total number of labeled cells for clones with the genotypes shown (n = 30 [WT]; 41 [*Igl^d*]; 7 [*N^{intra}*]; 13 [*Igl^d*; *N^{intra}*]). Error bars represent $\pm$ SEM. Student's t test (E) and Kruskal-Wallis test (F, K, and L) were performed to determine statistical significance. $p > 0.05$ was considered not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Red arrow, dividing cell; red arrowhead, cell doublet following cytokinesis; white arrows, invading cells. White scale bar, 50 μ m; red scale bar, 10 μ m.

Pilot Screen

During an initial pilot screen, candidate genes previously implicated in cancer were studied. These genes were well characterized and therefore were very likely to present a phenotype. Also included were negative controls, i.e., RNAi lines to genes that are not normally expressed in this tissue. We used transgenic UAS-RNAi lines, which together with pannier-Gal4 and MARCM allowed us to restrict gene knockdown (KD) to *Igl^d* mutant tissue on the notum of the fly (Figure 2A). We used RNAi lines from two near-genome-wide RNAi libraries (VDRC, Austria, and NIG, Japan) and where possible used two independent RNAi transgenes to knock down gene expression for each gene. In total, the pilot consisted of 67 RNAi lines targeting 46 well-known genes (see Table S1 for a list of pilot genes). These candidates included various oncogenes, tumor suppressor genes, matrix metalloproteinases, and regulators of cell morphogenesis, with a range of biological functions (Figure 2B).

We observed a wide range of phenotypes in the pilot screen including hyperproliferation, multilayering, invasion, and effects on subcellular structures (junctions, microvilli, basal protrusions; Figures 2C–2K). Negative controls failed to generate significant phenotypes. We saw a range of expected phenotypes, for example, increased clonal coverage following RNAi of the known tumor suppressor, *Tsc1* (a negative regulator of Tor signaling); reduced clonal coverage following RNAi of a known promoter of the cell cycle, *tkv* (promotes Dpp signaling); increased multilayering following RNAi of the polarity determinants *scrib*, *expanded*, and *dlg*; and smaller apices following RNAi of *Cdc42*, as has been observed previously (Georgiou et al., 2008) (Table S1).

Following the successful completion of the pilot screen, we went on to screen a total of 764 RNAi lines corresponding to 497 individual genes. Recent advances in genome characterization technologies have uncovered a plethora of candidate genes across numerous tumor types that have been found to be commonly mutated or misregulated in human cancers (Beroukhi et al., 2010; International Cancer Genome et al., 2010; Lawrence et al., 2014). However, other than being implicated by these new technologies, many are completely uncharacterized. By screening *Drosophila* orthologs of these previously implicated cancer genes we sought to determine which of these genes affect tumor behavior and drive tumor progression in our system.

Systematic High-Throughput Scoring and Quality Control

We generated a database, whereby we could systematically score specific aspects of tumor behavior, allowing us to record an extremely detailed analysis of how each gene KD affected tumor behavior (see Table S1 for full database). This database consists of 33 phenotypic categories where each animal with *Igl^d* + RNAi KD clones is scored relative to animals with *Igl^d* clones alone. Each category describes an aspect of tumor behavior. Categories include clone size and shape, number of dividing cells, number of invading cells, apex size, junction defects, cytoskeletal defects, multilayering, etc. The scoring system we employed reflected the fact that gene KD could either positively or negatively affect specific aspects of tumor behavior (Figure S1). A minimum of five animals were analyzed per gene KD, and each animal was scored blind by two researchers. An online searchable database with all results from the screen, including all high-resolution images for each RNAi line, is available at <https://flycancerscreen.nottingham.ac.uk>.

To verify that our high-throughput qualitative scoring system gave meaningful results that represented real changes in tumor behavior, we performed a careful quantitative analysis on a selection of genes chosen at random for categories that were amenable to a simple quantitative analysis. As shown in Figures S2A–S2D, a strong positive correlation was observed for all categories measured (0.91–0.97, Spearman correlation test). To further evaluate the quality of our dataset, we asked whether two independently generated RNAi lines targeting the same gene produced similar phenotypes. We compared scores across categories for each pair of RNAi lines and found that, of the 256 genes that were targeted by two independent RNAi lines, 224 (87.5%) gave statistically similar phenotypes (Figures S2E–S2J; Table S2).

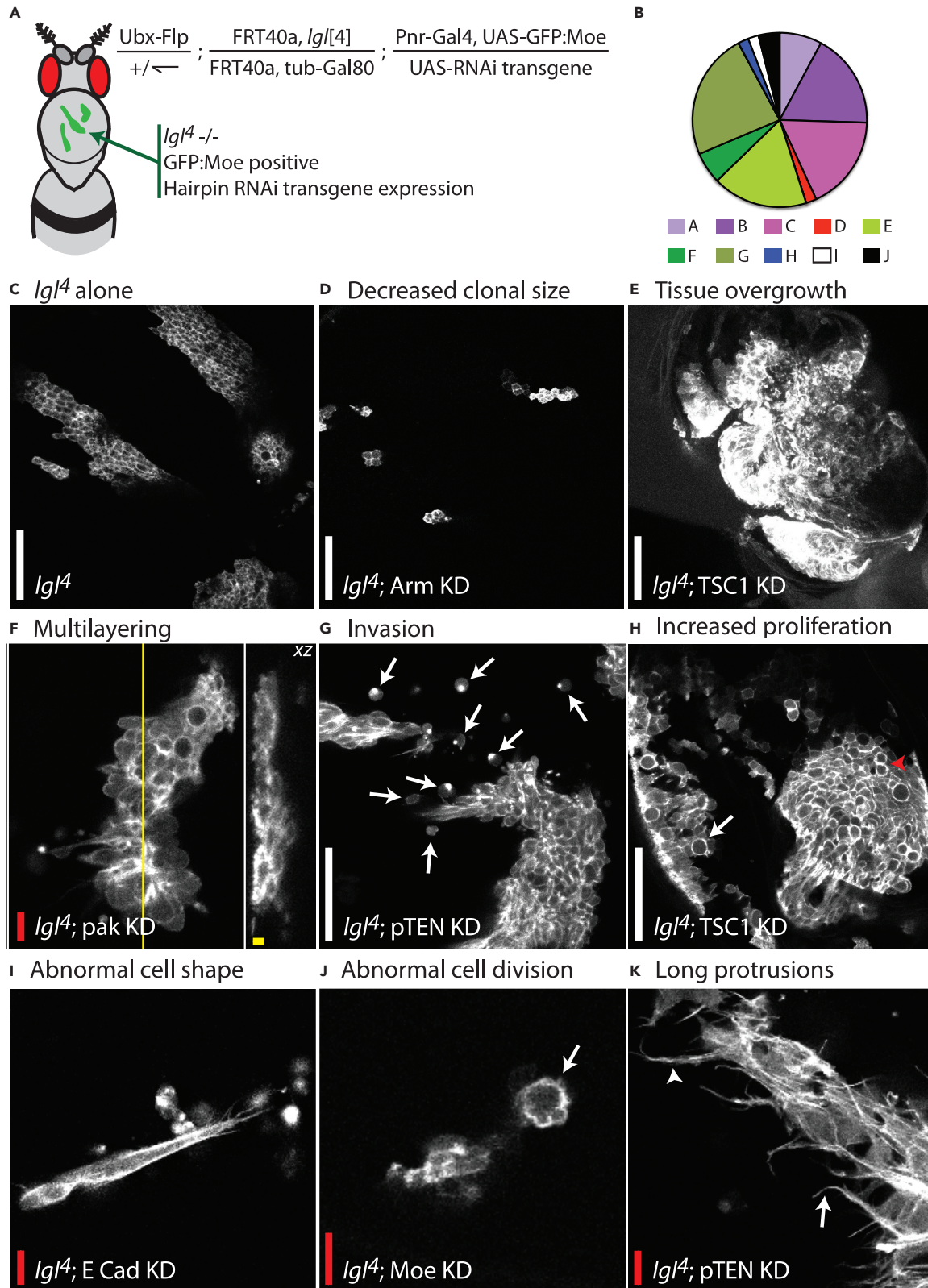


Figure 2. Pilot Screen Identifies Several Modulators of Tumor Behavior

(A) Schematic illustrating how clones with distinct genotypes were generated on the back of the fly. The MARCM system was employed to generate mutant clones specifically within the fly dorsal thorax, through the use of Ubx-Flp. This generated GFP:Moe-labeled *Igf⁴* homozygous mutant clones. RNAi transgene expression, and therefore gene KD, was restricted to the labeled *Igf⁴* mutant tissue.

(B) Pie chart illustrating the range of biological functions from those genes included in the pilot screen. (A) apicobasal polarity, (B) cell adhesion, (C) cytoskeleton, (D) axon guidance, (E) cell cycle, (F) gene expression, (G) signaling, (H) mitochondria, (I) others, (J) unknown.

(C–K) Examples of phenotypes observed within the pilot screen. In the pilot screen we observed effects on clonal size (D and E), tissue morphology (E and F), cell morphology (I and K), and cell behavior (G, H, and J). These are just a few examples of the many distinct phenotypes that we observed. Panel (C) shows *Igf⁴* clones for comparison. Arrows: (G) invading cells, (H) dividing cells, (J) a blebbing dividing cell, and (K) very long basal protrusions. Arrowheads: (H) cell doublet following cytokinesis and (K) long protrusions joining to form a fascicle. White scale bar, 50 μ m; red scale bar, 10 μ m; yellow scale bar, 10 μ m in xz plane.

Identification of Genes that Affect Tumor Behavior

We used an unbiased approach to identify candidate genes that increase or decrease specific aspects of tumor progression in our system. We calculated a mean score for each of the 764 RNAi lines across each of the 33 phenotypic categories (see <https://flycancerscreen.nottingham.ac.uk>). Using these averages, we determined the distribution of scores for all 33 categories. Genes with a mean score above or below the interquartile range from the median were selected as genes of interest. For categories with a two-tailed distribution we were able to identify genes that when knocked down either positively or negatively regulate a specific aspect of tumor behavior. For example, using this methodology we identified 66 RNAi lines that promote and 49 RNAi lines that inhibit cancer cell invasion (mean scores range from +0.73 to +1.5, and –0.55 to –1.2, respectively). See Table S3 for a full list of hits for all categories.

To identify genes that regulate similar or related cell behaviors, we clustered RNAi lines based on phenotypes presented across all categories. This resulted in the identification of 10 phenotypic clusters (Figure 3A). Analysis of the hierarchical clustering revealed, for example, that Cluster 8 shows decreased clonal tissue and increased tissue multilayering and cell body rounding (Figure 3A). Gene ontology (GO) term analysis shows enrichment in junction assembly, cell adhesion, cell differentiation, and fate specification factors (Table S4). A more general categorization of gene function reveals an increase in apicobasal polarity and cell adhesion factors (Figure S3). Therefore, Cluster 8 includes factors that are crucial to the maintenance of an ordered, monolayered, and polarized epithelium. Thus, cluster analysis reveals groups of genes with similar overall phenotypes that may share similar or related molecular functions. Within these groups lie several uncharacterized genes that we can now classify as tumor suppressors.

We additionally clustered categories based on phenotypes presented across all RNAi lines and identified three distinct category clusters (Figure 3B). Categories that clustered together included those related to (A) actin cytoskeleton regulation, (B) invasion and multilayering, and (C) cell proliferation and cell and tissue morphology. We were particularly interested in the identification of novel genes that promote cancer cell invasion. Interaction networks have become a powerful tool to identify novel disease-associated genes (Sevimoglu and Arga, 2014). To generate a functionally validated interaction map of invasive genes, we combined all hits in three categories that clustered strongly together (Figure 3B): invasion, multilayering, and cell body rounding. For each gene, we searched for physical or genetic interactions, validated by experimental data, including yeast two-hybrid, co-immunoprecipitation, and other interaction data from various databases (see Methods). We maintained interactions only between hit genes from these categories, together with lethals and “linker genes,” which linked hit genes from our screen by one interaction (Figure 4; Data S1). The resulting network includes 321 interactions between 140 genes, 99 of which have not been previously implicated in cancer cell invasion or migration, including nine genes that are completely uncharacterized.

Using MCODE (Molecular Complex Detection) software (Bader and Hogue, 2003) we found seven clusters of highly interconnected nodes (Figure 4). Complex 1 comprises core proteins involved in cytoskeleton organization, including Rac2, Scar, WASp, Arp2, and mbc. Adhesion proteins highly involved in cancer invasion are present in complex 6; Complex 5 is enriched in axon guidance molecules, whereas other identified complexes are enriched in proteins that have not been previously linked to cancer cell invasion, such as Complexes 4 and 7. By integrating hits in invasive categories from our screen, together with protein and genetic interaction data, we have therefore identified a large number of genes that are now implicated in cancer cell invasion.

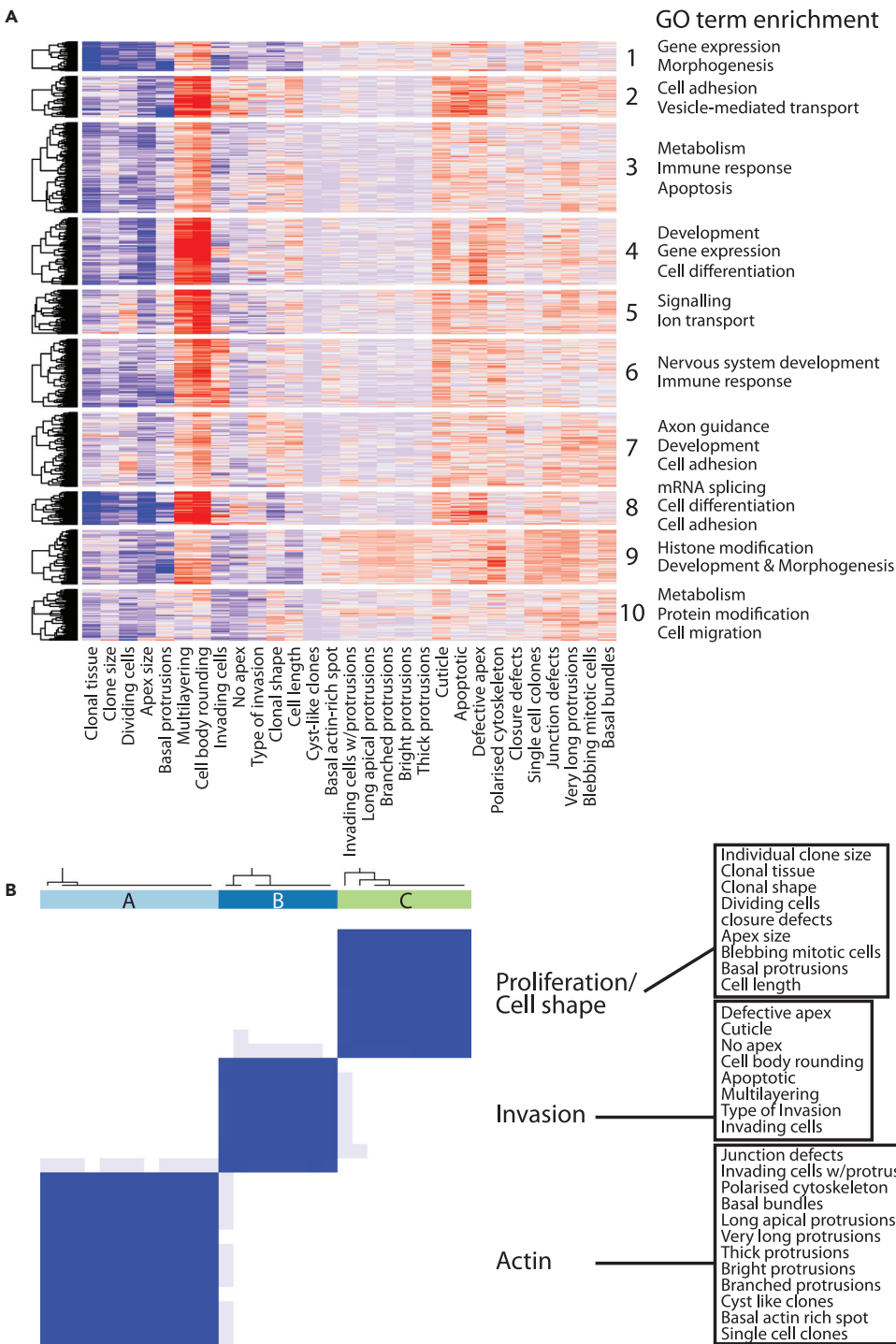


Figure 3. Clustering Analyses Identify 10 RNAi Line Clusters and Three Distinct Phenotypic Subgroups

(A) Heatmap representation of supervised clustering of 764 RNAi lines with average phenotype scores. Each row represents an RNAi line; each column represents a phenotype category. *A priori*, the model-based optimal number of $K = 10$ (phenotypic clusters) was determined. The clustering of rows and columns are based on Euclidean distance. Map colors represent row-scaled average scores: blue indicates the lowest score, light blue indicates an intermediate score, and red indicates the highest score. Each cluster was analyzed with regard to biological function by GO enrichment analysis. The most enriched representative GO categories are shown on the right-hand side of each cluster.

Figure 3. Continued

(B) Consensus clustering of average scores of 29 phenotypic categories reveals three distinct subgroups. Each column represents one phenotype. Heatmaps display consensus values between pairs of phenotypes by blue shading. High consensus corresponds to phenotypes that always occur in the same cluster and is shaded dark blue. See also [Table S4](#) and [Figure S3](#).

Characterization of Invading Cancer Cells

With the aim of characterizing the behavior of individual invading cells, we followed cells within mutant clones over time, before, during, and post-invasion. We found, in all genotypes studied, that pre-invasive cells would round up and form a characteristic actin-rich spot at one side of the cell before invasion ([Figure 5A](#), [Video S1](#)). By calculating the coefficient of determination using Spearman's rho (r_s) we observed a high-to-moderate positive correlation between a polarized actin accumulation and invasion in all genotypes studied, irrespective of whether the mutant clones were rarely invasive or highly invasive ([Figures 5B–5D](#)). The number of cells presenting this polarized phenotype within the epithelial sheet is therefore an indicator of invasive potential.

A major advantage of our *in vivo* model is that the directionality and speed of invading cells can be studied and quantified in real time ([Figures 5A–5I](#)). It was notable that in many cases, invading cells, although viable, have no directionality to their migration and randomly move about over a number of hours ([Figure 5A](#), [Video S1](#)). However, in some cases, as in the case of SA1KD, invading cells appear to be very motile ([Figures 5E–5I](#), [Video S2](#)). Single-cell tracking of *Igl⁴* and SA1KD-invading cells was performed to determine the X, Y, and Z trajectories and to calculate their speed and directionality. An illustration of representative trajectories is shown in [Figures 5F](#) and [5F'](#). To determine directionality, the trajectory of each cell was measured over 30 min. The total number of micrometres traveled was documented (length in [Figures 5G](#) and [5H](#)) as well as the distance an invading cell would have traveled if following a straight line (displacement in [Figures 5G](#) and [5H](#)). [Figure 5H](#) shows a significant increase in length and displacement for SA1KD cells (41.55 μm length, $p < 0.01$; 26.55 μm displacement, $p < 0.05$) when compared with *Igl⁴* cells (16.07 μm length; 4.16 μm displacement). There is no significant difference between length and displacement in SA1KD cells, indicating that their trajectories are directional. In addition, the speed of migration for SA1KD cells was 2.7-fold higher (1.46 $\mu\text{m}/\text{min}$, $p < 0.01$) when compared with *Igl⁴*-invading cells (0.53 $\mu\text{m}/\text{min}$; [Figure 5I](#)). It also became apparent that those cells that migrated in a fast, directional fashion did not possess a single actin-rich spot, but multiple dynamic actin-rich spots ([Figure 5E](#)), and quantification of migrating cells showed that those cells with multiple spots migrated at a significantly faster rate. We additionally found that a low proportion of *Igl⁴*-invading cells can possess multiple actin-rich spots, which also migrate in a directional fashion ([Figures 5J](#) and [5K](#)), indicating that this change in cytoskeletal organization and behavior is important to promote directional migration, irrespective of mutant background.

When imaging pre-invasive and invading cells in the xz plane, we found that cells that are still attached to, or within, the epithelial sheet show very limited lateral movement, and only migrate once they are fully detached from the sheet ([Figures 5L](#) and [5M](#)). We additionally found that invading cells detach from the epithelial sheet more readily in SA1KD clones than in *Igl⁴* clones, which corresponds with SA1KD clones being highly invasive, with invading cells that exhibit directional migration ([Figure 5N](#)).

It has previously been shown that WT epithelial cells delaminate from the pupal notum at early pupal stages, but this delamination is concentrated at the midline region and is rapidly followed by cell death ([Fujisawa et al., 2019](#); [Marinari et al., 2012](#)). This is in stark contrast to the behavior of invading cells within highly invasive tumors in our screen, where invasion is observed irrespective of the clone's position within the epithelial sheet, and invading cells do not undergo immediate cell death (we have imaged invading cells for up to 2 h without observing cell death; for example, see [Figure 5A](#) and [Video S1](#)). To specifically test for the viability of invading cells within highly invasive tumors, we used the genetically encoded apoptosis reporter iCasper ([To et al., 2015](#)). We expressed iCasper within WT clones, *Igl⁴* clones, and in clones for five strong hits for invasion from our screen, namely, *Igl⁴*; CG12268KD, *Igl⁴*; RhoGAP19DKD, *Igl⁴*; Sema1aKD, *Igl⁴*; CG10931KD, *Igl⁴*; CackD. We observed that in four of the five invasive genotypes tested, a high proportion (~70%) of invading cells were iCasper negative. Only WT, *Igl⁴* alone, and *Igl⁴*; CG12268KD mutant clones showed a high proportion of invading cells that were positive for apoptosis (~64%; [Figures 5O](#) and [5P](#)).

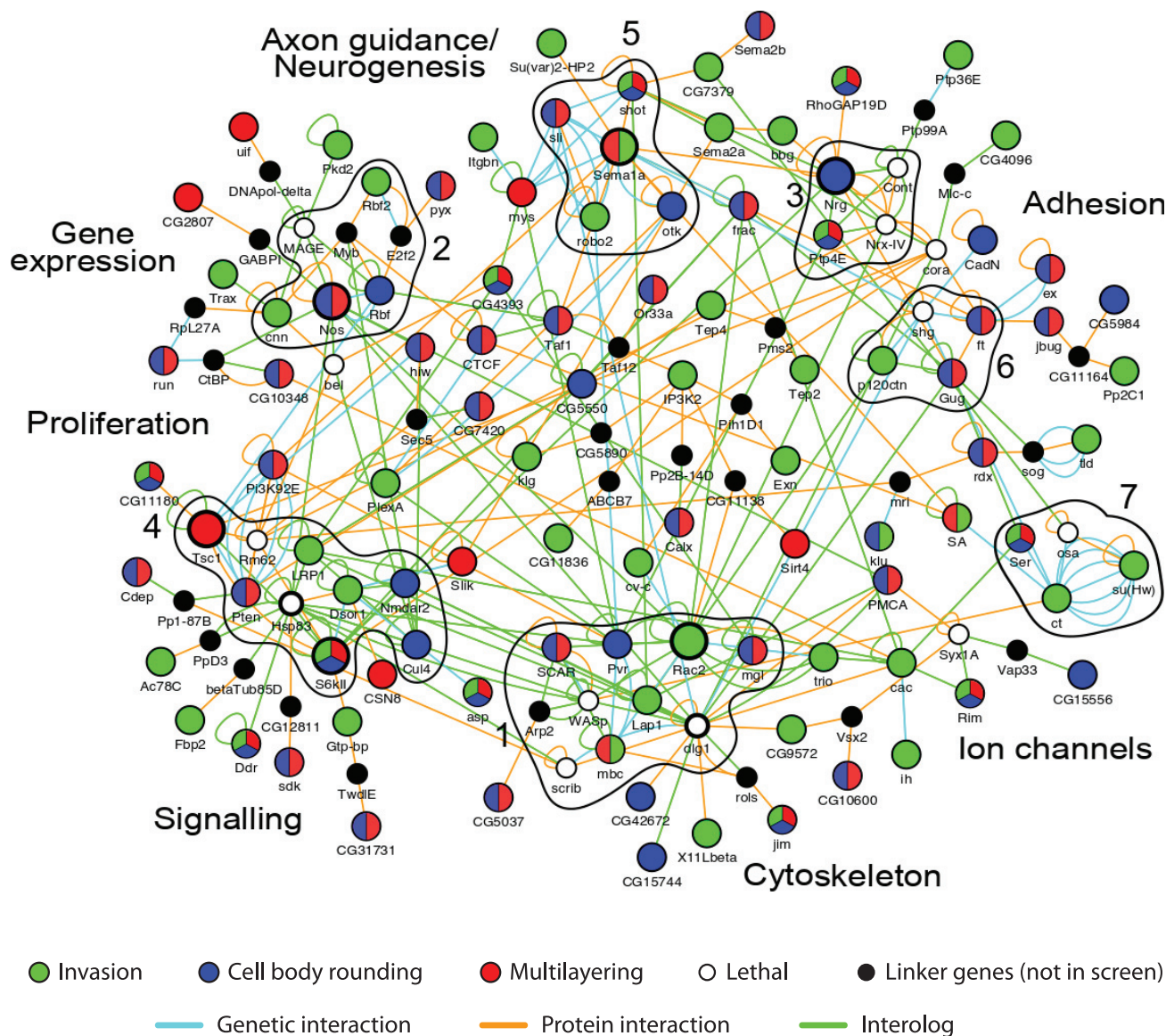


Figure 4. An Interaction Network of Invasion Suppressors

Interactions between genes for which KD enhanced the categories “invasion,” “multilayering,” and “cell body rounding” are shown. Each circle node represents a gene. Node color indicates phenotype observed in the screen: green, invasion; blue, cell-body rounding; red, multilayering; multi-coloured nodes, genes that were hits for more than one phenotype; white, lethal; black, “linker genes,” i.e., genes that were not part of the screen, but that connect screen hit genes by one interaction; nodes with a bold outline, hub genes in this network. Lines represent interactions: cyan, genetic; orange, protein-protein; green, interolog. MCODE complexes of highly interconnected genes are outlined in black. Significantly enriched GO terms are indicated. See also [Data S1](#).

Having identified a number of invasion suppressors in our screen, we wanted to test whether human orthologs of the fly genes within this category would also act in a similar way. We took a panel of five fly genes that (1) strongly promote invasion when their expression is knocked down, and (2) have high-confidence, high-scoring best match human orthologs ([Hu et al., 2011](#)). Genes included were RhoGAP19D, Rim, S6kl1, CG7379, and shot (their closest human orthologs are ARHGAP23, RIMS2, RPS6KA3, ING1, and DST). We designed small interfering RNAs (siRNAs) against these human genes to see if their loss would lead to similar effects in the MCF7 breast cancer cell line. We used an *in vitro* invasion assay to test whether gene KD would promote MCF7 invasion and/or migration. We found a significant increase in both invasion and migration following gene KD of RPS6KA3, ING1, and DST, and a significant increase in migration alone with gene KD of RIMS2 ([Figure S4](#)).

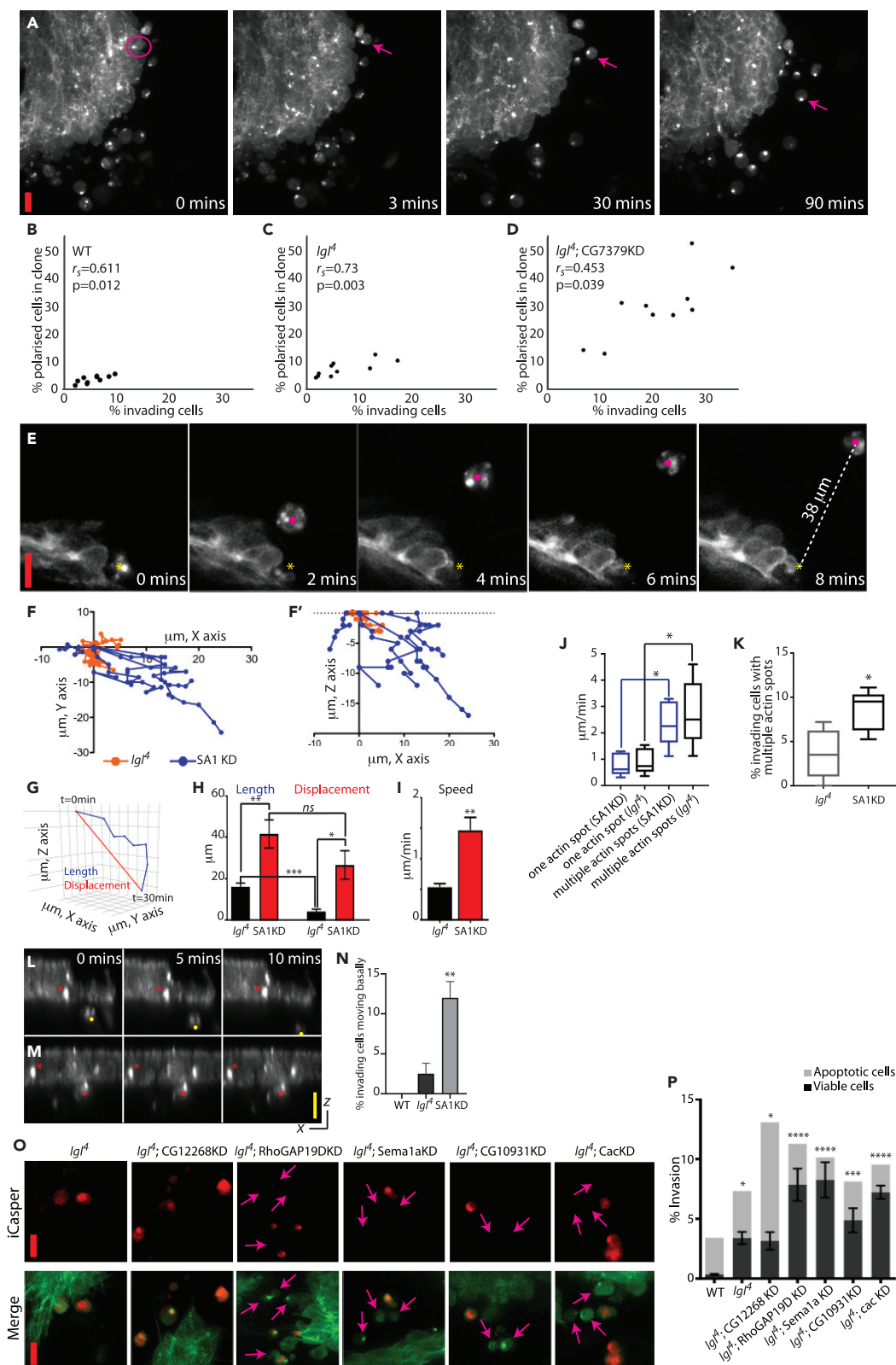


Figure 5. Characterization of Selected Invasion Suppressors

(A) An example of a highly invasive mutant clone (genotype: *Igf1^d*; CG7379KD) labeled with GFP:Moe. Highlighted is a pre-invasive cell that rounds up and forms a characteristic actin-rich spot at one side of the cell before invasion (0 min). The cell then detaches from the mutant clone and migrates away (arrow). (B–D) Correlation between the percentage of clonal cells with a polarized actin accumulation and the percentage of invading cells per animal (n = 10 animals/genotype). (B) WT, (C) *Igf1^d*, (D) *Igf1^d*; CG7379KD. The two parameters show a significant correlation, irrespective of whether the mutant clones were rarely invasive or highly invasive. (E–F') (E) Stills from a time-lapse showing the basal surface of a GFP:Moe-labeled SA1KD clone. Yellow star marks the initial location of an invading cell; magenta dot shows the location of the invading cell at the indicated time. The cell shown has moved 38 μ m in 8 min. (F and F') Representative single-cell trajectories from *Igf1^d* (orange) and SA1KD invading cells (blue) shown in xy (F) and xz (F'). Each cell was measured every 3 min for 30 min. (G) Illustration showing the two trajectories measured for each invading cell to determine directionality. Length (blue) follows the full trajectory of an invading cell. Displacement (red) measures a straight line from the initial to the final point. (H) Quantification of length and displacement from *Igf1^d* and SA1KD cells (n = 25 cells from 5 animals/genotype). Cells that have directionality have no significant difference between length and displacement. (I) Quantification of speed of migration, showing average micrometers traveled per minute (n = 25 cells from 5 animals/genotype). (J) Quantification of speed of migration (μ m/minute) for *Igf1^d* and SA1KD cells that present either a single actin spot or multiple actin spots (n = 5 cells/group). Those with multiple spots travel faster irrespective of genotype. (K) SA1KD cells have a significantly higher proportion of invading cells with multiple actin spots (n = 5 animals/genotype). (L and M) Orthogonal view of invading cells showing that cells only migrate once detached from the epithelial sheet (yellow dot, (L)). Red asterisk, pre-invasive cell within sheet; red dot, delaminated cell still attached to sheet (M). (N) Quantification of the percentage of pre-invasive cells that detach from the epithelial sheet and migrate, in WT, *Igf1^d*, and SA1KD clones (n = 3 animals/genotype). (O) iCasper (red) and GFP:Moe (green)-labeled mutant clones (genotypes specified above panels). Arrows highlight invading cells that are iCasper negative. (P) Four of the five invasive genotypes tested showed a high proportion of invading cells that were iCasper negative (n = 10 animals/genotype). Error bars = $\pm$ SEM. Student's t test or one-way ANOVA with Dunnett's post hoc test for multiple comparisons was performed to determine statistical significance. p > 0.05 was considered not significant, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Red scale bar, 10 μ m; yellow scale bar, 10 μ m in the xz plane. See also Figure S4.

These results provide strong evidence that our *in vivo* system can identify regulators of tumor progression and cancer cell invasion. Results show that in most cases invading cells are non-apoptotic, and that this model can provide additional insight on invading cell morphology and behavior, which can indicate a tumor's invasive potential. Results also suggest that the invasion hits identified in our genetic screen are likely to have relevance to human disease.

The Cohesin Complex Is an Invasion Suppressor

Cohesin is a multi-protein complex that forms a tripartite ring-like structure consisting of the proteins SMC1, SMC3, and RAD21 (Gruber et al., 2003). In addition, RAD21 binds to a stromalin protein (SA1 or SA2, also known as STAG1 or 2 in humans) (Losada et al., 2000; Zhang et al., 2013) (Figure 6A). Therefore two cohesin complexes can form, with cohesin genomic distribution subject to a great degree on the SA/STAG protein that binds to the tripartite ring (Kojic et al., 2018). Cohesin is evolutionarily conserved, with functional cohesin complexes found ubiquitously in all eukaryotic organisms, from yeast to humans (Michaelis et al., 1997; Losada et al., 2000). The cohesin complex is mainly known for its role in sister chromatid cohesion (SCC) (Michaelis et al., 1997); however, current understanding of the possible and numerous roles cohesin may play in tumor initiation and cancer progression is limited (de Koninck and Losada, 2016).

Four subunits of the cohesin complex were studied in our genetic screen: SMC1, SMC3, RAD21, and SA1. Knockdown of these subunits induced significant cytoskeletal changes to *Igf1^d* tumors, including increased multilayering, cell body rounding, and apex defects. In addition, SA1KD significantly enhanced the *Igf1^d* invasive phenotype, with other cohesin subunits having no effect on invasion (Figures 6B–6F). We next knocked down the expression of specific cohesin subunits in WT clones and found that SA1 and SA2KD strongly promoted invasion even in the absence of the *Igf1^d* mutation, whereas the other subunits did not; all subunits, however, promoted multilayering (Figures 6G–6I). Using iCasper we also saw that a high proportion of invading cells evaded apoptosis (Figures 6J–6L) and as shown earlier, showed fast directional migration (Figures 5E–5K; Video S2).

Our screen identified cohesin subunits as affecting epithelial architecture, cell shape, and in the case of SA subunits, promoting frequent cell delamination. These phenotypes therefore implicate effects on adhesion, polarity, and actin regulation as possible underlying influences on the observed cell behavior. We investigated cell-cell adhesion and polarity using antibodies to proteins that localize to the adherens

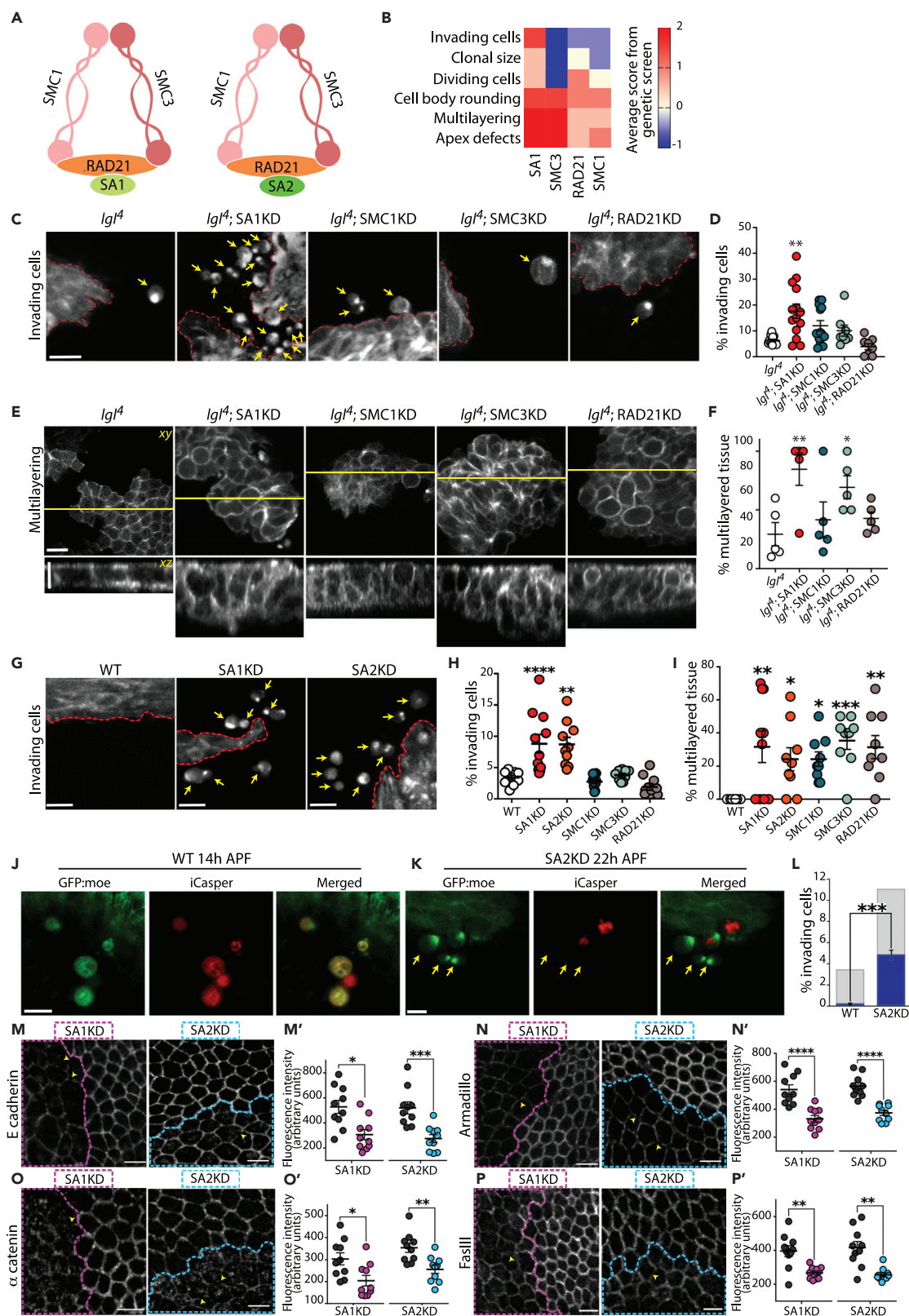


Figure 6. SA1 or SA2KD Promotes Invasion

(A) Somatic cells simultaneously express two different Cohesin rings, differentiated by the presence of either SA1/STAG1 or SA2/STAG2. (B) Heatmap illustrating qualitative scores given to cohesin subunits included in the genetic screen. A subset of categories is shown. Red, enhancement of a phenotype; yellow, no phenotype change; blue, inhibition of a phenotype. (C–F) GFP:moe positively marked *Ig^l* mutant clones with additional cohesin complex subunit KD, showing invading cells (arrows; (C)) and multilayering (E), quantified in (D) and (F); $n = 5$ animals/genotype. Red dashed line highlights edge of clone. Yellow line shows position of xz slice shown. (G) Basal confocal slice of GFP:moe positively marked WT, SA1, or SA2KD clones, highlighting invading cells (arrows). (H and I) Quantification of % invading cells (H) and % multilayering (I) following KD of each cohesin subunit, compared with WT. (J–L) Confocal images of the basal surface of iCasper (red) and GFP:Moe (green)-labeled WT clones (J) and SA2KD clones (K). Arrows highlight invading cells that are iCasper negative. Quantified in (L): Gray, % invading cells/total number of labeled cells; blue, % non-apoptotic invading cells/total number of labeled cells; $n = 50$ cells from 10 animals/genotype. Young WT pupae were used as a control (J) as older WT animals have little to no invading cells. (M–P) SA1 or SA2KD clones, highlighted by magenta and cyan dashed lines, respectively, show disrupted E-cadherin (M, quantified in M'), armadillo (N, quantified in N'), α -catenin (O, quantified in O'), and FasIII (P, quantified in P'), localization. Arrowheads highlight junctional breaks. Quantification shows fluorescence intensity at the level of the junction ($n = 100$ junctions from 10 animals for each genotype). Scale bars, 10 μ m. Error bars = $\pm$ SEM. Student's *t* test or one-way ANOVA with Dunnett's post hoc test for multiple comparisons was performed to determine statistical significance. $p > 0.05$ was considered not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. See also Figures S5 and S6.

junction (AJ), septate junction, and the sub-apical region. We generated SA1 and SA2KD clones and directly compared junction composition inside and outside the clones within the same tissue. A significant reduction in the cortical localization of E-cadherin, α -catenin, β -catenin, and FasIII was observed at the junctional level in both SA1 and SA2KD clones, when compared with the surrounding WT tissue, with evidence of junctional breaks, ectopic structures (puncta, tubules), and mislocalization of junction components (Figures 6M–6P), which are phenotypes that are commonly observed when junctional integrity is compromised (Georgiou et al., 2008). In contrast, KD had no effect on the polarity proteins investigated (dlg and aPKC; Figure S5). These results suggest that SA1 and SA2 act as invasion suppressors in part through the correct localization of junction determinants, thereby maintaining cell-cell junction integrity.

To determine if the role of SA1 and SA2 as invasion suppressors is conserved, we next studied the effect that the loss of their human orthologs, STAG1 and STAG2, would have on MCF7 cell invasion and migration using an *in vitro* invasion assay. LOF mutations of STAG2 are significantly elevated in metastatic breast cancer tumors when compared with lower grades (Repo et al., 2016), suggesting that STAG2 has a role in preventing tumor transition to malignancy. STAG2 is also commonly mutated in several cancer types, including bladder cancer and Ewing sarcoma (Aquila et al., 2018; Tirode et al., 2014). When analyzing each cohesin subunit in turn we found that only STAG1 and STAG2KD promoted invasion and migration, with the core components of the tripartite ring failing to affect cell behavior (Figures S6A–S6I) thereby mirroring the effect we see *in vivo* in the fly (Figures 6G and 6H).

Cohesin is known to influence gene expression. It has been shown in yeast and flies that substantial reductions in cohesin dosage of more than 85% are required to disrupt cohesion and chromosome segregation, whereas small to moderate reductions can affect gene expression (Dorsett and Merckenschlager, 2013). Therefore, the invasive effects that we see in SA/STAG mutants could be due to changes in the expression of genes that affect cell-cell junctions and/or the cytoskeleton. As STAG2 is the most abundant and most mutated cohesin gene in human cancers we performed a microarray gene expression analysis, comparing gene expression in MCF-7 cells post STAG2KD with untreated cells (unt) and with cells treated with non-targeting siRNA (non-T). Of 21,448 genes analyzed, the expression of 23 genes was significantly altered as a result of STAG2KD ($p < 0.01$, fold change [FC] ≥ 1.5 or ≤ -1.5 ; Figure S6 and Table S5). We additionally used RT-qPCR on a selection of genes (STAG2, PCDH1, EHD2 and AKR1B10) to verify the microarray results, with qPCR showing the same or stronger expression change in all cases (Figure S6N).

GO term analysis identified six biological processes that were significantly enriched within the 23 differentially expressed genes, including cell-cell adhesion, protein localization, and cell projection organization (Figure S6O). In addition, an interaction network was generated, using the Cytoscape plugin GeneMania, to display any genetic and physical interactions, verified by experimental data, between the differentially expressed genes and members of the AJ KEGG pathway (Figure S6P; Data S2). Ninety-five interactions between 20 differentially expressed genes and 20 AJ KEGG pathway genes indicate that the differentially expressed genes in STAG2KD cells extensively interact with members of the AJ pathway. Furthermore, EHD2 was significantly downregulated in STAG2KD cells. EHD2 has been linked to E-cadherin localization and expression, and lower EHD2 expression is associated with metastatic tumors (Shi

et al., 2015; Yang et al., 2015). EHD2 links endocytosis to the actin cytoskeleton (Guilherme et al., 2004) and could therefore be influencing E-cadherin's ability to recycle at the junction.

An additional GO term analysis was performed on differentially expressed genes found in two studies that depleted STAG2 expression in cell lines of epithelial origin (MCF10A, Kojic et al., 2018, and HCT116, Casa et al., 2020). Here we found statistically enriched terms including regulation of cell-cell adhesion, regulation of cellular protein localization, regulation of cell-matrix adhesion (Kojic et al., 2018), and positive regulation of cell migration (Casa et al., 2020).

Cohesin Loss of Function Induces the Formation of a Supracellular Actomyosin Ring

Although SA1KD, SA2KD, and SMC3KD promote multilayering (Figure 6I), at an apical level they present a phenotype very similar to WT, with cells presenting an organized geometric shape (Figures 7A, 7B, 7D, and 7E). By contrast, we see a very different phenotype for three cohesin LOF genotypes: *smc3^A* (an ethyl methane sulfonate-induced truncating mutation within *smc3*, K575term, Haelterman et al., 2014; Yamamoto et al., 2014); combined SA1 + SA2KD; and NipBKD (loss of NippedB prevents cohesin from interacting with DNA, Ciosk et al., 2000). These mutants induced a highly distinctive phenotype with drastic cytoskeletal changes, including the formation of a supracellular actin ring (Figures 7C and 7F–7H), eventually followed by clonal extrusion (Figure 7C). It therefore appears that a more severe disruption to cohesin function leads to a very different phenotype to that observed when a single SA subunit is KD. Here individual cell invasion is not observed, rather apical constriction and basal clonal extrusion occur, which are likely to have relevance to the poorly understood process of collective cell invasion in cancer. We further characterized the phenotype using both GFP:Moe to label actin and mCherry:spaghetti squash (*sqh*; the fly ortholog of the regulatory light chain of non-muscle myosin II). We found that the supracellular ring is enriched with actomyosin, which induces the invagination of the mutant tissue, forming a ball of cells with a central lumen (Figures S7B–S7D). We also found significantly elevated levels of E-cadherin within *smc3^A* clones (Figures S7D and S7F), which could also promote clonal invagination through differential adhesion properties between cell types (Steinberg, 2007).

Long time-lapse movies show that over a number of hours the actomyosin ring contracts, inducing a basal clonal extrusion from the epithelial sheet (Figure S7C). Using the caspase sensor, iCasper, we found no significant difference in the levels of apoptosis in *smc3^A* clones, irrespective of whether the clone was still connected to the epithelial sheet or had already extruded (Figure S7G). Furthermore, time-lapse imaging was performed on extruded clones with little increase in iCasper signal observed over 1 h post-extrusion (Figure S7H), indicating that the basal extrusion of *smc3^A* clones does not trigger extensive cell death.

Known mechanisms that trigger apical constriction during development include the apical localization of activated Rho1, which recruits and activates myosin II (Padash Barmchi et al., 2005). We found that Rho1 and *Sqh* are essential for the determination of *smc3^A* cell morphology and actin ring formation, as dominant-negative Rho (RhoN) and *Sqh*KD both inhibit actin ring formation and clonal extrusion, whereas phosphomimetic *Sqh* (*Sqh*-EE) significantly increases the prevalence of this phenotype (Figures 7I–7O).

To better understand the potential mechanism of action of SMC3 in apical constriction and actin ring formation, an enhancer/suppressor screen of genes involved in regulating the localization of myosin II and Rho1 to the apex of the cell was performed. Six candidate genes were KD and, where possible, overexpressed, both alone and in combination with the *smc3* mutation, to determine if these genes enhance or rescue the actin ring and clonal extrusion phenotype. Although four genes promoted actin ring formation in WT clones when overexpressed, only Mad had any significant effect within *smc3^A* clones. Mad overexpression within *smc3^A* clones significantly increased the number of actin rings and delaminated clones (1.196, *n* = 8, *p* < 0.05) when compared with *smc3^A* alone (0.393, *n* = 8), whereas MadKD in *smc3^A* tissue had the opposite effect (0.196, *n* = 8, *p* < 0.01; Figures 7P and 7Q).

Mad is the main effector of the *Drosophila* Dpp signaling pathway. An increase in Dpp signaling has been directly implicated in apical constriction and actin ring formation (Jidigam et al., 2015). Using a phospho-Mad antibody (pMad) we detected a significant increase in pMad levels in *smc3^A* clones and SA1 + SA2KD clones, specifically when these clones contained actin rings (Figures 7R–7T) suggesting that an

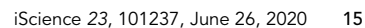


Figure 7. A More Severe Cohesin LOF Induces Actin Ring Formation

(A–G) GFP:moe positively marked WT (A), SMC3KD (B), *smc3^A* (C), SA1KD (D), SA2KD (E), SA1KD; SA2KD (F), and NipBKD (G) clones. Actin rich rings (yellow arrows) were observed in *smc3^A*, SA1, and SA2 simultaneous KD and NipBKD clones.

(H) Quantification of the number of actin rings per square millimeter of clonal tissue. Eight animals were analyzed for each genotype.

(I–O) GFP:moe positively marked clones ((I) *smc3^A*; (J) *smc3^A* + RhoV14; (K) *smc3^A* + RhoN; (M) *smc3^A* + Sqh EE; (N) *smc3^A* + SqhKD). Dominant-negative Rho (RhoN) and SqhKD inhibit actin ring formation in *smc3^A* clones; phosphomimetic Sqh (SqhEE) increases the number of clones with actin rings. Quantified in (L) and (O) showing the number of actin rings or delaminated clones per square millimeter of clonal tissue. Each dot represents one animal. *smc3^A* + RhoV14 resulted in very small unicellular clones (J) or no clones at all and could not be quantified.

(P and Q) Genes involved in apical constriction were either knocked down or overexpressed in GFP:moe positively marked clones, either on their own (P) or within *smc3^A* clones (Q). Quantification shows the number of actin rings or delaminated clones per square millimeter of clonal tissue. Each dot represents 1 animal.

(R and S) GFP:moe-labeled *smc3^A* (R) and SA1 + SA2KD (S) clones stained for the active form of the Dpp signaling effector, phosphorylated Mad (pMad).

(T) Quantification of mean fluorescence intensity from the nuclei of cells within clones, with and without actin rings, compared with WT tissue within the same animal. 35 nuclei from 7 animals were measured. Each dot represents one animal. Scale bars, 10 μ m. Error bars = $\pm$ SEM. Statistical analysis: Student's t test. $p > 0.05$ was considered not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

See also [Figure S7](#).

increase in Mad activity is necessary to induce apical constriction in cohesin LOF clones. It therefore appears that an upregulation of Dpp signaling is a key determinant for the collective invasion observed in cohesin LOF clones.

Given the known pleiotropic effects of the cohesin complex (on SCC, homologous recombination, genome organization and gene transcription, among others) and given our findings showing that cohesin subunits can regulate individual or collective cell invasion in an apparent dose-dependent manner, we studied the dynamics of chromosomal architecture in dividing cells *in vivo*. We generated WT, *smc3^A*, SA1KD, and SA2KD clones, which were labeled with both GFP:Moe and Histone:RFP and carried out live imaging of dividing cells within these clones. We found the vast majority of *smc3^A* mutant cell divisions were defective in chromosome alignment and/or chromosome separation during metaphase and anaphase, respectively. In contrast, the vast majority of divisions in SA1 and SA2KD cells appeared normal ([Figure S8](#); [Videos S3](#), [S4](#), [S5](#), and [S6](#)) adding to the growing body of evidence to suggest that only a major reduction of cohesin function leads to cohesion and segregation defects ([de Koninck and Losada, 2016](#)).

In summary, this work has (1) identified numerous genes that affect tumor behavior in a wide variety of ways; (2) generated a functionally validated network of invasion-suppressor genes; (3) identified the cohesin complex as an important invasion suppressor that can promote individual or collective invasion, dependent on severity of LOF; and (4) established the fly pupal notum as an excellent *in vivo* system to study tumor progression.

DISCUSSION

By combining the genetic amenability of *Drosophila melanogaster* with the power of RNAi transgenics, we were able to generate tumors with specific genotypes and to monitor tumor behavior in the living animal. The *in vivo* system we have developed offers a number of significant advantages and is particularly suitable to the study of tumor progression and invasion. It enables us to (1) monitor GFP:Moe-labeled tumors *in situ*, surrounded by WT tissue and the native local microenvironment; (2) image tumors in high spatial and temporal resolution over a number of hours or even days post-tumour induction; and (3) KD gene expression specifically within the developing tumor, allowing us to investigate the tumor promoting potential of numerous genes that would be developmentally lethal under classic mutation conditions.

Cancer genomes show extreme heterogeneity, with individual solid organ tumors possessing on average > 50 non-silent mutations in the coding regions of different genes ([Greenman et al., 2007](#); [Jones et al., 2008](#); [Ding et al., 2008](#); [Network, 2011](#)). Breast and colorectal cancers have been found to be the most heterogeneous, with an average of 84 and 76 mutations per tumor, respectively ([Sjoblom et al., 2006](#); [Wood et al., 2007](#)). Further complexity is evident when considering epigenetic alterations that can contribute to tumorigenesis and tumor progression ([Jones and Baylin, 2007](#)). The challenge is to identify those genes, from the many that have been implicated in human cancer, that drive cancer progression. We used our *in vivo* system to investigate a set of almost 500 genes, whose human orthologs have previously been implicated in cancer and have now identified numerous genes that either positively or negatively regulate specific aspects of tumor behavior within an epithelium in a living animal.

To understand tumor transition to malignancy, and to develop new therapeutic strategies, it will be key to paint a detailed picture of the complex signaling processes that occur during tumor progression. Our database incorporates 33 phenotypic categories and therefore offers a unique starting point to elucidate the molecular mechanisms of multiple aspects of tumor progression.

However, our primary focus was invasion, and our screen identified numerous genes that regulate epithelial cancer cell invasion. We generated a functionally validated network of invasive genes; GO term analysis of this network identified several terms that are significantly enriched, indicating processes that are likely to be important for invasion to take place. This includes adhesion, cytoskeletal remodeling, signaling, and intriguingly many axon guidance molecules. The Slit, Robo, and Semaphorin families have been previously implicated as both tumor and metastasis suppressors in breast cancer. SLIT/ROBO signaling has been postulated to prevent invasion by maintaining proper cell-cell adhesion, thereby inhibiting the detachment of tumor cells (Yuasa-Kawada et al., 2009). Many other axon guidance genes have been found to be invasion suppressors in our screen, as have uncharacterized genes that genetically interact with axon guidance genes, opening up an intriguing avenue of future research. It is clear that a loss of polarity and a disruption to normal adhesion are pivotal to promoting the process of invasion. Axon guidance proteins, being heavily involved in developmental processes that require cell movement, could be promoting invasive characteristics via these two fundamental processes.

Our *in vivo* system is furthermore particularly suited to imaging the invasive process. Our observation of characteristic cell shape changes (cell rounding and a polarized actin enrichment) that accompany invasion has been previously reported and associated with invasion (Sahai and Marshall, 2003; Yamazaki et al., 2005). However, an important avenue of future research will be to investigate the morphological and molecular processes that underlie the differential behavior between invading cells with and without directional migration. Cell body rounding would indicate an amoeboid type migration, but the characteristic blebbing of amoeboid migration is only clearly obvious in those cells undergoing directional migration. The use of a membrane (rather than actin-associated) marker together with high-resolution microscopy would help to determine whether the extent of membrane blebbing is an important attribute for directionality in this system. An additional consideration is the genetic simplicity of these tumors. It is evident that, in the fly, where there is less redundancy in key regulatory genes, we are able to generate multilayered, invasive tumors, with just two key mutations, but for many invasion suppressors further cooperative mutations are likely to be required to promote directional migration. Extracellular matrix composition and the presence or absence of a chemotactic gradient are also important considerations for directed migration and will be influencing cell behavior here (Talkenberger et al., 2017).

Our work on the cohesin complex provides an example of how specific phenotypes observed in our screen can inform downstream characterization analyses and provides further validation that our screen is picking up important regulators of tumor progression.

Cohesin was initially identified for its role in SCC in yeast (Guacci et al., 1997; Michaelis et al., 1997) and *Xenopus* (Losada et al., 1998) but has subsequently been found to be involved in homologous recombination-mediated DNA repair, higher-order chromatin structure and transcriptional regulation (Nasmyth and Haering, 2009; Mizuguchi et al., 2014; Phillips-Cremins et al., 2013; Hadjur et al., 2009; Seitan et al., 2013; Zuin et al., 2014). How cohesin performs these multiple roles is not fully understood, but is thought to be largely due to cohesin's ability to hold DNA strands in either *trans* (during cell division) or *cis* (generating chromatin loops) (de Koninck and Losada, 2016). This wide variety of functions complicates our understanding of how cohesin mutations may contribute to cancer progression. Inactivating mutations in genes that encode either the core cohesin subunits or the regulatory proteins that affect cohesin function (e.g., PDS5A/B, WAPL, CDCA5, NIPBL, MAU2, etc.) are common in numerous cancer types, including bladder, melanoma, colorectal, lung, Ewing sarcoma, and myeloid malignancies. Importantly, there is no clear correlation between the presence of cohesin mutations and aneuploidy in many tumor types, with recent studies implicating effects on chromatin structure, transcription, DNA repair, and stem cell/progenitor differentiation as important phenotypes that could promote cancer progression (Hill et al., 2016; de Koninck and Losada, 2016). Although cohesin is essential for cell viability, mutations are likely to reduce the amount of total functional cohesin within the cell, which will affect these diverse cohesin-mediated tasks in different ways, depending on the subunit that is mutated, the nature of the mutation, and the cell type affected. Our work shows that, as each specific mutation impacts cohesin

function in different ways, effects on tumor cell behavior can range from defects in epithelial architecture to the promotion of either individual or collective invasion; the phenotype observed will depend on whether the mutation leads to a modification or a disruption of cohesin function, and the degree of any such disruption.

We found loss of cohesin function to induce different phenotypes related to actin cytoskeleton rearrangement. KD of one subcellular localization subunit, SA1 or SA2, increased invasion, multilayering, and apex defects. Reduced expression of the core subunits, SMC1, RAD21, and SMC3, increased multilayering and apex defects, yet had no effect on invasion. A more severe loss of cohesin function (an LOF *smc3* allele, SA1 + SA2 simultaneous KD, or NipBKD) induced clonal extrusion and collective invasion. Differences in cohesin subunit function (SA1 and SA2 provide subcellular localization; SMC1, SMC3, and RAD21 form the core of the ring) (Gruber et al., 2003) and isoform redundancy (SA1/SA2, SMC1A/SMC1B) (Losada et al., 2000; Revenkova et al., 2004), in combination with the specific dose required for each subunit to efficiently perform its role in either gene expression regulation or SCC (Laugsch et al., 2013), could be key to understanding the different effects observed in this study. Several recent studies have shown that individual loss of SA1 or SA2 has different effects compared with loss of all cohesin (Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017) and that the two SA subunits are not fully functionally interchangeable (Kojic et al., 2018). Therefore, loss of one specific SA subunit will have drastic effects on how cohesin interacts with chromatin and on gene expression. Our *in vivo* experiments in the fly and transcriptomics experiments *in vitro* suggest that loss of SA1 or SA2 induces single-cell invasion by affecting cohesin-mediated gene expression during interphase, with strong effects on junction stability. Our live cell imaging of SA1 and SA2KD cells provides further evidence to suggest that aneuploidy is unlikely to make a major contribution to this phenotype. By contrast, a severe loss of cohesin function due to a loss of functional SMC3 does lead to chromosomal instability, which ultimately leads to a misregulation of DPP signaling and increased E-cadherin levels, followed by clonal extrusion. This phenotype could be due to a combination of chromosomal instability, aneuploidy, and chromatin rearrangement defects.

Limitations of the Study

One limitation of the screen, as is the case for any cancer screen, is the fact that the results presented here describe tumor behavior within a specific tissue and anatomical location (the fly notum) and against a specific genetic background (the underlying mutation being *Igf⁴*). In the fly, just as in humans, one would expect tumors with the same genotype to behave differently in different tissues, and additionally expect different combinations of mutations to result in different phenotypes. Despite this, work carried out in the human breast cancer cell line MCF7 shows that the majority of hits tested give the same phenotypes and thereby will have relevance to human disease. This is most clearly seen when testing cohesin subunits in the fly and in MCF7 cells: STAG1 and STAG2 both promote invasion when their expression is knocked down, whereas other cohesin subunits do not, recapitulating the effect seen within the fly screen.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Marios Georgiou (marios.georgiou@nottingham.ac.uk).

Materials Availability

All unique/stable reagents generated in this study will be made available on request, but we may require a payment and/or a completed Materials Transfer Agreement if there is potential for commercial application.

Data and Code Availability

The accession number for the microarray data reported in this paper is GEO: GSE137773. An online searchable database with all results from the screen, including raw high-resolution images for each RNAi line, is available at <https://flycancerscreen.nottingham.ac.uk>.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101237>.

ACKNOWLEDGMENTS

We wish to thank the fly community for their generosity with reagents, especially the Bloomington, VDRC, and NIG stock centers, without whom this project would not have been possible. We thank Anna Grabowska for the MCF-7 cell line, the School of Life Sciences Imaging (SLIM) for invaluable help with the confocal microscopes, and Louise Cheng and Peter Shaw for critical reading of the manuscript. We thank Zsuzsa Markus for fly husbandry and additionally thank Keith Spriggs for help with preparing image files for the online resource. This work was supported by Cancer Research UK, United Kingdom (grant numbers C36430, A12891). B.C.C. was supported by a University of Nottingham/CONACyT award, United Kingdom and Mexico; A.D.R. was supported by a Nottingham Vice-Chancellor's Scholarship for Research Excellence Award, United Kingdom.

AUTHOR CONTRIBUTIONS

Conceptualization, M.G.; Methodology, M.G., B.C.C., A.C., N.A.M., and Z.E.C.; Investigation, B.C.C., Z.E.C., A.C., N.A.M., A.D.R., U.N., Y.N.F., M.R.H., M.C.U., and A.B.; Formal Analysis, A.L., M.C.U., B.C.C., and A.D.R.; Writing – Original Draft, M.G.; Writing – Review & Editing, M.G., B.C.C., A.C., A.D.R., and A.L. Funding Acquisition, M.G.; Resources, M.G., R.R., and S.T.M.; Supervision, M.G. and S.T.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 31, 2019

Revised: March 30, 2020

Accepted: June 2, 2020

Published: June 26, 2020

REFERENCES

- Aquila, L., Ohm, J., and Woloszyńska-Read, A. (2018). The role of STAG2 in bladder cancer. *Pharmacol. Res.* 131, 143–149.
- Bader, G.D., and Hogue, C.W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2.
- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905.
- Bilder, D. (2004). Epithelial polarity and proliferation control: links from the *Drosophila* neoplastic tumor suppressors. *Genes Dev.* 18, 1909–1925.
- Brumby, A.M., and Richardson, H.E. (2003). Scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in *Drosophila*. *EMBO J.* 22, 5769–5779.
- Brumby, A.M., and Richardson, H.E. (2005). Using *Drosophila melanogaster* to map human cancer pathways. *Nat. Rev. Cancer* 5, 626–639.
- Casa, V., Moronta Gines, M., Gade Gusmao, E., Slotman, J.A., Zirkel, A., Josipovic, N., Oole, E., Van, I.W.F.J., Houtsmuller, A.B., Papantonis, A., and Wendt, K.S. (2020). Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. *Genome Res.* 30, 515–527.
- Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* 5, 243–254.
- Cohen, M., Georgiou, M., Stevenson, N.L., Miodownik, M., and Baum, B. (2010). Dynamic filopodia transmit intermittent Delta-Notch signaling to drive pattern refinement during lateral inhibition. *Dev. Cell* 19, 78–89.
- Couto, A., Mack, N.A., Favia, L., and Georgiou, M. (2017). An apical-basal gradient of Rac activity determines protrusion form and position. *Nat. Commun.* 8, 15385.
- de Koninck, M., and Losada, A. (2016). Cohesin mutations in cancer. *Cold Spring Harb Perspect. Med.* 6, a026476.
- Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., Mclellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B., et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455, 1069–1075.
- Dorsett, D., and Merckenschlager, M. (2013). Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. *Curr. Opin. Cell Biol.* 25, 327–333.
- Fujisawa, Y., Kosakamoto, H., Chihara, T., and Miura, M. (2019). Non-apoptotic function of *Drosophila* caspase activation in epithelial thorax closure and wound healing. *Development* 146, 169037.
- Gateff, E. (1978). Malignant neoplasms of genetic origin in *Drosophila melanogaster*. *Science* 200, 1448–1459.
- Georgiou, M., and Baum, B. (2010). Polarity proteins and Rho GTPases cooperate to spatially organise epithelial actin-based protrusions. *J. Cell Sci.* 123, 1089–1098.
- Georgiou, M., Marinari, E., Burden, J., and Baum, B. (2008). Cdc42, Par6, and aPKC regulate Arp2/3-mediated endocytosis to control local adherens junction stability. *Curr. Biol.* 18, 1631–1638.
- Gonzalez, C. (2013). *Drosophila melanogaster*: a model and a tool to investigate malignancy and identify new therapeutics. *Nat. Rev. Cancer* 13, 172–183.
- Greenman, C., Stephens, P., Smith, R., Dalgleish, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.
- Gruber, S., Haering, C.H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. *Cell* 112, 765–777.

- Guacci, V., Hogan, E., and Koshland, D. (1997). Centromere position in budding yeast: evidence for anaphase A. *Mol. Biol. Cell* 8, 957–972.
- Guilherme, A., Soriano, N.A., Bose, S., Holik, J., Bose, A., Pomerleau, D.P., Furcinitti, P., Leszyk, J., Corvera, S., and Czech, M.P. (2004). EHD2 and the novel EH domain binding protein EHP1 couple endocytosis to the actin cytoskeleton. *J. Biol. Chem.* 279, 10593–10605.
- Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410–413.
- Haelterman, N.A., Jiang, L., Li, Y., Bayat, V., Sandoval, H., Ugur, B., Tan, K.L., Zhang, K., Bei, D., Xiong, B., et al. (2014). Large-scale identification of chemically induced mutations in *Drosophila melanogaster*. *Genome Res.* 24, 1707–1718.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Hill, V.K., Kim, J.S., and Waldman, T. (2016). Cohesin mutations in human cancer. *Biochim. Biophys. Acta* 1866, 1–11.
- Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* 12, 357.
- Hurst, D.R., and Welch, D.R. (2011). Metastasis suppressor genes at the interface between the environment and tumor cell growth. *Int. Rev. Cell Mol. Biol.* 286, 107–180.
- International Cancer Genome, C., Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., et al. (2010). International network of cancer genome projects. *Nature* 464, 993–998.
- Jidigam, V.K., Srinivasan, R.C., Patthey, C., and Gunhaga, L. (2015). Apical constriction and epithelial invagination are regulated by BMP activity. *Biol. Open* 4, 1782–1791.
- Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. *Cell* 128, 683–692.
- Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., et al. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321, 1801–1806.
- Khan, S.J., Bajpai, A., Alam, M.A., Gupta, R.P., Harsh, S., Pandey, R.K., Goel-Bhattacharya, S., Nigam, A., Mishra, A., and Sinha, P. (2013). Epithelial neoplasia in *Drosophila* entails switch to primitive cell states. *Proc. Natl. Acad. Sci. U S A* 110, E2163–E2172.
- Kojic, A., Cuadrado, A., de Koninck, M., Gimenez-Llorte, D., Rodriguez-Corsino, M., Gomez-Lopez, G., Le Dily, F., Marti-Renom, M.A., and Losada, A. (2018). Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.* 25, 496–504.
- Laugsch, M., Seebach, J., Schnittler, H., and Jessberger, R. (2013). Imbalance of SMC1 and SMC3 cohesins causes specific and distinct effects. *PLoS One* 8, e65149.
- Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501.
- Lee, T., and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci.* 24, 251–254.
- Lo, P., Hawrot, H., and Georgiou, M. (2012). Apicobasal polarity and its role in cancer progression. *Biomol. Concepts* 3, 505–521.
- Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev.* 12, 1986–1997.
- Losada, A., Yokochi, T., Kobayashi, R., and Hirano, T. (2000). Identification and characterization of SA/SCC3p subunits in the Xenopus and human cohesin complexes. *J. Cell Biol.* 150, 405–416.
- Marinari, E., Mehonic, A., Curran, S., Gale, J., Duke, T., and Baum, B. (2012). Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature* 484, 542–545.
- Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91, 35–45.
- Mirzoyan, Z., Sollazzo, M., Allocca, M., Valenza, A.M., Grifoni, D., and Bellosa, P. (2019). *Drosophila melanogaster*: a model organism to study cancer. *Front. Genet.* 10, 51.
- Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.M., Taneja, N., Folco, H.D., Fitzgerald, P., Dekker, J., Mirny, L., Barrowman, J., and Grewal, S.I.S. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432–435.
- Moberg, K.H., Bell, D.W., Wahrer, D.C., Haber, D.A., and Hariharan, I.K. (2001). Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* 413, 311–316.
- Nagata, R., and Igaki, T. (2018). Cell competition: emerging mechanisms to eliminate neighbors. *Dev. Growth Differ.* 60, 522–530.
- Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* 43, 525–558.
- Network, T.C.G.A.R. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615.
- Padash Barmchi, M., Rogers, S., and Hacker, U. (2005). DRhoGEF2 regulates actin organization and contractility in the *Drosophila* blastoderm embryo. *J. Cell Biol.* 168, 575–585.
- Pagliarini, R.A., and Xu, T. (2003). A genetic screen in *Drosophila* for metastatic behavior. *Science* 302, 1227–1231.
- Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Rao, S.S.P., Huang, S.C., Glenn ST Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin loss eliminates all loop domains. *Cell* 171, 305–320.e24.
- Repo, H., Loytyniemi, E., Nykanen, M., Lintunen, M., Karra, H., Pitkanen, R., Soderstrom, M., Kuopio, T., and Kronqvist, P. (2016). The expression of cohesin subunit SA2 predicts breast cancer survival. *Appl. Immunohistochem. Mol. Morphol.* 24, 615–621.
- Revenkova, E., Eijpe, M., Heyting, C., Hodges, C.A., Hunt, P.A., Liebe, B., Scherthan, H., and Jessberger, R. (2004). Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nat. Cell Biol.* 6, 555–562.
- Rudrapatna, V.A., Cagan, R.L., and Das, T.K. (2012). *Drosophila* cancer models. *Dev. Dyn.* 241, 107–118.
- Sahai, E., and Marshall, C.J. (2003). Differing modes of tumour cell invasion have distinct requirements for Rho/ROCK signalling and extracellular proteolysis. *Nat. Cell Biol.* 5, 711–719.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56.
- Seitan, V.C., Faure, A.J., Zhan, Y., Mccord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* 23, 2066–2077.
- Sevimoglu, T., and Arga, K.Y. (2014). The role of protein interaction networks in systems biomedicine. *Comput. Struct. Biotechnol. J.* 11, 22–27.
- Shi, Y., Liu, X., Sun, Y., Wu, D., Qiu, A., Cheng, H., Wu, C., and Wang, X. (2015). Decreased expression and prognostic role of EHD2 in human breast carcinoma: correlation with E-cadherin. *J. Mol. Histol.* 46, 221–231.
- Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274.
- Sporn, M.B. (1996). The war on cancer. *Lancet* 347, 1377–1381.

- Steeg, P.S. (2003). Metastasis suppressors alter the signal transduction of cancer cells. *Nat. Rev. Cancer* 3, 55–63.
- Steinberg, M.S. (2007). Differential adhesion in morphogenesis: a modern view. *Curr. Opin. Genet. Dev.* 17, 281–286.
- Stephens, R., Lim, K., Portela, M., Kvasnakul, M., Humbert, P.O., and Richardson, H.E. (2018). The scribble cell polarity module in the regulation of cell signaling in tissue development and tumorigenesis. *J. Mol. Biol.* 430, 3585–3612.
- Talkenberger, K., Cavalcanti-Adam, E.A., Voss-Bohme, A., and Deutsch, A. (2017). Amoeboid-mesenchymal migration plasticity promotes invasion only in complex heterogeneous microenvironments. *Sci. Rep.* 7, 9237.
- Tapon, N., Ito, N., Dickson, B.J., Treisman, J.E., and Hariharan, I.K. (2001). The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. *Cell* 105, 345–355.
- Tirode, F., Surdez, D., Ma, X., Parker, M., le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetete-Lalami, S., Rusch, M., et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* 4, 1342–1353.
- To, T.-L., Piggott, B.J., Makhijani, K., Yu, D., Jan, Y.N., and Shu, X. (2015). Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. *Proc. Natl. Acad. Sci. U S A* 112, 3338–3343.
- Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* 318, 1108–1113.
- Woodhouse, E.C., Fisher, A., Bandle, R.W., Bryant-Greenwood, B., Charboneau, L., Petricoin, E.F., 3rd, and Liotta, L.A. (2003). *Drosophila* screening model for metastasis: Semaphorin 5c is required for l(2)gl cancer phenotype. *Proc. Natl. Acad. Sci. U S A* 100, 11463–11468.
- Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599.
- Xu, T., and Rubin, G.M. (1993). Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* 117, 1223–1237.
- Yamamoto, S., Jaiswal, M., Charng, W.L., Gambin, T., Karaca, E., Mirza, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A *drosophila* genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 159, 200–214.
- Yamazaki, D., Kurisu, S., and Takenawa, T. (2005). Regulation of cancer cell motility through actin reorganization. *Cancer Sci.* 96, 379–386.
- Yan, J., Yang, Q., and Huang, Q. (2013). Metastasis suppressor genes. *Histol. Histopathol.* 28, 285–292.
- Yang, X., Ren, H., Yao, L., Chen, X., and He, A. (2015). Role of EHD2 in migration and invasion of human breast cancer cells. *Tumour Biol.* 36, 3717–3726.
- Yuasa-Kawada, J., Kinoshita-Kawada, M., Rao, Y., and Wu, J.Y. (2009). Deubiquitinating enzyme USP33/VDU1 is required for Slit signaling in inhibiting breast cancer cell migration. *Proc. Natl. Acad. Sci. U S A* 106, 14530–14535.
- Zhang, N., Jiang, Y., Mao, Q., Demeler, B., Tao, Y.J., and Pati, D. (2013). Characterization of the interaction between the cohesin subunits Rad21 and SA1/2. *PLoS One* 8, e69458.
- Zoranovic, T., Manent, J., Willoughby, L., Matos de Simoes, R., la Marca, J.E., Golenkina, S., Cuiping, X., Gruber, S., Angjeli, B., Kanitz, E.E., et al. (2018). A genome-wide *Drosophila* epithelial tumorigenesis screen identifies Tetraspanin 29Fb as an evolutionarily conserved suppressor of Ras-driven cancer. *PLoS Genet.* 14, e1007688.
- Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., Van, I.W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U S A* 111, 996–1001.

Supplemental Information

A Genetic Analysis of Tumor Progression in *Drosophila* Identifies the Cohesin Complex as a Suppressor of Individual and Collective Cell Invasion

Brenda Canales Coutiño, Zoe E. Cornhill, Africa Couto, Natalie A. Mack, Alexandra D. Rusu, Usha Nagarajan, Yuen Ngan Fan, Marina R. Hadjicharalambous, Marcos Castellanos Uribe, Amy Burrows, Anbarasu Lourdusamy, Ruman Rahman, Sean T. May, and Marios Georgiou

Figure S1

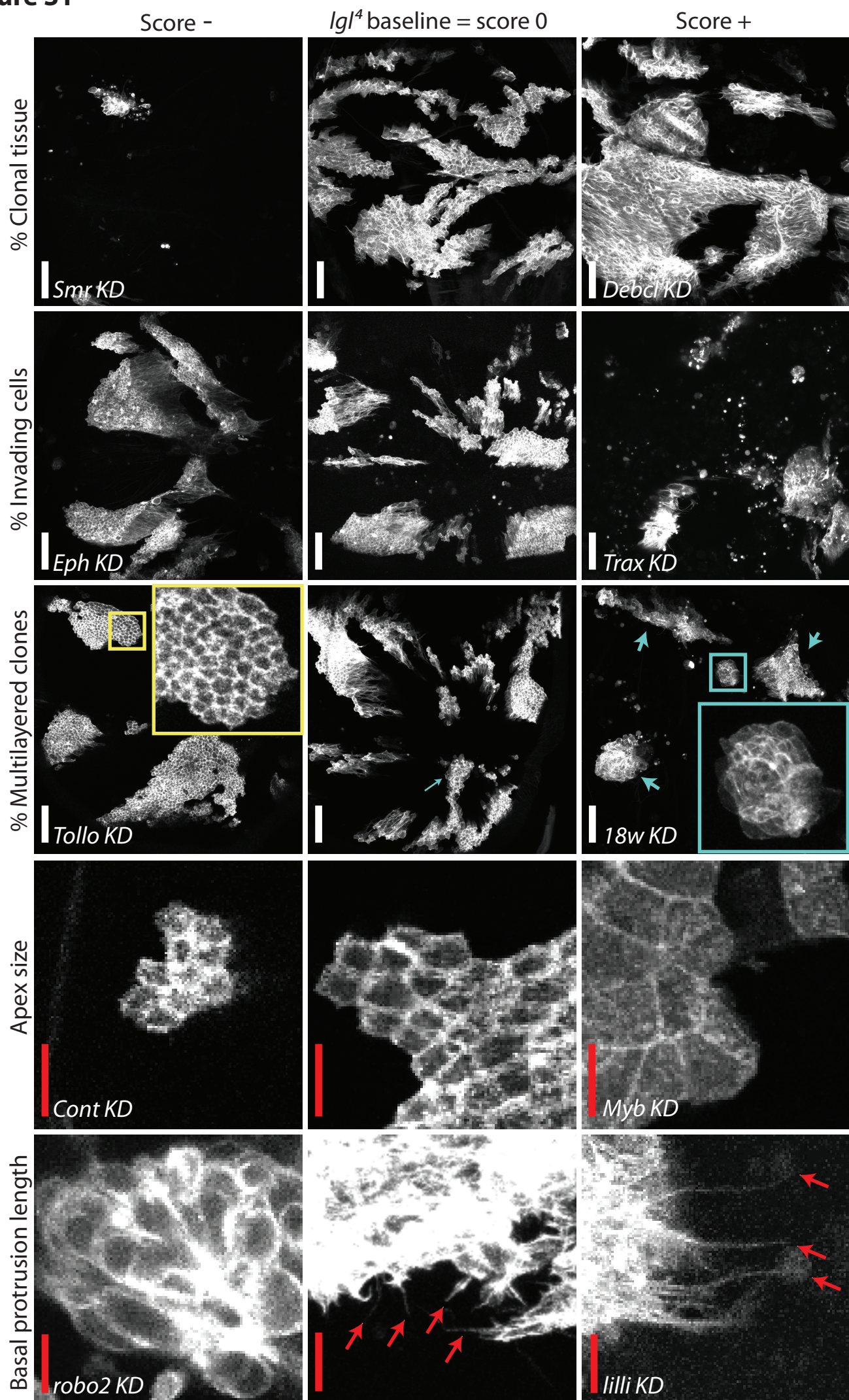
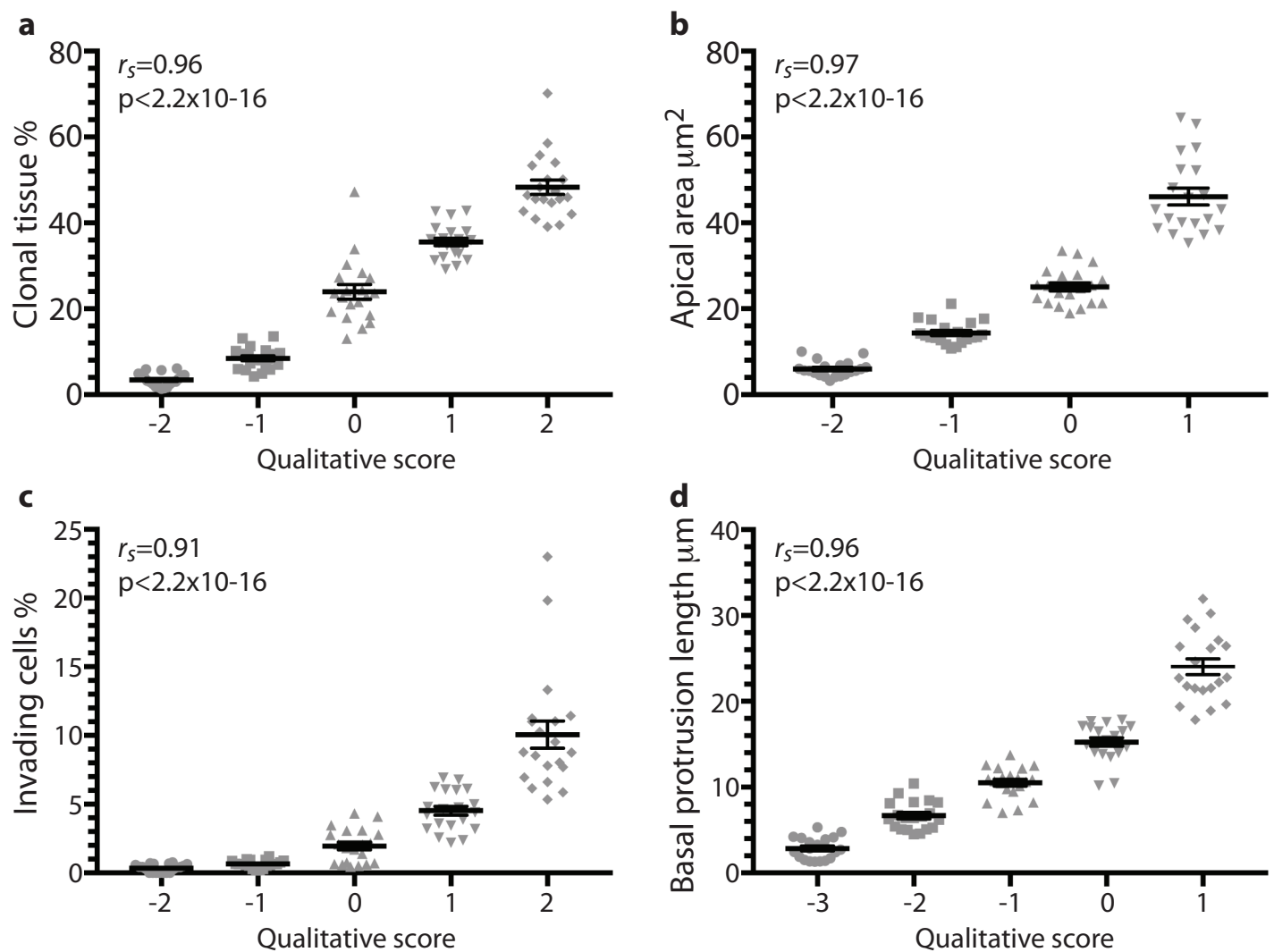


Figure S1: Semi-quantitative scoring system employed to systematically record phenotypes across phenotypic categories. Related to Table S1 and Figures 2-3.

Rows show five example categories from the 33 categories within the database to illustrate our scoring method: percentage of clonal tissue within the fly pupal notum (i.e. GFP-positive area versus GFP-negative area); percentage of invading cells against total GFP-positive tissue; percentage of multilayered clones against total GFP-positive tissue; average apex size of epithelial cells; average length of basal protrusions. Columns show scores given: central column shows a score of 0 for the *lgl⁴* mutant background; right column shows positive scores of +1 or +2, which indicate an increase in phenotype compared to the *lgl⁴* baseline; left column shows negative scores of -1 or -2, which indicate a decrease in phenotype compared to *lgl⁴* baseline. Panel names on right and left columns refer to the specific gene knocked down in an *lgl⁴* mutant background. Yellow box indicates a magnified monolayered clone. Turquoise box indicates a magnified multilayered clone; turquoise arrows highlight multilayered clones within the notum (others are monolayered). Red arrows point to basal protrusions. White scale bar: 10µm; red scale bar: 5µm.

Figure S2



Invasion

Multilayering

Long basal protrusions

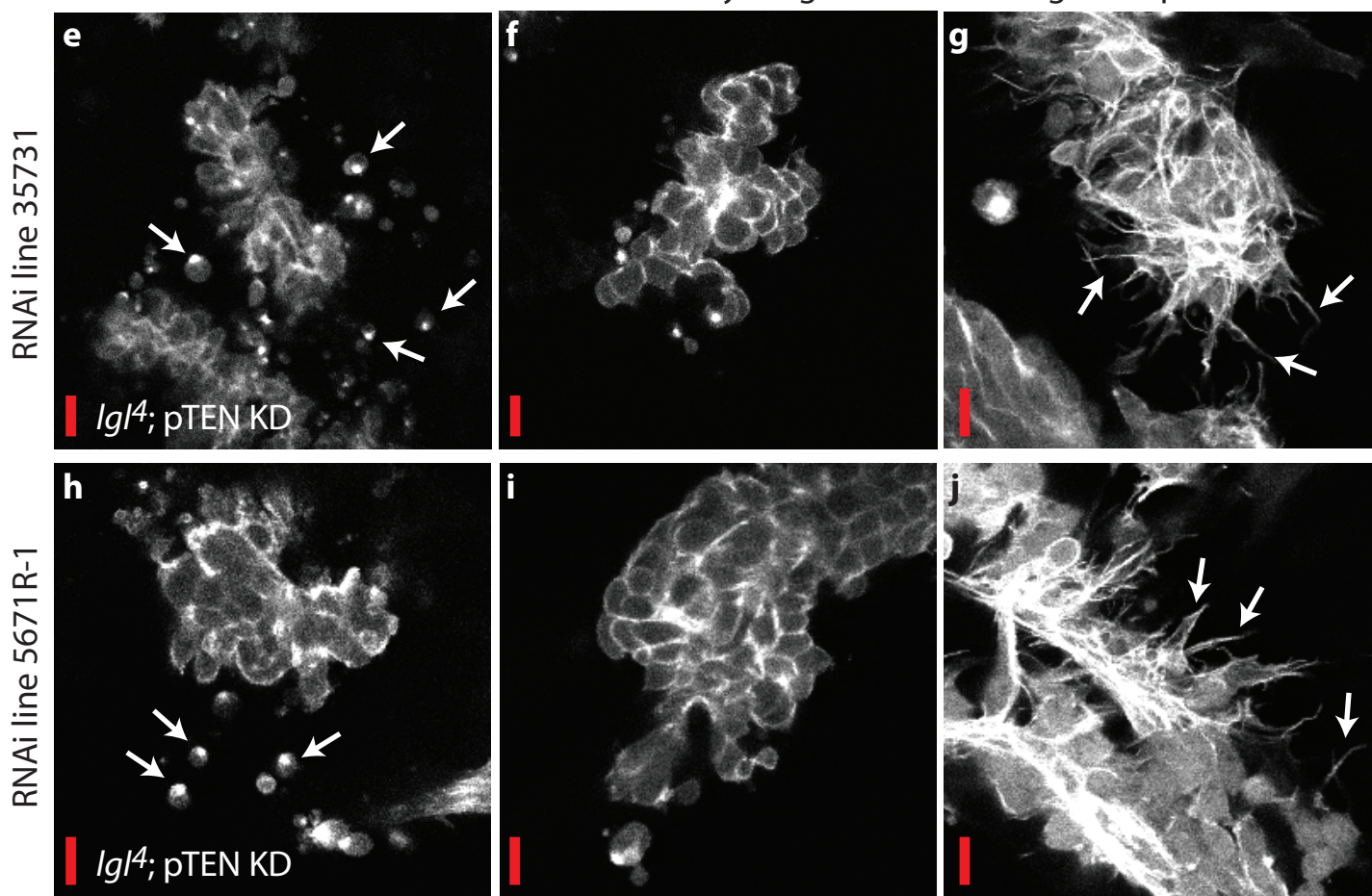


Figure S2: Quality control. Related to Table S2 and Figure 3.

(a-d) Graphs comparing qualitative scores given to individual animals against the quantitative measure for a range of phenotypes, including: (a) percentage clonal tissue; (b) apical area; (c) percentage invading cells (d) basal protrusion length (n= 20 animals for each category of qualitative score). (e-j) A comparison of the phenotypes observed when two independent RNAi lines (35731-GD and 5671R-1) were used to target the same gene (pTEN). Both showed strong effects on invasion (e and h), multilayering (f and i), and basal protrusion length (g and j). Arrows highlight: invading cells (f and i); long basal protrusions (g and j). Scale bars = 10 μ m. Error bars represent $\pm$ s.e.m.

Figure S3

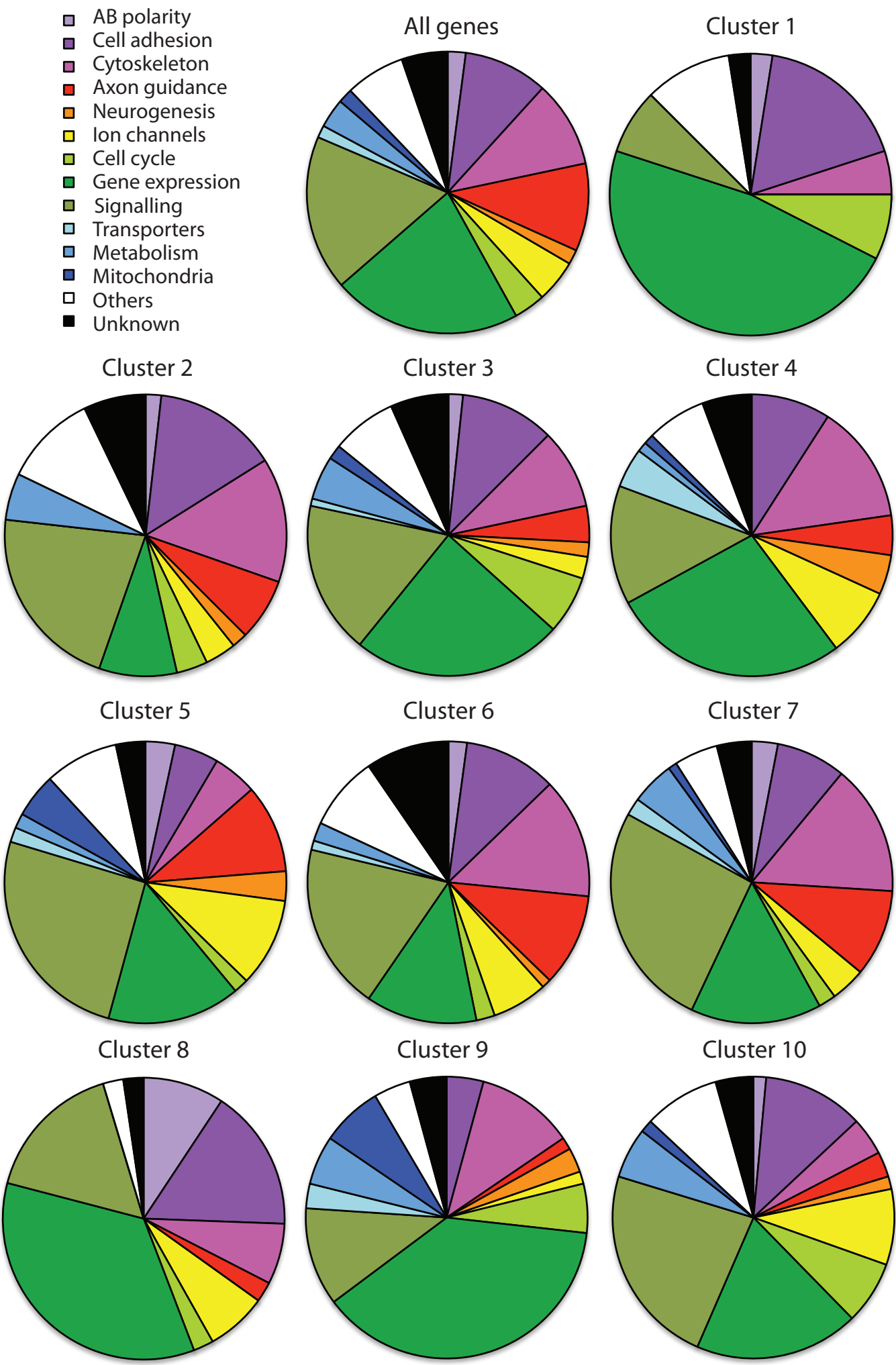


Figure S3: An analysis of gene function for genes within phenotypic clusters. Related to Figure 3. Pie charts illustrating the range of biological functions for all 497 genes included in the screen (all genes), as well as those functions that were enriched within phenotypic clusters (clusters 1-10).

Figure S4

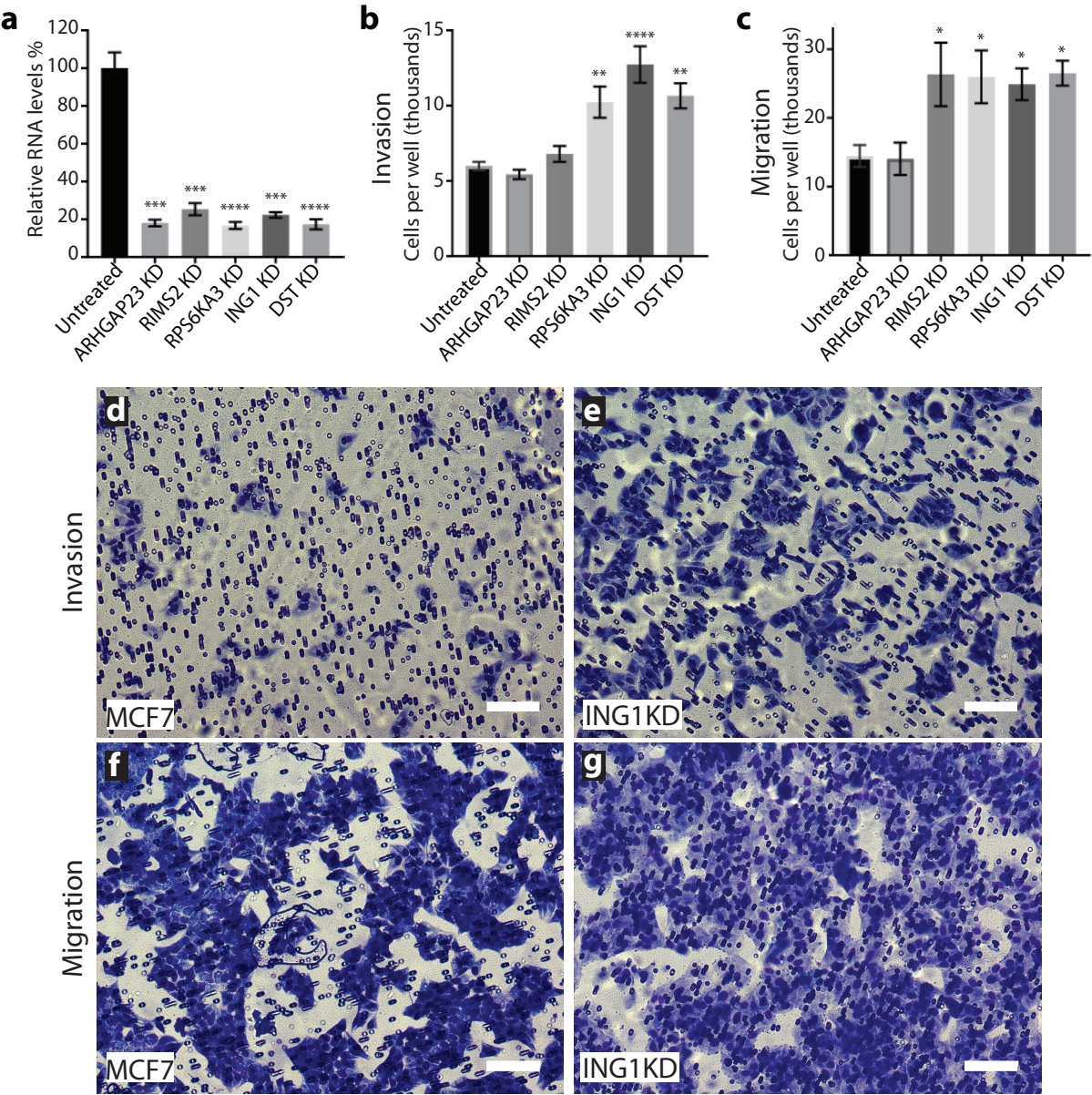


Figure S4: In vitro transwell migration and invasion assays for human orthologues of a selection of highly invasive fly genotypes. Related to Figure 5.

(a-c) Quantification of invasion (b) and migration (c), showing the number of MCF7 cells (a human breast adenocarcinoma cell line), untreated or transfected with siRNA targeting the indicated gene, which migrated through an 8µm pore membrane. For the invasion assay the membrane was coated with a layer of Matrigel (n=3 for each assay). Relative mRNA levels, determined by qRT-PCR, are shown in (a). Error bars represent $\pm$ s.e.m. One-way ANOVA with Dunnett's post hoc test for multiple comparisons was performed to determine statistical significance. **(d-g)** Representative images from the invasion and migration assays, showing untreated MCF7 cells and ING1KD cells. Staining intensity (purple) is directly proportional to the number of cells that have invaded through the ECM (d-e) or migrated through the 8µm pore membrane (f-g). Scale bar: 100µm.

Figure S5

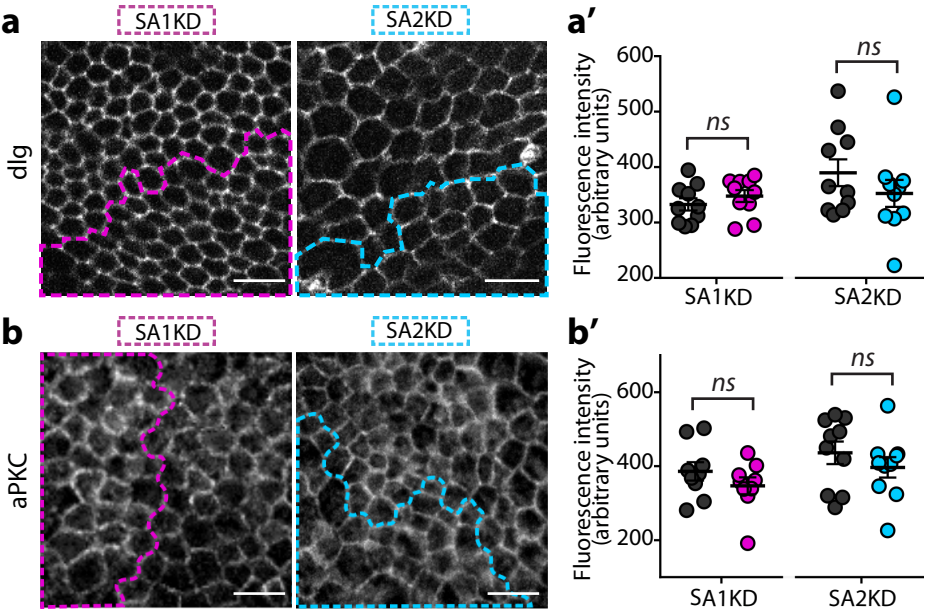


Figure S5: SA1 or SA2KD does not affect Dlg or aPKC localisation. Related to Figure 6.
(a-b) SA1 or SA2KD clones, highlighted by magenta and cyan dashed lines, respectively, show normal dlg (a) and aPKC (b) localisation. **(a'-b')** Quantification shows fluorescence intensity at the level of highest intensity (n=100 junctions from 10 animals for each genotype). Scale bars: 10µm. Error bars = $\pm$ s.e.m. Statistical analysis: Student's T test.

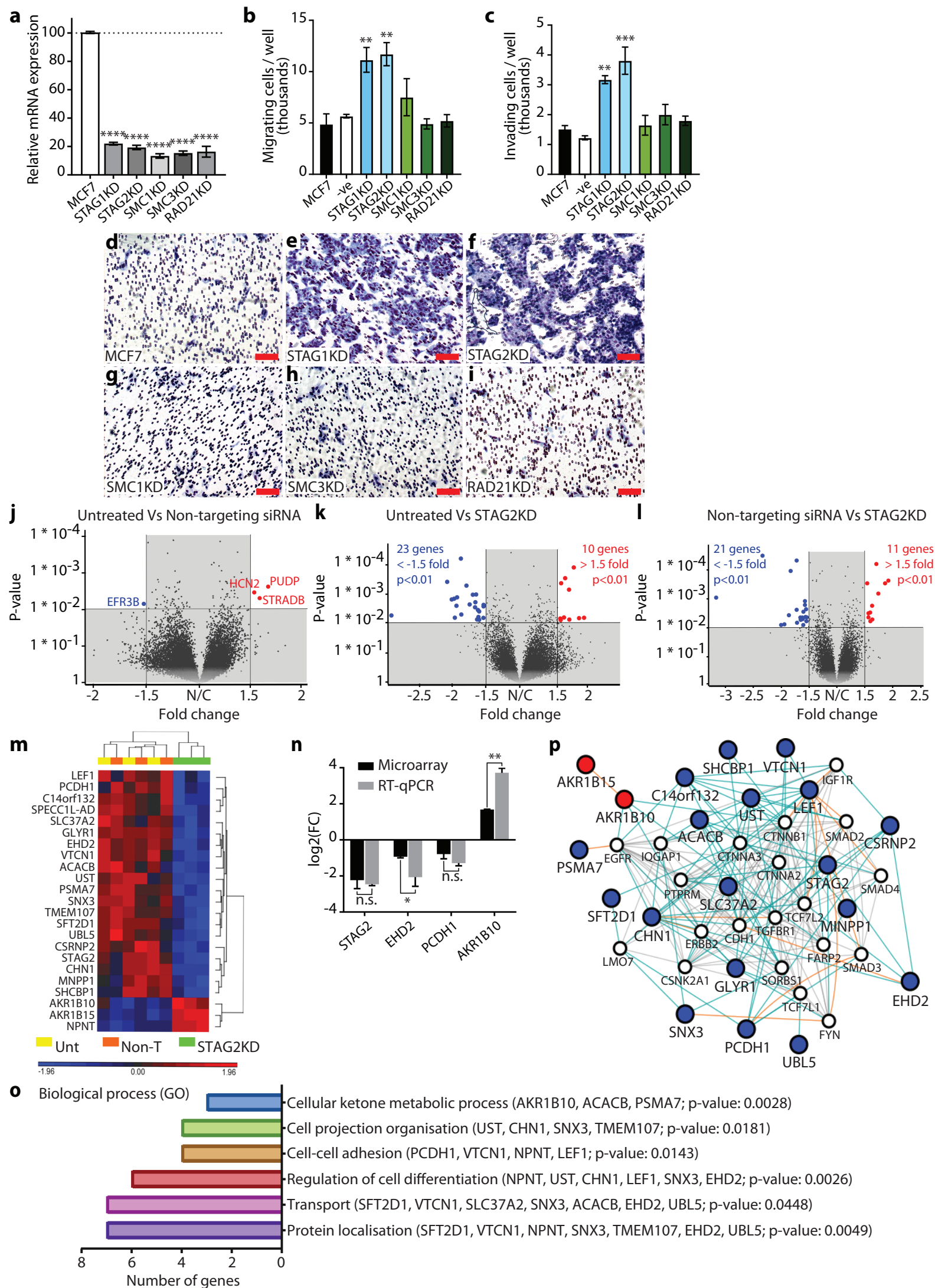
Figure S6

Figure S6: STAG1 and STAG2 are invasion suppressors in human cells. Related to Figure 6 and Table S5.

(a-i) In vitro transwell migration and invasion assays with the human breast adenocarcinoma cell line, MCF7, with or without the knockdown of human somatic cohesin complex subunits. **(a)** Relative mRNA levels following KD, determined by qRT-PCR. **(b-c)** Quantification of migration (b) and invasion (c) showing the number of MCF7 cells, untreated or transfected with siRNA targeting the indicated gene, which migrated through an 8µm pore membrane. For the invasion assay, the membrane was coated with a layer of Matrigel. n=3 for each assay. **(d-i)** Representative images of the invasion assay, showing dramatically increased invasion specifically when STAG subunits are knocked down. Staining intensity (purple) is directly proportional to the number of cells that have invaded through the ECM. Scale bar=100µm. **(j-l)** Volcano plot of significance level versus fold change in genetic expression from treated and untreated MCF7 cells. Each grey dot represents a gene that had no significant change in expression i.e. that had a p value greater than 0.01 and an expression fold change between -1.5 and 1.5. Each blue and red dot represents a gene that was downregulated or upregulated, respectively, compared to the control. Differentially expressed genes between STAG2KD cells and untreated MCF7 cells (unt) or MCF7 cells treated with non-targeting siRNA (nonT) are shown in (k) and (l) respectively; (j) is the negative control. To reduce the possibility of false positives, only genes that were differentially expressed in both STAG2KD against unt and nonT were selected. **(m)** Heat-map representation of unsupervised clustering of the 23 differentially expressed genes by STAG2KD in MCF7 cells. Each column represents a sample. Yellow: unt; orange: nonT; green: STAG2 siRNA treated. Colour code represents log₂ of the fold change of expression: blue, downregulated; red, upregulated. Horizontal and vertical clusters were created based in Euclidean distance. **(n)** Bar chart comparing microarray and RT-qPCR for four genes with differential expression. **(o)** Gene Ontology (GO) terms reaching Bonferroni corrected significance for enrichment amongst genes that had significantly altered their expression by STAG2KD. **(p)** Interaction network of differentially expressed genes by STAG2KD and adherens junction KEGG pathway proteins. The GeneMANIA plug-in for Cytoscape was used to generate an interaction network based on previously documented interactions. White nodes mark AJ components; red nodes: genes upregulated by STAG2KD; blue nodes: genes downregulated by STAG2KD. Orange lines: physical interactions; cyan lines: genetic interactions; grey lines: interactions between AJ genes. Error bars represent ± s.e.m. Student's T test or One-way ANOVA with Dunnett's post hoc test for multiple comparisons was performed to determine statistical significance.

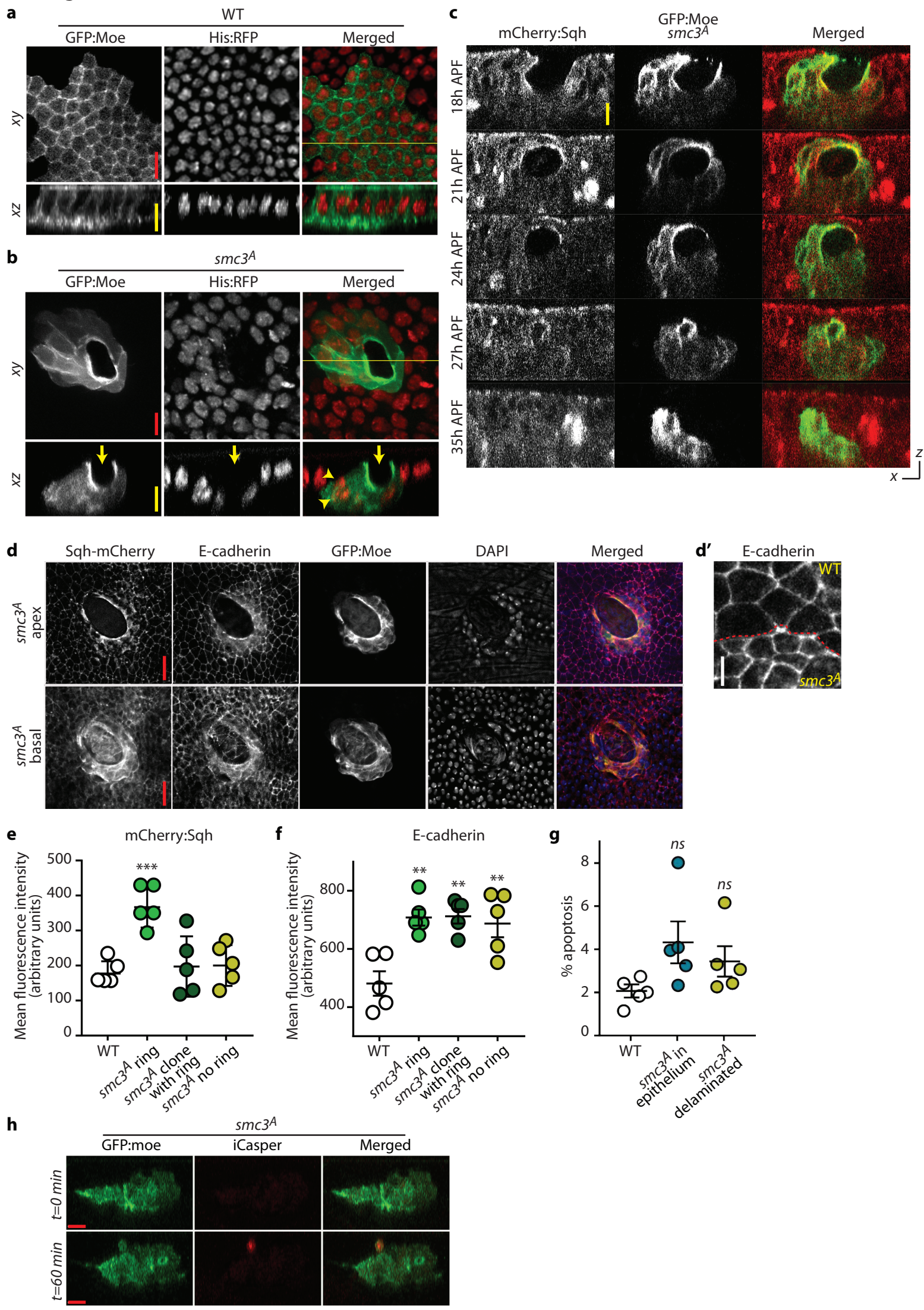
Figure S7

Figure S7: *smc3* LOF induces a supracellular actomyosin ring, which leads to invagination and clonal extrusion. Related to Figure 7.

(a-b) GFP:moe positively marked clones (green) and ubiquitously expressed Histone:RFP (red). Wild type tissue organisation is uniform; the orthogonal (xz) view shows a monolayer of cells (a). *smc3^A* clones frequently contain a supracellular actin-rich ring (b). No Histone:RFP signal is observed inside the ring. The xz view shows that the actin ring induces invagination (arrows); the nuclei of *smc3^A* cells are displaced basally (arrowheads). Yellow line shows position of xz slice shown. **(c)** Lateral view of a GFP:moe positively marked *smc3^A* clone (green) with additional mCherry:Sqh labelling (red). Animal was imaged live for approximately 15 hours. At 18h APF, the actin ring induces tissue invagination. At 21h APF, the actin ring closes, forming a central lumen. The mutant clone is extruded from the epithelium, and the actin ring contracts over time, closing the internal lumen (35h APF). **(d)** A dissected *Drosophila notum* containing a GFP:moe positively marked *smc3^A* clone with additional mCherry:Sqh labelling. The tissue is also stained for E-cadherin and DAPI. An apical confocal section shows no protein localisation or DAPI nuclear stain within the actin-rich ring. A more basal confocal section shows E-cadherin and DAPI staining, marking the adherens junctions and nuclear material of invaginated cells. **(d')** A higher magnification image of the E-cadherin stain, comparing *smc3^A* tissue with neighbouring WT tissue. **(e-f)** Quantification of mean fluorescence intensity of mCherry:Sqh (e) and E-Cadherin (f) in WT clones and *smc3^A* clones, with or without actin-rich rings. If a ring was present, measurements were taken at the ring, or at junctions outside the ring. 50 measurements from 5 animals were quantified, each dot represents one animal. Note the increased Sqh levels specifically at the ring, whereas cortical E-cadherin is significantly increased throughout the mutant clone. **(g)** Quantification of iCasper positive cells within WT clones, *smc3^A* clones that are still connected to the epithelial sheet, and extruded *smc3^A* clones. n=5 animals, all the labelled cells within each animal were quantified, each dot represents one animal. **(h)** Confocal images of a delaminated iCasper (red) and GFP:Moe (green) labelled *smc3^A* clone showing minimal apoptosis over 1h post extrusion. Red scale bars: 10µm; white scale bar: 5µm; yellow scale bars: 10µm in the xz plane. Error bars = ± s.e.m. Statistical analysis: Student's T test.

Figure S8

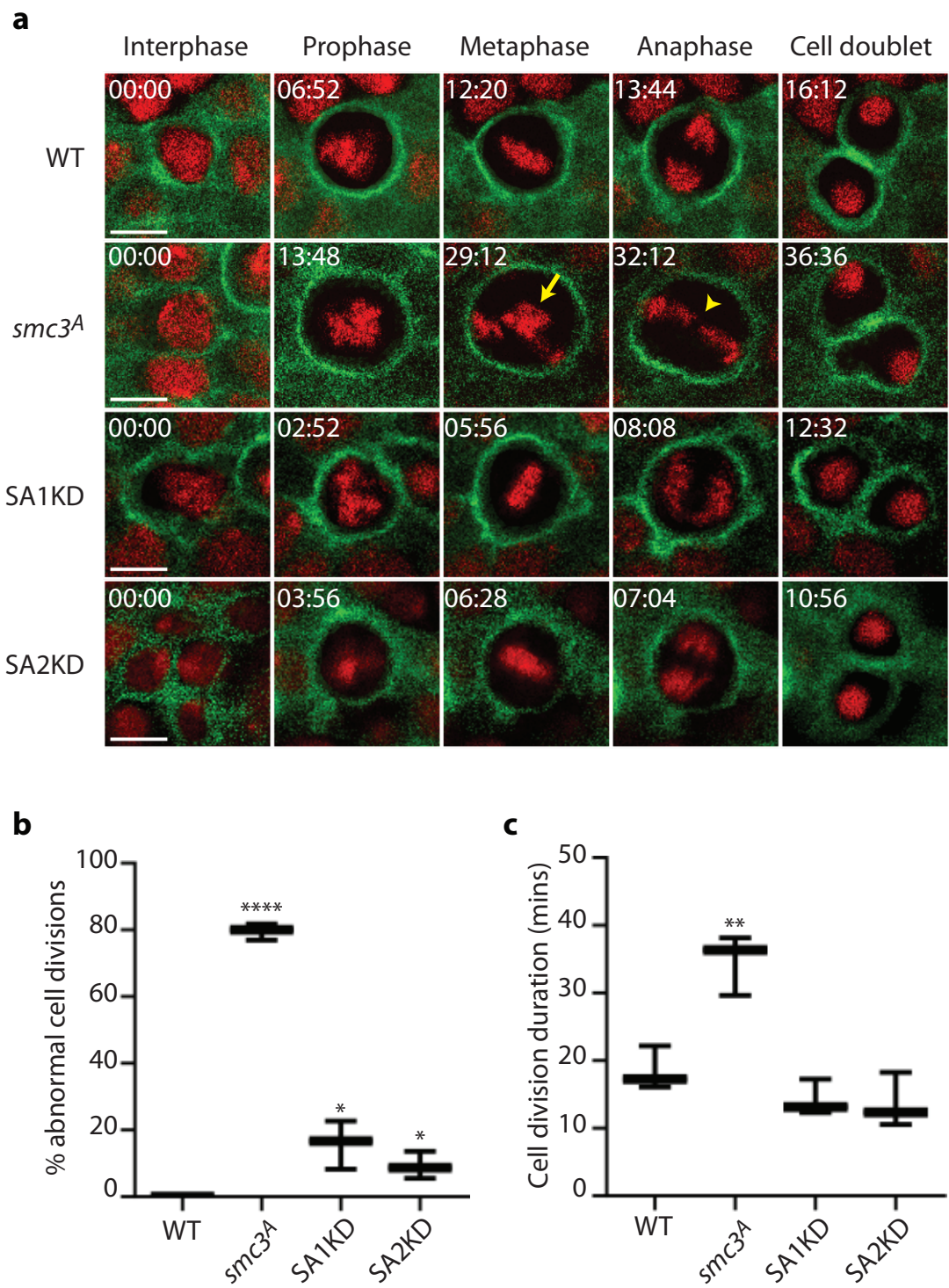


Figure S8: Live imaging shows frequent defective chromosome dynamics in *smc3^A* clones during cell division. Related to Movies S3-S6 and Figures 6-7.

(a) In vivo imaging of WT, *smc3^A*, SA1KD, and SA2KD clones, labelled with GFP:Moe and Histone:RFP, allowing live cell analysis of chromosomal dynamics during cell division. Scale bar: 5µm. Time stamp in top left corner of each panel. Arrow highlights defective chromosome alignment; arrowhead defective chromosome separation. **(b-c)** Quantification of percentage abnormal cell divisions (b) and cell division duration (c). n=40 cells from 4 animals for each genotype. Error bars = $\pm$ s.e.m. Statistical analysis: Student's T test.

Transparent Methods

Transgenic *Drosophila* stocks and crosses

Flystocks and crosses were raised on standard cornmeal and yeast medium at 25°C.

The following stocks were used: Ubx-FLP (#42718), *l(2)gl⁴* neoFRT40A /CyO (#36289), tubP-GAL80 neoFRT40A (#5192), Pnr-GAL4 (#25758), UAS-GMA (#31776), tubP-GAL80 neoFRT19A (#5132), UAS-p35 (#6298), *sna*[Sco]/CyO; PBac[y[+mDint2] w[+mC]=UAS-iCasper-noGFP-T2A-HO1]VK00005/TM6B (#64186), *sqh*-mCherry (#59024) were obtained from Bloomington *Drosophila* Stock Center (Indiana, USA).

UAS-RNAi stocks were obtained from VDRC Stock Center (Vienna, Austria) and NIG (Japan).

w¹¹¹⁸; P{GD12222}v27988 and *w¹¹¹⁸*; P{GD12222}v27989 were obtained from the *Drosophila* Genetic Resource Center (DGRC).

Ubx-FLP; FRT40A/Cyo-GFP, Pnr-GAL4, UAS-GFP:Moe/TM6b,

Ubx-FLP; *l(2)gl⁴*, neoFRT40A /Cyo-GFP; Pnr-GAL4, UAS-GFP:Moe/TM6b,

Ubi-His:RFP/CyO, and

w; tub-Gal80, FRT40A; MKRS/TM6b are lab stocks.

The following strains were recombined in the lab to create new strains: Pnr-GAL4 and UAS-GFP:Moe (chromosome III), Ubx-FLP and *Dlg⁵²* (chromosome X), Pnr-GAL4 and neoFRT82b *scrib¹* (chromosome III).

We designed an F2 RNAi screen as follows: tubP-GAL80 FRT40A; MKRS/TM6b Tb Hu virgin flies were crossed to UAS-RNAi lines on chromosome III; male and humoral F1 progeny of this cross of genotype tubP-GAL80 FRT40A /+; UAS-RNAi/TM6b Tb Hu were crossed to Ubx-FLP; *lg⁴* FRT40A /CyO, Kr-GFP; Pnr-GAL4, UAS-GFP:Moe/TM6b Tb Hu virgin females; male and female F2 progeny of genotype Ubx-FLP/+; *lg⁴* FRT40A / tubP-GAL80 FRT40A; Pnr-GAL4, UAS-GFP:Moe/UAS-RNAi were mounted for confocal imaging.

Dissections and live imaging

White pupae of the appropriate genotype were selected and aged at 29°C for 18 to 22 hours. Notum from aged pupae were dissected in PBS and the tissue fixed in 4% formaldehyde for 20 min at room temperature, before being blocked and permeabilised with PBS containing 0.2% BSA, 5% NGS, 0.1% Triton X-100.

For live imaging, aged animals with the appropriate genotype were prepared by cutting a window in the pupal case, attached to a slide with double-sided sticky tape. A coverslip with a drop of injection oil was then placed over the notum, supported by coverslips at either end to allow imaging on inverted confocal microscopes. Live imaging was performed with either a Leica SP2 inverted confocal microscope equipped with a × 40/1.25 NA oil objective with PL APO correction, a Zeiss LSM880 inverted confocal microscope equipped with a × 40/ 1.30 NA oil Ph3 M27 objective or a Zeiss LSM5 Exciter AxioObserver equipped with an EC Plan-Neofluar × 40/1.30 oil lens. Z-series were acquired using 1 μm z-sectioning. For time-lapse experiments z-series were acquired every 3 minutes for 2 hours.

Images from fixed samples were acquired on a Zeiss LSM880 inverted confocal microscope using a × 40/ 1.30 NA oil Ph3 M27 objective. Z-series were acquired using 0.5 μm z-sectioning.

Antibodies

We used the following primary antibodies at the indicated dilutions for this study: rat anti-E-Cad [1:100, DSHB (DCAD2)], mouse anti-Armadillo (1:100, DSHB), rat anti- α -Catenin [1:100, DSHB (DCAT1)], mouse anti-Fasciclin III [1:400, DSHB (7G10)], mouse anti-Discs Large [1:100, DSHB (4F3)], rabbit anti-PKC zeta [1:50, Santa Cruz (sc-216-G)], rabbit anti-pMad [1:250, Abcam (ab52903)]. Secondary antibodies from Molecular Probes were Alexa Fluor 488, 546 and 633.

Cell Culture

MCF-7 human breast carcinoma cells were obtained from Dr Anna Grabowska (Faculty of Medicine & Health Sciences, University of Nottingham). Cells were cultured in RPMI medium without phenol red (Invitrogen) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 20mM L-glutamine (Sigma) and grown in T75 culture flasks at 37°C in a 5% CO₂ atmosphere.

To knockdown gene expression, MCF-7 cells were incubated for 48 hours with transfection media containing DharmaFect1 (Dharmacon) transfection reagent and ON-TARGET plus SMARTpool siRNA (Dharmacon) against ArhGAP23, DST, RPS6KA3, RIMS2, ING1, STAG1, STAG2, SMC1, SMC3 and RAD21. ON-TARGETplus Non-targeting siRNA (Dharmacon) was used as a negative control and siGLO controls (Dharmacon) for transfection optimisation experiments.

After the transfection, cells with and without the knockdown were harvested and plated at appropriate density in hanging inserts with 8µm pore (Millipore). For invasion experiments, the inserts were covered with 75µg extracellular matrix (Corning). Cells were plated in 1% FBS RPMI and 10% FBS RPMI was used underneath the inserts. The amount of migrating or invading cells was assessed after a 48 hours incubation. Migrating or invading cells were detached from the bottom of the inserts using accutax cell detachment solution (Merk Millipore). The fluorescence intensity of the cell suspension was measured using 480/520 nm filters and optimal gain on a BMG fluorescent plate reader after a 15 minutes incubation with CyQuant GR dye (Thermo Fisher Scientific). Images of the migrating or invading cells were acquired using a 10x/ 0.3 NA objective on a Zeiss Axioplan/Axiophot brightfield microscope after inserts were stained using a Shandon Kwik-Diff stain kit (thermo fisher scientific) following the manufacturer's instructions. A background correction was performed using Gimp Software.

Total cellular RNA was extracted from un-transfected cells, cells featuring knockdown and cells transfected with ON-TARGETplus Non-targeting Control siRNAs (Dharmacon) using TRIzol (Thermo Fisher Scientific) following the manufacturer's instructions. The concentration and purity of RNA were determined by spectrophotometry. cDNA was synthesized using RevertAid First Strand cDNA Synthesis Kit [Thermo Fisher Scientific (k1621)] following the manufacturer's instructions. Quantitative real-time PCR (RT-PCR) was performed with A C1000 thermal cycler CFX96 RT system. Non-universal primers were designed using primer 3. Samples were labelled using IQ SYBR Green qPCR Master Mix (BioRad). Primer specificity and efficiency was confirmed by standard and melting curve analyses. Relative transcript levels were calculated relative to the transcript level of the reference gene GAPDH using the fold change method:

$$\Delta Cq = \text{mean } Cq (\text{target gene}) - \text{mean } Cq (\text{housekeeping gene})$$

$$E\Delta Cq = \frac{\text{power (PCR efficiency for the housekeeping gene, mean } Cq (\text{target gene}))}{\text{power (PCR efficiency for the target gene, mean } Cq (\text{housekeeping gene}))}, FC = \frac{E\Delta Cq}{\text{mean } E\Delta Cq}.$$

Primer list

Primer name	Sequence 5'	Sequence 3'
GAPDH human	ATGTTTCGTCATGGGTGTGAA	GTCTTCTGGGTGGCAGTGAT
ING1 human	TACTGTCTGTGCAACCAGGT	TTCTCCAGGGCTTTGTCCAT
ARHGAP23 human	CCGATGAATGGAGTGCCTT	ATGAAGTGGCGCAGAGTGAA
DST human	TACTGCCCTGGTCACTCTCAT	TACATCTGCTACCGGCCAC
RPS6KA3 human	ACCACATCTAGTAAAGGGTGCC	TTCACAGGGCTGTTGAGGTG
RIMS2 human	ACAGCAAGAACAGAAGGGTGA	CACATAACCTTGTTTGAGCGT
STAG1	GGCCAGCCGAAGTTAGAAGA	CCATTAGACCCCGAACAGCA
STAG2	AGACATGCCTGAGCAGATTGT	AAACGCAGCAAGTCCTCCTT
SMC1	ATCCACAGCTACAAGCCAGC	TTGCAAGTCGACTGCTCCTT
SMC3	TGGAGTTGGAATTAGGGTGTCA	GTCCACCTGAAAGCTGTTGC

RAD21	TGTATCAATGGGTGGGCCTG	GGCTCCAATGCAAATGCTTCT
PCDH1	GTCATCCAGGTGAAGGCCAA	CGGTCCTTAGCAAGCACTGA
EHD2	GGTGCGAGTTCACGCTTACA	AGATGACGGGCAGTTTGAGG
AKR1B10	CAGCAACAGAGAGCAGGACG	TGCCAAGAGGAGACTTCCAA

Interaction map

The interaction network was generated by integrating publicly available data to obtain a unified map of genes involved in cancer invasion. The data resources used were Droid (www.droidb.org, version 2018_08), Flybase (version FB2018_03) and BioGRID (version 3.5.165). PubMed ID was used as a unique identifier for each interaction, to prevent the inclusion of redundant information. Interactions were included into further consideration only if both interactors were hits in the categories of 'invasion', 'multilayering', 'cell body rounding', 'lethals' or if the interactor was not part of the screen yet linked an otherwise not interacting gene from any of these categories. The resulting network was drawn using Cytoscape 3.6 (http://www.cytoscape.org). MCODE v1.5.1 algorithm was used to identify subnetwork complexes of highly interconnected nodes.

Transcriptomics

Whole-genome transcriptome analysis of transfected and un-transfected (both un-treated and treated with non-targeting siRNA) MCF-7 cells was conducted at the Nottingham Arabidopsis Stock Centre (NASC). The RNA concentration and quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA) and the RNA 600 Nano Kit (Caliper Life Sciences, Mountain View, CA). Samples with a minimum RNA concentration of 100 ng/μl and RNA Integrity Number (RIN) ≥ 8 were used for gene expression analysis. Single stranded complementary DNA was prepared from 200 ng of total RNA as per the GeneChip™ WT PLUS Reagent Kit (Applied Biosystems and Affymetrix). Total RNA was first converted to cDNA, followed by *in vivo* transcription to make cRNA. Single stranded cDNA was synthesised, end labelled and hybridised for 16 h at 45°C to Clariom™ S Assay arrays (Thermo Fisher Scientific).

Gene expression data were analysed using Partek Genomics Suite 6.6 software (Partek Incorporated). The raw CEL files were normalised using the RMA background correction with quantile normalisation, log base 2 transformation and mean probe-set summarisation with adjustment for GC content. Differentially expressed genes (DEG) were identified by a two-way ANOVA. DEG were considered significant if p-value was ≤ 0.01 and fold change of >1.5 or <-1.5.

Regulatory network prediction was accomplished by use of GeneMANIA plug-in for Cytoscape. Based on this database, an analysis of genes interacting with adherens junction protein coding genes was carried out.

GO term enrichment

Genes were analysed for GO biological process enrichment, using the online annotation database: DAVID Bioinformatics Resources 6.8 (Huang et al., 2008). Functional annotation clustering with default settings was used; medium stringency and Benjamini–Hochberg correction was applied. The enriched GO terms with FDR <0.05 were selected and displayed in the interaction map.

Calculations and statistical analysis

Excel and XLstat were used to perform calculations, generate graphs and calculate statistical significance with Student's t-test, Mann-Whitney U test, Kruskal-Wallis test, and Spearman Correlation test where P>0.05 was considered not significant, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 and $r_s \in [0.5, 1.0]$ was considered to suggest a strong, $r_s \in [0.3, 0.5]$ a moderate, $r_s \in [0.1, 0.3]$ a weak and $r_s \in [0, 0.1]$ a very weak or non-existent correlation.

Supplemental References

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4, 44.