Uncovering the signaling landscape controlling breast cancer cell migration identifies novel metastasis driver genes

Esmee Koedoot [1,4], Michiel Fokkelman[1,4], Vasiliki-Maria Rogkoti[1,4], Marcel Smid[2], Iris van de Sandt[1], Hans de Bont[1], Chantal Pont[1], Janna E. Klip[1], Anna. M. Timmermans[2], Erik A.C. Wiemer[2], Peter Stoilov[3], John A. Foekens[2], Sylvia E. Le Dévédec[1], John W.M. Martens[2], Bob van de Water[1,5]

[1] Division of Drug Discovery and Safety, LACDR, Leiden University, the Netherlands

[2] Department of Medical Oncology and Cancer Genomics Netherlands, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands

[3] Department of Biochemistry and Cancer Institute, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, USA

[4] These authors contribution equally to the manuscript.

[5] To whom correspondence should be addressed

[5]Address of correspondence:

Bob van de Water
Division of Drug Discovery and Safety
Leiden Academic Centre for Drug Research
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands
Tel: +31-71-5276223
Fax: +31-71-5274277
b.water@lacdr.leidenuniv.nl

The authors have declared that no conflict of interest exists.

Abstract

TNBC is an aggressive and highly metastatic breast cancer subtype. Enhanced TNBC cell motility is a prerequisite of TNBC cell dissemination. Here we applied an imaging-based RNAi phenotypic cell migration screen using two highly motile TNBC cancer cell lines (Hs578T and MDA-MB-231) to provide a repository of signaling determinants that functionally drive TNBC cell motility. We screened ~4,200 target genes individually and discovered 133 and 113 migratory modulators of Hs578T and MDA-MB-231, respectively, which were linked to signaling networks predictive for breast cancer progression. The splicing factors PRPF4B and BUD31 and the transcription factor BPTF were essential for cancer cell migration, amplified in human primary breast tumors and associated with metastasis-free survival. Depletion of PRPF4B, BUD31 and BPTF caused primarily down-regulation of genes involved in focal adhesion and ECM-interaction pathways. PRPF4B was essential for TNBC metastasis formation *in vivo*, making PRPF4B a candidate for further drug development.

Introduction

Breast cancer is the most prevalent cancer in women. The triple-negative breast cancer (TNBC) subtype which lacks expression of the estrogen, progesterone and HER2 accounts for ~15% of breast cancer. TNBC is the most aggressive BC subtype with an overall 5 year relapse of 40% due to primary and secondary metastatic spread[1,2]. Transcriptomics has classified four different molecular subtypes of TNBC genes[3] while proteomics studies revealed markers of disease progression[4]. More recently, genome-wide sequencing of large numbers of human breast cancers have defined the somatically acquired genetic variation in breast cancer including TNBC[5–7] as well as the genetic evolutionary programs associated with local and distant metastatic spread[8,9]. Although the roles of individual genes in breast cancer metastasis[10–14] have been described, and few pooled shRNA screens have identified novel regulators of cancer metastasis[15,16] a systematic analysis of the functional consequences of genetic aberrations in TNBC for progression towards metastatic disease is still lacking.

The most critical cell biological hallmark of TNBC metastasis is the increased plasticity of TNBC cells[17–19]. Many TNBC cell lines have adapted a mesenchymal phenotype and demonstrate a high migratory behavior in association with an increased metastatic spread[12,20–25]. Cell migration is involved in various steps of the metastatic cascade, including local invasion, intravasation, extravasation, and colonization of secondary sites[26]. Understanding the fundamentals of TNBC cell migration is critical for our comprehension of the development of metastatic disease. The control of cell migration is complex and involves components of the cell adhesion machinery as well as modulators of the actin cytoskeleton that coordinate cellular motility behavior[11,27]. The functionality of these machineries is controlled by signaling pathways and associated transcriptional programs that coordinate the expression of these cell adhesion and cytoskeletal modulators as well as their post-translational modification and activity[28–30]. While the signaling pathways that control TNBC proliferation have been uncovered using genome-wide cell survival screens[31,32] the role of individual cell signaling components that define TNBC motility behavior is less clear.

Here we systematically unraveled the global cell signaling landscape that functionally control TNBC motility behavior through a phenotypic imaging-based RNAi-screen to identify genes involved in the regulation of different migratory phenotypes. We discovered novel genes including several transcriptional modulators, e.g. PRPF4B, BPTF and BUD31, that define TNBC migratory programs and metastasis formation, which are associated with poor clinical outcome of breast cancer and share signaling networks underlying prognostic gene signatures for primary breast cancer.

Results

A systematic high throughput signaling RNAi screen for TNBC cell migration

To unravel the critical signaling components that drive TNBC cell migration, we selected two of the most highly motile TNBC cell lines Hs578T and MDA-MB-231 for microscopy-based RNAi screening using the PhagoKinetic Track (PKT) assay (Fig. 1A, Suppl. Fig. 1, Supplementary movies 1 and 2; and described in detail previously[33]). Briefly, cells were seeded in fibronectin-coated 96-well plates containing a thin layer of discrete latex beads that are phagocytosed after cell attachment and migration, leaving behind migratory tracks. Image-analysis of the track characteristics including net area, major and minor axis, axial ratio and roughness allowed quantitative assessment of cell migration behavior. To unravel the signaling landscape that regulates mesenchymal tumor cell migration, we focused our screening effort on the complete set of cell signaling components, covering all kinases, phosphatases, (de)ubiquitinases, transcription factors, G-protein coupled receptors, epigenetic regulators and cell adhesion-related molecules (4198 individual target genes in total, SMARTpool siRNAs (pool of 4 single siRNAs per target)). Quantitative output data was normalized (robust Z-score) to mock transfected control cells. High and low Z-scores of individual parameters already showed the effect of siRNA knockdown on cell motility, i.e. low net area or low axial ratio suggests inhibition of cell migration whereas high axial ratio and high major axis indicated enhanced motility (Fig. 1C/D). Even though the quantification provided eight parameters, all the different migratory phenotypes observed within the data were not fully represented by single parameters. Therefore, migratory phenotypes were manually classified, and primary hits were identified by setting thresholds on Z-score for the most dominant parameters of each phenotype (described in the methods section). Candidate knockdown with decreased cell migration accompanied with an increase in cell size did not always result in a decreased net area, but were in this way selected using the round track phenotype. The migratory phenotypes were visualized by principal component analysis (PCA)-based clustering (Fig. 1E/F). For all phenotypes together, we defined in total 2807 hits in total: 1501 primary hits for Hs578T and

1306 for MDA-MB-231. Cytoskeletal genes, which are known to be important in cell migration and might provide regulatory feedback loops, were not enriched in these primary hit lists (Suppl. Fig. 2). Importantly, there was no correlation between proliferation (number of tracks) and any of the phenotypic parameters, suggesting that hit selection was mainly based on effects on migration (Suppl. Fig. 3). However, we cannot exclude that effects on cell proliferation might be contributing to the inhibition of cell migration for some of the candidates, in particular for those candidate genes that show a decrease in track number. We identified 129 overlapping hits showing similar effects on cell migration upon knockdown in both cell lines, suggesting these were bona-fide cell line-independent drivers of tumor cell migration. Hence we selected these overlapping hits for validation by single siRNA sequences. Additionally, to obtain a larger coverage of genes regulating cell migration that would uncover a more cell type specific migratory behavior, we also selected the top 153 hits in each cell line for validation. Only genes that have been defined as druggable were validated, resulting in validation of 451 unique targets (129 overlapping hits, 153 Hs578T unique hits and 153 MDA-MB-231 unique hits) (Suppl. Fig. 1 and Suppl. Table 1).

To validate the primary hits, we repeated the PKT screen assays with both SMARTpool and four single siRNA sequences (Fig. 2A and Suppl. Fig. 1). In total, 217 (77%) hits were validated in the Hs578T and 160 (57%) in the MDA-MB-231 (significant effects for at least 2 singles and SMARTpool, for Hs578T see Fig. 2B; for MDA-MB-231 see Suppl. Fig. 4; all validated genes are in Suppl. Table 2; reproducibility is shown in Suppl. Fig. 5 and Supp. Fig. 6). Mock (no siRNA addition) and siDNM2 were used as negative and positive control, respectively. The majority of validated hits was found in the phenotypic classes of reduced cell migration (Fig. 2C) and there was an overlap of 65 validated candidate genes that showed inhibition of cell migration in both cell lines (Fig. 2D). This relatively low overlap of candidates cannot be attributed to cell line specific mutations, copy number alterations or differences in expression levels of the candidates (Suppl. Fig. 7). Annotation of protein classes for each set of validated hits (Hs578T, MDA-MB-231, and overlap) showed that most of the hits were transcription factors (Figure 2E i) also after correction for library size (Figure 2E ii), suggesting that

transcriptional regulated gene networks are critical drivers of TNBC cell migration behavior and consequently metastasis formation.

Transcriptional determinants are critical drivers of BC migratory phenotypes.

To further confirm the migratory phenotype detected in the static PKT assay, we evaluated the effect of all validated hits on cell migration using a live microscopy cell migration assay with GFP-expressing Hs578T and MDA-MB-231 cells (Suppl. Table 3 and 4). Live cell imaging with Hs578T cells confirmed 133 of the 217 hits to inhibit cell migration. Similarly, for the MDA-MB-231 cells, 113 candidate genes (out of 160 validated hits) were confirmed to regulate cell migration. Upon knockdown, 31 PKT overlap candidates inhibited cell migration in this assay in both cell lines (Fig. 3A and 3B and Suppl. Movies 3-14), including various transcriptional and post-transcriptional regulators such as RUNX1, MTF1, PAX7, ZNF141, SOX14, MXD1, ZNF446, TARDBP, TBX5, BPTF, TCF12, TCERG1, ZDHHC13, BRF1, some of which are directly involved with splicing (BUD31 and PRPF4B) or histone modification (HDAC2 and HDAC10). For cell line specific validated hits, we filtered candidate genes for which the expression was associated with clinical breast cancer metastasis-free survival (MFS) in a patient dataset (the Public-344 cohort, GSE5237 and GSE2034, Suppl. Table 5). Many of the hits associated with poor outcome inhibited cell migration in both Hs578T and MDA-MB-231 (Fig. 3A and 3B, see Suppl. Table 3 and 4 for all candidate genes), combined with the overlap candidates resulting in 43 genes that were common denominators of cell migration. Single cell migratory trajectories were plotted for genes affecting cell migration in both cell lines (Fig. 3C). Taken together, our tertiary screen established high confidence in 61% and 71% of our candidate genes (Hs578T and MDA-MB-231 respectively). Furthermore, we confirmed the general role of our candidates in random cancer cell migration by validation of several main candidates by inducible CRISPR-Cas9 knockout in a live cell migration assay (Supp. Fig. 8), siRNA knockdown followed by live cell migration assays in 2 additional TNBC cell lines (Suppl. Fig. 9) and siRNA knockdown followed by a traditional scratch assay (Supp. Fig. 10). All of our

tested candidates were also affecting FBS-directed cell migration in MDA-MB-231, but not per se in Hs578T (Suppl. Fig. 10). This suggests that effects on random cell migration cannot always be directly extrapolated to directed cell migration, especially in the Hs578T cell line.

Functional drivers of tumor cell migration partner in networks predictive for BC progression.

To better understand the regulatory networks driving BC cell migration, we used the larger lists of our PKT validated candidate genes (217 for Hs578T and 160 for MDA-MB-231) to inform on protein-protein interaction (PPI) networks that are involved in Hs578T and MDA-MB-231 cell migration. KEGG pathway analysis was performed on the first-order networks of our candidate genes and revealed that similar pathways were affecting cell migration in both cell lines, despite that the networks were constructed from different candidate genes (Fig. 4A and 4B, Suppl. Table 6). We identified cancer-related pathways such as 'pathways in cancer' and 'focal adhesion' but also immune related pathways such as 'osteoclast differentiation' and 'chemokine signaling'. Interestingly, KEGG pathway analysis-based correction of the annotation of the PKT validated candidates not only confirmed the potent role of 'Transcriptional misregulation in cancer', but also immune-related and splicing pathways in cancer cell migration (Suppl. Fig. 11A). To further investigate the connection of our candidate genes to cell migration and invasion, we correlated our signaling networks of three established gene signatures associated with metastatic behavior and cell migration: the Human Invasion Signature (HIS)[34], the Lung Metastasis Signature (LMS)[35,36], and a 440-gene breast cancer cell migration signature (manuscript in preparation, Rogkoti et al.). Next, these three independent gene signatures were used to generate minimum interaction PPI networks, which only contained connecting nodes and seed proteins. Both Hs578T and MDA-MB-231 networks show a solid overlap with the 440-gene signature-derived network, with 156 and 145 genes overlapping (Hs578T and MDA-MB-231, respectively) (Suppl. Fig. 11B). This notion is further strengthened by the overlap of the Hs578T and MDA-MB-231 PPI networks with the LMS and

HIS signature-based networks: 58 (LMS) and 90 (HIS) genes in overlap with Hs578T network, and 53 (LMS) and 77 (HIS) genes with the MDA-MB-231 network. Furthermore, each gene signature derived network showed enrichment for the same KEGG pathways as the PPI networks based on our candidate genes (Suppl. Table 6). Given the high degree of overlap between these three gene signature-based networks and lists of candidate genes, we constructed a single zero-order network based on the combination of candidate genes affecting cell migration in Hs578T and MDA-MB-231 (65 genes, Fig. 4C). This revealed a sub-network linking 8 transcriptional regulators of which most already have been related to cancer progression, including HDAC2, BPTF, BRF1, TAF11, TCF12 and FOS[37–39], but also a prominent role for SMADs that are normally driven by TGF-β[40,41]. However, TGF-β treatment showed limited effects on TNBC cell migration (Suppl. Fig. 12), suggesting that migration is not dependent on TGF-β and SMADs were crucial for TNBC cell migration. Next we systematically investigated the effect of knockdown of the 217 PKT validated hits for morphological changes in the highly polarized Hs578T cell line by actin cytoskeleton staining, confocal imaging and quantitative single cell analysis (Suppl. Table 7). Hierarchical clustering grouped our PKT validated hits in nine different clusters (Fig. 4D, see Suppl. Fig. 13 for all gene names). Both clusters 2 and 9 contained control knockdown samples but also many genes that affected Hs578T cell migration, suggesting that a decrease in migration does not necessarily coincide with an overall change in cell morphology. Not surprisingly, inhibition of cell migration was associated with a wide variety of cellular morphologies: genes in cluster 1 and 3 decreased overall cell area, while genes in cluster 4, 5 and 6 increased overall cell area. Regarding the parameter cell spikes (reflecting the number of cell protrusions), we observed the same variation: some genes reduced cell protrusion formation (cluster 4, 5 and 8), while others enhanced protrusion formation (cluster 7). In vivo, loss of cell adhesion and increased motility are both prerequisites for metastasis formation. However, in vitro, our candidates could inhibit cell migration via different mechanisms: 1) increased cell-cell adhesions, 2) decreased cell-matrix interactions and 3) decreased actin turnover that can also result in multiple cell phenotypes. For example, down-regulation of cell-matrix interactions can decrease the cell

area, while decreased actin turnover might more affect the focal adhesions and spikes. In this way, our combined data suggests that cell shape and cellular motility are affected independently and indicates different genetic programs that define BC cell migration behavior.

Modulators of cell migration are associated with BC metastasis-free survival

To further relate our candidate genes to breast cancer progression and metastasis formation in patients, we compared our genes to three prognostic signatures for breast cancer metastasis[42–44]. Despite the minimal overlap of genes, these prognostic gene signatures have many related pathways in common[42] and minimum interaction PPI networks generated based on Wang's 76 genes, Yu's 50 genes, and the NKI-70 showed a robust overlap with our Hs578T and MDA-MB-231 cell migration networks based on PKT screen candidates (217 for Hs578T and 160 for MDA-MB-231) (Fig. 5A and Suppl. Table 8). These three gene expression signatures are strongly predictive of a short time to metastasis, implying but not yet statistically proven that our candidate genes are part of biologically functional regulatory networks and pathways critical in early onset of breast cancer metastasis.

Moreover, we investigated the percentage of mutations, amplifications and deletions (together % altered) of the 43 candidate genes that affected migration in both cell lines (see Fig. 3A and 3B) in 29 cancer types using publicly available data from The Cancer Genome Atlas (TCGA) (Fig. 5B). We identified clusters of candidate genes highly altered in multiple cancer types, amongst which breast cancer. This alteration rate was not related to tumor type aggressiveness (Supp. Fig. 14). For the main factors including BPTF, BUD31, CACNG1, RUNX1, GRK1, PTPRS, PRPF4B, and PBX1, most of these genetic alterations in breast cancer are dominated by amplifications (Fig. 5C), suggesting that enhanced expression levels of these candidates might be involved in breast cancer initiation or progression. Moreover, 11 of these candidates showed higher amplification rates (not significant) in the aggressive primary tumors that already metastasized at diagnosis (Suppl. Fig. 15) and 14 candidates

amongst which BUD31 and PRPF4B demonstrated a significant higher amplification rate in the TNBC compared to the ER positive subtype (Suppl. Fig. 16). Consequently, we also evaluated the association of the gene expression of these 43 candidate genes with MFS in the Public-344 cohort (Fig. 5D, Suppl. Table 9). Interestingly, high expression levels of both splicing factors PRPF4B and BUD31 are associated with earlier metastasis formation in triple-negative and ER-positive tumors respectively (Fig. 5D), but not in the other subtypes (Supp. Fig. 17). Non-core splicing factor PRPF4B is a serine/threonine-protein kinase regulating splicing by phosphorylation of other spliceosomal components[45]. BUD31 is a core splicing factor essential for spliceosome assembly, catalytic activity and associates with multiple spliceosome sub-complexes and has shown to be a MYC target in MYC-driven cancer cells[46]. We also identified the transcription factor BPTF, known for its role in chromatin remodeling and mammary stem cell renewal and differentiation[47,48], that is highly amplified in many cancer types and significantly positively correlated to MFS in breast cancer patients irrespective of the subtype (Fig. 5D, Suppl. Fig. 17). We further focused on splicing factors PRPF4B, BUD31 and transcription factor BPTF, since these were newly identified modulators of cell migration associated with BC MFS and/or highly amplified in BC.

PRPF4B, BUD31 and BPTF modulate expression and activity of cell-matrix adhesion components

In order to identify the mechanisms by which PRPF4B, BUD31 and BPTF were affecting BC cell migration, we performed knockdown of these candidate genes in MDA-MB-231 and Hs578T cells, followed by next generation sequencing (NGS)-based transcriptome analysis (3 days after siRNA transfection, 20.10^6 paired end reads per sample). For all three candidate genes, knockdown efficiency was >90% in Hs578T cells and >80% in MDA-MB-231 cells (Suppl. Fig. 18A, RT-qPCR validation Supp. Fig. 19). We identified differentially expressed genes (DEGs; log2FC <-1 or >1; adjusted p-value <0.05) for siPRPF4B, BUD31 and BPTF. Notably, expression levels of other validated screen candidates available in our RNA

sequencing dataset were not specifically affected by knockdown of PRPF4B, BUD31 or BPTF, indicating that these genes uniquely modulate transcriptional programs that affect cell migration (Suppl. Fig. 20). Knockdown of BUD31 had the broadest effect on gene expression and caused down-regulation of 1119 genes in Hs578T and 929 in MDA-MB-231, with ~50% affected genes overlapping between the two cell lines (Suppl. Fig. 18B-D). PRPFB4 and BPTF knockdown most significantly affected the transcriptome of the Hs578T cells, primarily resulting in down-regulation of gene expression. There was limited overlap in the DEGs between PRPFB4, BUD31 and BPTF (Suppl. Fig. 18E). Since PRPF4B and BUD31 are both splicing factors, we also investigated the effects of knockdown of these candidates on alternative splicing patterns (Suppl. Fig. 21). As expected, depletion of BUD31, a core spliceosomal protein, increased the intron retention compared to controls (Suppl. Fig. 21A and 7C)[46]. As might be expected from a non-core splicing factor, siPRPF4B only increased a small number of introns retained (Suppl. Fig. 21B and 21C). All tested intron retention events were validated with RT-PCR (Suppl. Fig. 22), indicating that the computational pipeline we used is reliable. We were unable to detect any changes in exon inclusion or 3' or 5' alterative splice site usage after knockdown of either PRPF4B or BUD31. Since siBUD31-induced intron retention was related to reduced expression of the specific gene of interest, we focused on the differentially expressed genes for further analysis. Although the general relation between intron retention and decreased gene expression was previously confirmed[49–51], future studies have to validate these direct causal relationships in response to PRPF4B and BUD31 knockdown. To identify pathways affected by siPRPFB4, siBUD31 and siBPTF we performed KEGG pathway over-representation analysis using the significantly down-regulated genes for all hits in both cell lines separately using ConsensusPathDB[52]. Although the overlap in DEGs between different cell lines and knockdown conditions was rather limited, the ECM-receptor interaction was over-represented in all conditions (Fig. 6A, Suppl. Fig. 23A). Moreover, knockdown of PRPF4B, BUD31 and BPTF resulted in down-regulation of the focal adhesion pathway in both cell lines, except for BPTF in Hs578T. Inhibition of these pathways likely directly contributes to the observed decrease in cellular migration. We also observed candidate specific responses such

as immune signaling for PRPF4B (Fig. 6A), cell adhesion for BPTF and metabolic and PI3K related pathways for BUD31 (Suppl. Fig. 23A). Next to down-regulated ECM-receptor interactions, deregulated TNF signaling was validated for all three knockdowns (Suppl. Fig. 24). Gene set enrichment analysis (GSEA)[53] confirmed the strong down-regulation of the ECM-receptor interaction pathway (Fig. 6B, Suppl. Fig. 23B). Clustering of all genes involved in ECM-receptor interaction (Fig. 6C, see Suppl. Fig. 25 for all gene names) or focal adhesion (Suppl. Fig. 26) demonstrated the involvement of many different pathway components of which some were overlapping between PRPFB4, BUD31 and BPTF (Fig. 6C and 6D). Many of these components were integrins, for which a prominent role has already been described in various steps in the metastatic cascade such as migration and invasion, anchorage-independent growth and colonization[54]. Moreover, increased integrin adhesions can activate actin contraction[55], regulate E-cadherin internalization and cell adhesion[56] and are related to EMT and cancer cell dissemination[57]. A similar down-regulation was observed at the protein level for several key components in both cell lines (Fig. 6E, Suppl. Fig. 27C). The effects on differential expression of cell matrix adhesion components such as integrins and focal adhesion kinase was also reflected in the different organization of focal adhesions and the F-actin network for both PRPFB4, BUD31 and BPTF (Fig. 6F and Suppl. Figure 27A-B). In summary, both splicing factors PRPF4B and BUD31 as well as the transcription factor BPTF modulate the expression of various focal adhesion-associated proteins and ECM-interaction signaling components in association with distinct cytoskeletal reorganization and decreased BC cell migration.

PRPF4B is essential for breast cancer metastasis formation *in vivo*

Finally, we investigated whether we could reproduce our *in vitro* findings in an *in vivo* mouse model for BC progression. Using our previously established orthotopic xenograft model, we predicted a decrease in BC metastasis formation upon splicing factor PRPF4B depletion. We selected PRPF4B because its depletion strongly inhibited both random and directed cell

migration in both Hs578T and MDA-MB-231 cells (see Fig. 3C and Suppl. Fig. 10) and, moreover, a role of PRPF4B in promoting TNBC metastasis formation has so far not been demonstrated. We established stable PRPF4B knockdown in the metastatic MDA-MB-417.5 cell line that expresses both GFP and luciferase[12,13,35] and contains similar basal PRPF4B levels as its parental MDA-MB-231 cell line (Supp. Fig. 28A-B). shPRPF4B MDA-MB-417.5 cells demonstrated ~40% PRPF4B knockdown at RNA as well as protein level (Fig. 7A-C) and similar as siRNA knockdown, decreased wound healing and intron retention of DGKZ and MAF1 (Supp. Fig. 28). shPRPF4B cells showed an equal primary tumor growth compared to the two shCtrl cell lines (Fig. 7D) which ensured identical time window for tumor cell dissemination from the primary tumor and outgrowth of macro-metastasis (Suppl. Fig. 29A-B). Interestingly, PRPF4B was higher expressed in the borders of the primary tumor (Supp. Fig. 30), supporting its potential role in invasion and metastasis formation. Bioluminescence imaging demonstrated that lung metastatic spread was less abundant in the PRPF4B knockdown group compared to control group (Suppl. Fig. 29C). Both bioluminescent imaging of the lungs *ex vivo* and counting of macro-metastases in the ink injected right lung revealed a significant decrease in metastasis formation in mice engrafted with shPRPF4B cells (Fig. 7F and 7G), which was also confirmed by a decreased lung weight (Suppl. Fig. 29D). *Ex vivo* bioluminescence imaging of the liver, spleen, heart, kidney, uterus and axillar lymph node also showed a decreased metastatic burden by shPRPF4B cells (Fig. 7E) confirmed by a decreased liver and spleen weight (Suppl. Fig. 29E and 29F). Altogether, this demonstrates that PRPF4B knockdown impairs general metastasis formation without showing organ-specificity.

Discussion

TNBC is the most aggressive form of breast cancer with 40% of patients dying from metastatic disease. New insights into TNBC migration is highly needed for the identification of potential new drug targets to modulate TNBC dissemination and possibly revert metastatic disease. In the present study, we applied a multi-parametric, high-content, imaging-based RNAi-screen covering ~4200 cell signaling components to unravel regulatory networks in TNBC cell migration and discovered 133 and 113 migratory modulators in the Hs578T and MDA-MB-231 cell lines, respectively. Splicing factors PRPF4B and BUD31 and transcription factor BPTF were critical for TNBC cell migration, associated with metastasis-free survival and affect genes involved in focal adhesion and ECM-interaction pathways. Moreover, PRPF4B knockdown reduced TNBC metastasis formation in vivo, making it an important target for future drug development.

Since our screening effort focused on a broad set of cell signaling components in two TNBC cell lines (Hs578T and MDA-MB-231), we were able to cover a high number of genes and networks in different migration modes. Indeed, the PKT assay allowed us to quantitatively assess different migration phenotypes, as the track morphology reveals the effect on migration, persistence and membrane activity. PCA analysis of the PKT data showed that three migratory phenotypes (small, small round, and round tracks) grouped closely together, resembling different types of inhibition of cell migration. Enhanced cell migration proved to be difficult to validate, probably due to the increased migratory phenotype in Hs578T and MDA-MB-231 reaching a physiological ceiling. Additional screening with TNBC cell lines with lower migratory potential would provide a platform to discover the spectrum of genes that may act as suppressors of TNBC cell migration and metastasis formation. The third live microscopy validation resulted in a conclusive selection of validated candidate tumor cell migration modulators: 133 and 113 genes in Hs578T and MDA-MB-231, respectively. The majority of these candidate genes displayed inhibition of cell migration and are most interesting for translation to cancer metastasis. Importantly, candidate migratory regulators, including SRPK1

and TRPM7, have previously been shown to impair cell migration and metastasis formation[12,14], supporting the robustness of our candidate drug target discovery strategy.

Our work provides a comprehensive resource detailing the role of individual signaling genes in cell migration. Previously, a cell migration screen in H1299 (non-small cell lung carcinoma) identified 30 candidate migration modulating genes[12]. Surprisingly, there was little overlap with our validated genes, with the exception of SRPK1. Similarly, little overlap in hits was found with a wound-healing screen in MCF10A cells[58]. These differences are likely due to the coverage and size of the screening libraries with the current screen covering ~4200 genes compared to ~1400 genes in the previously published data. Moreover, TNBC cell migration might be driven by different genetic program than non-small cell lung carcinoma and MCF10A cells. Also, the MCF10A screen focused on collective cell migration in epithelial cells, which is distinct from single cell mesenchymal migration in our two TNBC cell lines. Since ECM is a major component of the local tumor microenvironment, all our migration assays were performed on fibronectin coated plates. This coating significantly increased the migratory behavior of our cells (Supp. Fig. 31), which could also result in different candidates compared to the previous reported MCF10A screen. Moreover, none of the previously identified host-regulators of metastases in mouse[59] were validated in our screen. This might be due to the small overlap (only 7 of 23 regulators were in the library), but also suggests that candidates from this study are more involved in the first steps in the metastatic cascade such as escaping the primary tumor and intravasation rather than later steps such as extravasation, colonization and metastatic outgrowth.

While our two individual TNBC cell lines showed some overlap in genes that define cell migration, clearly also many distinct molecular determinants were defined for each cell line. This suggest that also the genetic differences between cancer cell lines lead to different molecular networks that can modulate the migratory behavior of tumor cells. For ultimate therapeutic strategies to modulate cancer dissemination the identification of common

denominators of migratory and metastatic behavior is essential. Our work contributes to the definition of some of these common players.

The importance of our candidate TNBC cell migration modulators was supported by comparative bioinformatics-based network analysis. Our comparative network analyses demonstrated that the cell migration PPI networks in Hs578T and MDA-MB-231 cells are profoundly similar to PPI networks derived from cell migration (Lung Metastasis Signature) [35], cell invasion (Human Invasion Signature)[34], as well as several BC prognostic signatures [42–44]. Moreover, we identified candidates that are highly amplified and/or mutated in many cancer types as well as candidates specifically related to breast cancer metastasis formation. Interestingly, two of these candidates, PRPF4B and BUD31, were splicing factors suggesting modulation of the expression of gene networks through alternative splicing. This is in line with some recent studies in which multiple splicing factors and events were related to cancer progression[60–62]. For example, cells undergo a complete switch in alternative splicing during EMT, catalyzed by the epithelial splicing factor ESRP[63]. Moreover, exon 11 skipping of key EMT component E-cadherin is downregulated during EMT and negatively correlated with its wild type variant[64] and the transmembrane glycoprotein CD44 important in cell migration and invasion has multiple splice variants related to invasion and metastasis formation[65]. In total, 43 splicing factors were represented in our primary screen of which 11 were selected for further validation in MDA-MB-231 (Supp. Fig. 32), again suggesting a prominent role for splicing regulation in breast cancer cell migration. Moreover, others have linked more splicing factors such as the hnRNPs, SRSF1, SRPK1 and PTBP1 to breast cancer migration and metastasis formation[12,66–69], making it very interesting to systematically evaluate the role of splicing in breast cancer cell migration in future studies.

The transcription factor BPTF was one of the major hubs in the interaction network of candidate genes, which is also highly amplified in primary tumors of BC patients and its expression levels are negatively related to MFS. Next generation sequencing revealed that knockdown of PRPF4B, BUD31 and BPTF all resulted in decreased expression of genes involved in focal

adhesions and ECM interactions, thus directly driving the observed more rounded and less polarized phenotype in conjunction with decreased cell migration.

Our list of candidate genes that regulate TNBC cell migration expectedly also contributes to cancer metastasis. We selected PRPF4B to assess whether our in vitro screening efforts can be translated to in vivo inhibition of metastasis formation. Depletion of PRPF4B strongly inhibited migration of both Hs578T and MDA-MB-231 cells in vitro. Also decreased levels of PRPF4B almost completely prevented spontaneous metastasis formation from the orthotopic primary tumor to distant organs, indicating that PRPF4B seems essential in metastasis formation. PRPF4B is a pre-mRNA splicing factor 4 kinase involved in the phosphorylation of PRP6 and PRP31 and splicing complex formation. Yet, in our hands depletion of PRPF4B showed little effect on intron retention patterns, suggesting more subtle splicing effects. However, PRPF4B knockdown did affect the expression of various components of the focal adhesion and ECM signaling pathways. Likely, this effect is an important contributor to the reduced migratory and metastatic behavior of TNBC cells. Hence, PRPF4B could be a relevant drug target to combat TNBC dissemination and future research should focus on the development of a specific PRPF4B inhibitor; the x-ray structure of the catalytic domain of PRPF4B suggest this is feasible[70]. For other potential candidates such as BPTF, the correlation with metastasis free survival added evidence that more candidate genes are likely involved in cancer metastasis. Since these associations do not prove a causal relationship, it would be very interesting to investigate these candidates in similar in vivo studies.

Our list of highly confident candidate migratory modulating genes provides ample opportunities for new hypotheses and additional studies in the field of cell migration. 16 G-protein coupled receptors were defined, which is particularly relevant as the pathway of GPCR-signaling is one of top over-represented pathways in ER-negative tumors[42]. Depletion of several ubiquitinases and proteasome components (DTX3L, UBE2E1, RNF31, RNF115, USP2, USP42, PSMC3, and PSMD10) were also found to inhibit tumor cell migration. Protein homeostasis and proteasome function have recently been suggested as a target for proliferation and growth in basal-like TNBC [71]. Transcription factors and regulators make up the largest group of candidate

metastasis genes. Some of these factors are part of a large interactive network including HDAC2, BPTF, BRF1, TAF11, TCF12 and FOS. The downstream targets of these transcription factors are potentially driving BC cell migration, and/or other biological processes that are critical for metastasis formation. Indeed BPTF knockdown affected focal adhesion and ECM components in a similar fashion as PRPF4B and BUD31 knockdown did.

In conclusion, in the present study we used imaging-based phenotypic screening to identify candidate metastatic genes for TNBC that have translational relevance. Understanding the gene networks that are controlled by the various candidate genes provides further insights in the biological programs that define BC cell migration behaviour and lead to important novel drug targets to combat cancer metastasis.

Methods

Transient siRNA-mediated gene knockdown

Human siRNA libraries were purchased in siGENOME format from Dharmacon (Dharmacon, Lafayette, CO, USA). Transient siRNA knockdown was achieved by reverse transfection of 50 nM single or SMARTpool (containing 4 different siRNAs) siRNA in a 96-well plate format using the transfection reagent INTERFERin (Polyplus, Illkirch, France) according to the manufacturer's guidelines. Medium was refreshed after 20 h and transfected cells were used for various assays between 65 to 72 h after transfection. Mock and/or siKinasePool were used as negative control. For the validation screen, 4 different single siRNAs and one SMARTpool siRNA mix (same four single siRNAs) were tested independently. For all other experiments, SMARTpool siRNAs were used.

Stable shRNA-mediated gene knockdown

MDA-MB-417.5 cells were transduced with lentiviral shRNA constructs coding for a non-targeting control sequences shCtrl #1 (SHC002), shCtrl #2 (SHC202V) or a sequence targeting the coding region of PRPF4B (target sequence: GCTTCACATGTTGCGGATAAT (TRCN0000426824)) (Mission/Sigma-Aldrich, Zwijndrecht, The Netherlands). The cells were selected by puromycin (sc-108071, Santa Cruz Biotechnology, Heidelberg, Germany). Knockdown efficiency was verified by RT-qPCR, Western Blot and immunofluorescent staining.

Phagokinetic track (PKT) assay

PKT assays were performed as described before[33,72]. For each transfection, duplicate bead plates were generated (technical replicates); transfection of each siRNA library was also performed in duplicate (independent biological replicate). Procedures for transfection, medium refreshment and PKT assay were optimized for laboratory automation by a liquid-handling robot (BioMek FX, Beckman Coulter).

PKT imaging and analysis

Migratory tracks were visualized by acquiring whole well montages (6x6 images) on a BD Pathway 855 BioImager (BD Biosciences, Franklin Lakes, NJ, USA) using transmitted light and a 10x objective (0.40 NA). A Twister II robotic microplate handler (Caliper Life Sciences, Hopkinton, MA, USA) was used for automated imaging of multiple plates. Montages were analyzed using WIS PhagoTracker[33]. Migratory tracks without cells or with more than 1 cell were excluded during image analysis. Quantitative output of PhagoTracker was further analyzed using KNIME. Wells with <10 accepted tracks were excluded. Next, data was normalized to mock to obtain a robust Z-score for each treatment and each parameter. After normalization, an average Z-score of the 4 replicates was calculated. Knockdowns with <3 images were removed, as well as knockdowns with <60 accepted tracks for Hs578T and <150 accepted tracks for MDA-MB-231. Phenotypic classes were determined based on Z scores of the track parameters: small (net area Z score < -4), small round (net area <8000, axial ratio < 1.7, net area Z score < -1, axial ratio Z-score < -3), big round (net area > 8000, axial ratio < 1.7, net area Z-score > 1, axial ratio Z-score < -4), long rough (axial ratio > 2.4, major axis > 200, roughness > 5, axial ratio Z-score > 1, major axis Z-score > 1) and long smooth (axial ratio > 2.1, major axis > 180, roughness < 5).

All PKT screening data will be available in the Image Data Resource database upon publication and is currently available upon request.

Live single cell migration assay and analysis

Hs578T-GFP and MDA-MB-231-GFP cells were transfected with siRNAs as described above and after 65 h, knockdown cell suspensions were seeded in fibronectin-coated black 96-well glass plates (SensoPlate, Greiner Bio-One, Frickenhausen, Germany). As controls, siRNA targeting the GTPase dynamin 2 (DNM2) was used for reduced cell migration and mock was used as transfection control. For the TGF-β and EGF stimulation experiments, cells were

seeded in fibronectin-coated black 96-well glass plates (SensoPlate, Greiner Bio-One, Frickenhausen, Germany). Cells were starved in serum-deprived RPMI for 2 hours after which cells were stimulated with 10 ng/ml TGF-β (Immunotools) or 100 ng/ml EGF (Peprotech). Imaging was directly started after treatment. Live microscopy was performed on a Nikon Eclipse Ti microscope using a 20x objective (0.75 NA, 1.00 WD), 2x2 binning and 2x2 montageTwo positions per well were selected and GFP images were acquired every 12 min for a total imaging period of 12 h using NIS software (Nikon, Amsterdam, The Netherlands). Image analysis was performed using CellProfiler (Broad Institute)[73]. For live cell migration, images were segmented using an in-house developed watershed masked clustering algorithm [74], after which cells were tracked based on overlap between frames. Tracking data was organized and analyzed using in-house developed R-scripts to obtain single cell migration data. Only data originating from cells that were tracked for a minimum of 2 h was used. Two negative control wells with low and high cell densities, comparable to the knockdown populations, were selected for statistical comparison, and knockdowns were required to be statistically significant compared to both controls.

Imaging-based phenotypic screen

Hs578T cells were fixed and permeabilized in 1% formaldehyde and 0.1% Triton X-100 in PBS and blocked in 0.5% bovine serum albumin (BSA, A6003, Sigma Aldrich) in PBS. Cells were stained with Rhodamine Phalloidin (R415, Molecular Probes) and imaged using an Nikon Eclipse TE2000-E inverted confocal microscope (Nikon Instruments, Amsterdam, The Netherlands) using a 20x Plan Apo objective, 408 and 561 nm lasers.2x digital zoom, 2x2 stitching images were captured at 4 positions per well. Nuclei and actin cell body were detected by CellProfiler (Broad Institute) and parameters were measured using the measure object size and shape module in CellProfiler. Images with more than 150 cells were filtered out. Using KNIME, nuclei without a clear cell body were rejected and single cell data was normalized to the median of 2 mock control wells per plate. For the heatmap, all features were mock normalized and clustering was performed on complete linkage and Euclidean distance.

Next generation sequencing

RNA was collected 72 hours after knockdown using the RNeasy plus mini kit (Qiagen) according to the manufacturer's guidelines. DNA libraries were prepared with the TruSeq Stranded mRNA Library Prep Kit and sequenced according to the Illumina TruSeq v3 protocol on an Illumina HiSeq2500 sequencer. 20.10^6 100 base pair paired-end reads were generated and alignment was performed using the HiSat2 aligner (version 2.2.0.4) against the human GRCh38 reference genome. Gene expression was quantified using the HTseq-count software (version 0.6.1) based on the ENSEMBL gene annotation for GRCH38 (release 84). Count data was normalized and log2 fold changes and adjusted P-values were calculated using the DESeq2 package[75]. Calculated log2 fold changes were used to perform ranked GSEA [53]. DEGs were selected by effect size (log2 fold change bigger than 1 or smaller than -1) and adjusted p-value (smaller than 0.05) and used for over-representation analysis for KEGG pathways using ConsensusPathDB[52].

For the intron retention analysis, RNA-seq reads were mapped to the current human genome (GRCh38) using Hisat 2[76]. Differential intron retention analysis was carried out in R using DexSeq package[77,78]. In DexSeq the difference of intron inclusion were determined based the counts from the intron and the counts from the two adjacent exons. The sizes of the exons were limited to 100nt immediately adjacent to the intron to reduce artifacts deriving from alternative promoters, alternative splice sites and alternative poly-adenylation sites. Deferentially retained introns were selected by effect size (relative change in inclusion more than 0.1) and statistical significance of the change (adjusted p-value less than 0.001). Alternative exon inclusion was analyzed using the rMATS package version 3.0.8[79]. Differentially spliced exons were selected by effect size (relative change in inclusion more than 0.1) and statistical significance of the change (adjusted p-value less than 0.001).

RNA sequencing data is available in Sequence Read Archive with accession number SRP127785.

Network analysis

Protein annotation of the primary hits was retrieved from QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, USA). Protein-protein interaction (PPI) networks were generated separately for all signatures using NetworkAnalyst (www.networkanalyst.ca)[80]. Candidate genes were used as seed proteins to construct first-order, minimum interaction and zero-order networks based on the InnateDB Interactome. KEGG pathway analysis was performed on the first-order PPI networks. The connection between multiple PPI networks was visualized by a Chord diagram using NetworkAnalyst.

Orthotopic mouse model for metastasis assessment

1 x 10^6 MDA-MB-417.5 shCtrl #1, shCtrl #2 or shPRPF4B cells diluted in 100 µL matrigel (9.2 mg/ml, 354230, batch 4321005, Corning, Amsterdam, The Netherlands) were injected in the fourth mammary fat pad of 7 to 9 week old female Rag2$^{-/-}$ Il2rg$^{-/-}$ mice (n=8 per group: coefficient of variation = 0.2, effect size = 30%, power π = 0.90 bij α = 0.05). Primary breast tumors were surgically removed when they reached the size of 7×7 mm. Next, bioluminescent imaging was used to follow metastasis formation over time. Mice were sacrificed 50 or 51 days after surgery and metastasis formation of all organs was assessed by bioluminescent imaging followed by weighing the lungs, liver and spleen. The right lung was injected with ink in order to count the number of lung macrometastases. Animals that passed away due to unknown reasons before the end of the study were removed from the analysis. Animals were randomly distributed over the different groups. The experiment was performed without blinding.

Breast cancer patient gene expression profiles

Gene expression data of a cohort of 344 lymph node-negative BC patients (221 estrogen receptor-positive (ER-positive) and 123 estrogen receptor-negative (ER-negative)), who had not received any adjuvant systemic treatment, was used and is available from the Gene

Expression Omnibus (accession no. GSE5327 and GSE2034). Clinical characteristics, treatment details and analysis were previously described[42,43,81–83]. Stata (StataCorp) was used to perform Cox proportional hazards regression analysis, with gene expression values as continuous variable and MFS as end point.

Statistical Analysis

Sample sizes were based on previously published similar experiments. When not indicated, all experiments were performed in biological triplicates. Normality of migration measurements and *in vivo* data was tested using Kolmogorov–Smirnov's test, d'Agostino and Pearson's test and Shapiro–Wilk's test using GraphPad Prism 6.0 (GraphPad Software, San Diego, CA). A data set was considered normal if found as normal by all three tests. Data sets following a normal distribution were compared with Student's t-test (two-tailed, equal variances) or one-way ANOVA (for comparison of more than 2 groups) using GraphPad Prism 6.0. Data sets that did not follow a normal distribution were compared using Mann–Whitney's test or a non-parametric ANOVA (Kruskal–Wallis with Dunn's multiple comparisons post-test) using GraphPad Prism 6.0. Results were considered to be significant if p-value < 0.05.

Study Approval

The study involving human BC patients was approved by the Medical Ethical Committee of the Erasmus Medical Center Rotterdam (Netherlands) (MEC 02.953). This retrospective study was conducted in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http:/www.federa.org/). Mouse experiments and housing were performed according to the Dutch guidelines for the care and use of laboratory animals (UL-DEC-11244).

Data availability

All datasets used to generate the results presented in this study are publicly available. RNA sequencing data is available in Sequence Read Archive with accession number SRP127785. All PKT screening data will be available in the Image Data Resource database upon publication

and is currently available upon request. All other remaining data are available within the article or supplemental data or is available upon request.

Author contributions

MF, EK, VMR, SELD and BvdW conceived and designed the experiments. MF, EK, VMR, SELD, IvdS, CP and JK performed the experiments. EK and MF analyzed the data. EACW, MS, JAF and JWMM provided the clinical data. PS performed the intron retention analysis. HdB provided technical support for the microscopes. EK and MF wrote the manuscript. BvdW, SELD, JAF and JWMM reviewed and corrected the manuscript.

Acknowledgements

This project was supported by the EU FP7 Systems Microscopy Network of Excellence (grant no. 258068), the ERC Advanced grant Triple-BC (grant no. 322737) and the Dutch Cancer Society project (grant nr 2011-5124).

References

1. Badve, S. *et al.* Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. *Mod. Pathol.* **24,** 157–167 (2011).

2. Fulford, L. G. *et al.* Basal-like grade III invasive ductal carcinoma of the breast: Patterns of metastasis and long-term survival. *Breast Cancer Res.* **9,** 1–11 (2007).

3. Lehmann, B. D. *et al.* Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One* **11,** e0157368 (2016).

4. Liu, N. Q. *et al.* Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* **17,** 155–164 (2012).

5. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534,** 47–54 (2016).

6. Morganella, S. *et al.* The topography of mutational processes in breast cancer genomes. *Nat. Commun.* **7,** 1–11 (2016).

7. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–4 (2012).

8. Yates, L. R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32,** 169–184.e7 (2017).

9. Yates, L. R. & Desmedt, C. Translational genomics: Practical applications of the genomic revolution in breast cancer. *Clin. Cancer Res.* **23,** 2630–2639 (2017).

10. Liu, N. Q. *et al.* Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *J. Natl. Cancer Inst.* **106,** (2014).

11. Sieg, D. J. *et al.* FAK integrates growth-factor and integrin signals to promote cell migration. *Nat. Cell Biol.* **2,** (2000).

12. Roosmalen, W. Van *et al.* Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *J. Clin. Invest.* **125,** 1648–1664 (2015).

13. de Graauw, M. *et al.* Annexin A2 depletion delays EGFR endocytic trafficking via cofilin activation and enhances EGFR signaling and metastasis formation. *Oncogene* **33,** 2610–2619 (2014).

14. Middelbeek, J. *et al.* TRPM7 is required for breast tumor cell metastasis. *Cancer Res.* **72,** 4250–61 (2012).

15. Knott, S. *et al.* Asparagine bioavailability governs metastasis in a model of breast cancer. *Nature* **554,** 378–381 (2018).

16. Wagenblast, E. *et al.* A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520,** 358–362 (2015).

17. Friedl, P., Sahai, E., Weiss, S. & Yamada, K. M. New dimensions in cell migration. *Nat. Rev. Mol. Cell Biol.* **13,** 743–747 (2012).

18. Petrie, R. J. & Yamada, K. M. At the leading edge of three-dimensional cell migration.

J. Cell Sci. **125,** 5917–5926

19. Liu, Y. J. *et al.* Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells. *Cell* **160,** 659–672 (2015).

20. Anders, C. K. & Carey, L. A. Biology, Metastatic Patterns, and Treatment of Patients with Triple-Negative Breast Cancer. *Clin. Breast Cancer* **9,** S73-81 (2009).

21. Perou, C. M. Molecular Stratification of Triple-Negative Breast Cancers. *Oncologist* **15,** 39–48 (2010).

22. Harney, A. S. *et al.* Real-time imaging reveals local, transient vascular permeability and tumor cell intravasation stimulated by Tie2hi macrophage-derived VEGFA. **5,** 932–943 (2015).

23. Beerling, E. *et al.* Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep.* **14,** 2281–2288 (2016).

24. Kedrin, D. *et al.* Intravital imaging of metastatic behavior through a mammary imaging window. *Nat. Methods* **5,** 1019–1021 (2008).

25. Riaz, M. *et al.* MiRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast Cancer Res.* **15,** (2013).

26. Chaffer, C. L. & Weinberg, R. A. A Perspective on Cancer Cell Metastasis. *Science (80-.).* **331,** 1559–1564 (2011).

27. Sahai, E., Garcia-Medina, R., Pouysségur, J. & Vial, E. Smurf1 regulates tumor cell plasticity and motility through degradation of RhoA leading to localized inhibition of contractility. *J. Cell Biol.* **176,** 35–42 (2007).

28. Nieto, M. A., Huang, R. Y. Y. J., Jackson, R. A. A. & Thiery, J. P. P. EMT: 2016. *Cell* **166,** 21–45 (2016).

29. Wolf, K. *et al.* Compensation mechanism in tumor cell migration: Mesenchymal-amoeboid transition after blocking of pericellular proteolysis. *J. Cell Biol.* **160,** 267–277 (2003).

30. Ye, X. & Weinberg, R. A. Epithelial-Mesenchymal Plasticity: A central regulator of cancer progression. doi:10.1016/j.tcb.2015.07.012

31. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170,** 577–592.e10 (2017).

32. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170,** 564–576.e16 (2017).

33. Fokkelman, M. *et al.* PhagoKinetic Track Assay: Imaging and Analysis of Single Cell Migration. *Bio-protocol* **6,** (2016).

34. Patsialou, A. *et al.* Selective gene-expression profiling of migratory tumor cells in vivo predicts clinical outcome in breast cancer patients. *Breast Cancer Res.* **14,** R139 (2012).

35. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436,** 518–524 (2005).

36. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic

spread. *PNAS* **104,** 6740–6745 (2007).

37. Bland, K. I., Konstadoulakis, M. M., Vezeridis, M. P. & Wanebo, H. J. Oncogene Protein Co-Expression Prognostic Discriminants for Breast Carcinoma. *Ann. Surg. J. B. Lippincott Co.* **221,** 706–720 (1995).

38. Müller, B. M. *et al.* Differential expression of histone deacetylases HDAC1, 2 and 3 in human breast cancer - overexpression of HDAC2 and HDAC3 is associated with clinicopathological indicators of disease progression. *BMC Cancer* **13,** 1–8 (2013).

39. Lee, C. C. *et al.* TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. *J. Biol. Chem.* **287,** 2798–2809 (2012).

40. Zi, Z., Chapnick, D. A. & Liu, X. Dynamics of TGF-β/Smad Signaling. *FEBS Lett.* **586,** 1921–1928 (2012).

41. Moustakas, A., Souchelnytskyi, S. & Heldin, C. Smad regulation in TGF- β signal transduction. *J. Cell Sci.* **114,** 4359–4369 (2001).

42. Yu, J. X. *et al.* Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer* **7,** 182 (2007).

43. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365,** 671–79 (2005).

44. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).

45. Schneider, M. *et al.* Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat. Struct. Mol. Biol.* **17,** 216–222 (2010).

46. Hsu, T. Y.-T. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525,** 384–388 (2015).

47. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).

48. Frey, W. D. *et al.* BPTF maintains chromatin accessibility and the self-renewal capacity of mammary gland stem cells. *Stem Cell Reports* **9,** 23–31 (2017).

49. Wong, J. J. *et al.* Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell* **154,** 583–595 (2013).

50. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24,** 1774–1786 (2014).

51. Bergeron, D., Pal, G., Beaulieu, Y. B. & Chabot, B. Regulated Intron Retention and Nuclear Pre-mRNA Decay Contribute to PABPN1 Autoregulation. *Mol. Cell. Biol.* **35,** 2503–2517 (2015).

52. Kamburov, A., Wierling, C., Lehrach, H. & Herwig, R. ConsensusPathDB - A database for integrating human functional interaction networks. *Nucleic Acids Res.* **37,** D623–D628 (2009).

53. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102,** 15545–15550 (2005).

54. Hamidi, H. & Ivaska, J. Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer* **18,** 533–548 (2018).

55. Martinez-rico, C., Pincet, F., Thiery, J. & Dufour, S. Integrins stimulate E-cadherin-mediated intercellular adhesion by regulating Src-kinase activation and actomyosin contractility. *J. Cell Sci.* **123,** 712–722 (2010).

56. Canel, M. *et al.* Quantitative in vivo imaging of the effects of inhibiting integrin signalling via Src and FAK on cancer cell movement ; effects on E-cadherin dynamics. *Cancer Res.* **70,** 9413–9422 (2010).

57. Yilmaz, M. & Christofori, G. EMT , the cytoskeleton , and cancer cell invasion. *Cancer metastasis rev* **28,** 15–33 (2009).

58. Simpson, K. J. *et al.* Identification of genes that regulate epithelial cell migration using an siRNA screening approach. *Nat. Cell Biol.* **10,** 1027–1038 (2008).

59. van der Weyden, L. *et al.* Genome-wide in vivo screen identifies novel host regulators of metastatic colonization. *Nature* **541,** 233–236 (2017).

60. Marabti, E. El & Younis, I. The Cancer Spliceome : Reprograming of Alternative Splicing in Cancer. *Front. Mol. Biosci.* **5,** 1–11 (2018).

61. Oltean, S. & Bates, D. O. Hallmarks of alternative splicing in cancer. *Oncogene* **33,** 5311–5318 (2014).

62. Marzese, D. M., Manughian, A. O., Javier, P. & Dave, I. J. O. Alternative splicing and cancer metastasis : prognostic and therapeutic applications. *Clin. Exp. Metastasis* **35,** 393–402 (2018).

63. Warzecha, C. C. *et al.* An ESRP-regulated splicing programme is abrogated during the epithelial – mesenchymal transition. *EMBO J.* **29,** 3286–3300 (2010).

64. Sharma, S., Liao, W., Zhou, X., Wong, D. & Lichtenstein, A. Exon 11 skipping of E-cadherin RNA downregulates its expression in Head and Neck cancer cells. *Mol. Cancer* **10,** 1751–1759 (2011).

65. Brown, R. L., Reinke, L. M., Damerow, M. S. & Perez, D. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J. Clin. Invest.* **121,** 1064–1074 (2011).

66. Pelisch, F. *et al.* Involvement of hnRNP A1 in the matrix metalloprotease-3-dependent regulation of Rac1 pre-mRNA splicing. *J Cell Biochem* **113,** 2319–2329 (2012).

67. He, X. *et al.* Involvement of polypyrimidine tract-binding protein (PTBP1) in maintaining breast cancer cell growth and malignant properties. *Oncogenesis* **3,** e84 (2014).

68. Shimoni-Sebag, A., Lebenthal-Loinger, I., Zender, L. & Karni, R. RRM1 domain of the splicing oncoprotein SRSF1 is required for MEK1-MAPK-ERK activation and cellular transformation. *Carcinogenesis* **34,** 2498–2504 (2013).

69. Anczuków, O. *et al.* The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol.* **19,** 220–228 (2012).

70. Gao, Q. *et al.* Evaluation of cancer dependence and druggability of PRP4 kinase using cellular, biochemical, and structural approaches. *J. Biol. Chem.* **288,** 30125–30138

(2013).

71. Petrocca, F. *et al.* A Genome-wide siRNA Screen Identifies Proteasome Addiction as a Vulnerability of Basal-like Triple-Negative Breast Cancer Cells. *Cancer Cell* (2013). doi:10.1016/j.ccr.2013.07.008

72. Naffar-Abu-Amara, S. *et al.* Identification of novel pro-migratory, cancer-associated genes using quantitative, microscopy-based screening. *PLoS One* **3,** 1–9 (2008).

73. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).

74. Yan, K., Verbeek, J. & Verbeek, F. in *ISoLA 2012, PARTII, LNCS 7610* 25–41 (2012). doi:10.1007/978-3-319-47166-2_67

75. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

76. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).

77. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22,** 2008–2017 (2012).

78. Reyes, A. *et al.* Drift and conservation of differential exon usage across tissues in primate species. *Proc. Natl. Acad. Sci.* **110,** 15377–15382 (2013).

79. Shen, S. *et al.* rMATS : Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *PNAS* 5593–5601 (2014). doi:10.1073/pnas.1419161111

80. Xia, J., Benner, M. J. & Hancock, R. E. W. NetworkAnalyst -integrative approaches for protein–protein interaction network analysis and visual exploration. *Nucleic Acids Res.* **42,** W167–W174 (2014).

81. Sotiriou, C. *et al.* Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98,** 262–272 (2006).

82. Desmedt, C. *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* **13,** 3207–3214 (2007).

83. Schmidt, M. *et al.* The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68,** 5405–5413 (2008).

Figure Legends

Figure 1. ***A phenotypic, imaging-based, RNAi screen identifies novel regulators of tumor cell migration.*** (A) Live cell imaging of Hs578T and MDA-MB-231 phagokinetic tracks. Scale bar is 100 µm. (B) Schematic representation of PKT screen. Transfection was performed in 96-well plates and controls were included in each plate. After 65 h, transfected cells were washed, trypsinized, diluted and seeded onto PKT assay plates. Plates were fixed after 7 h of cell migration and whole-well montages were acquired using transmitted light microscopy. For each siRNA knockdown, a robust Z-score was calculated for each PKT parameter. (C) The three most dominant quantitative PKT parameters are shown for Hs578T and (D) MDA-MB-231. Representative images of migratory tracks for genes with strong effect are shown below each graph and highlighted for enhancement (red) and inhibition (green). (E) Principal component analysis of migratory phenotypes in Hs578T and (F) MDA-MB-231. Migratory phenotypes were identified manually and corresponding Z-scores were determined. Only hits in each phenotypic class and mock control are plotted, and representative images of each phenotype are shown. (G) Overlap of hits in each phenotypic class in both Hs578T and MDA-MB-231.

Figure 2. ***Candidate migratory gene validation by deconvolution PKT screen.*** (A) Representative images of valid candidate genes using 4 individual single siRNAs in three phenotypic classes are shown. Scale bar is 100 µm. (B) 282 selected genes were tested in a deconvolution PKT screen with 4 single siRNAs per gene. SMARTpool and single siRNA Z-scores of Net Area and Axial Ratio are shown for the 'overlap hits' that were validated in Hs578T. (C) Distribution of validated candidate genes in the five phenotypic classes. (D) Overlap of validated hits between Hs578T and MDA-MB-231. (E) Distribution of the validated genes involved in tumor cell migration in Hs578T, MDA-MB-231 over the different libraries based on validated hits effective in both cell lines. Statistical significance was determined using a Fisher's exact test. * $p < 0.05$

Figure 3. ***Candidate genes directly affect TNBC cell migratory behavior.*** (A) Quantification of single cell migration speed of Hs578T-GFP cells after knockdown of validated hits. Hs578T-GFP cells were transfected with siRNAs and cell migration was assessed by live microscopy. Hits were required to show significant and consistent effects in both replicates to be considered as candidate genes (right panel). Median ± 95% confidence interval is shown and cell populations were compared by Kruskal-Wallis test with Dunn's post correction test. (B) Same as in A, with MDA-MB-231-GFP cells. (C) Single cell trajectories of

cell migration upon knockdown of PRPF4B, MXD1, BUD31 or BPTF in (i) Hs578T and (ii) MDA-MB-231 cell lines.

Figure 4. ***Regulatory networks drive tumor cell migration***. (A) Enrichment of KEGG Pathways in PPI networks generated from Hs578T candidate genes and (B) MDA-MB-231 candidate genes. NetworkAnalyst was used to generate PPI networks. (C) Zero-order interaction network of combined Hs578T and MDA-MB-231 candidate genes reveals a highly connected subnetwork of clinically associated genes (in blue). Candidate genes inhibiting cell migration in both cell lines are shown in red; central hubs are highlighted in green. The degree of connectivity (number of connections) is displayed on the right. (D) Phenotype-based clustering of the PKT validated candidate genes based on morphological changes in the Hs578T cell line. Per parameter, log2FC compared to mock control was calculated. Clustering was performed based on Euclidean distance and complete linkage.

Figure 5. ***Candidate modulators of TNBC cell migration are related to breast cancer metastasis-free survival and breast cancer progression***. (A) Prognostic gene signatures were used to generate minimum interaction PPI networks and compared to our candidate TNBC cell migration gene networks. Candidate genes affecting cell migration feed into similar networks essential for BC progression and metastasis formation. (B) Hierarchical clustering (Euclidean distance, complete linkage) of genetic modifications (mutations, deletions and amplifications combined) of 43 candidate genes in 29 cancer types. Data was derived from The Cancer Genome Atlas. Annotation shows the expression of the candidate in relation to BC metastasis-free survival in different BC subtypes. P-values were calculated using Cox proportional hazards regression analysis, with gene expression values as continuous variable and metastasis-free survival as end point. Genes marked in red and blue are highlighted in C. Red genes were selected for further analysis. (C) Contribution of different genetic modifications to the rate in several highly mutated or amplified candidates. (D) Kaplan Meier curves of for expression of PRPF4B, BUD31 and BPTF and relation to metastasis-free survival in ERpos or TNBC breast cancer. Gene expression data of lymph-node negative BC patient cohort without prior treatment using optimal split was used.

Figure 6. ***PRPF4B, BUD31 and BPTF depletion affects expression of focal adhesion and ECM-receptor signaling related genes.*** (A) Over-representation analysis of genes with decreased expression levels (log2FC change < -1) after PRPF4B knockdown in Hs578T and MDA-MB-231 cells using pathways annotated in the KEGG database. (B) Gene Set Enrichment Analysis (GSEA) identifies the ECM-receptor interaction pathway as significantly enriched in downregulated genes after PRPF4B knockdown in Hs578T and MDA-MB-231

cell lines. (C) Hierarchical clustering (Euclidean distance, complete linkage) of log2FC in expression levels of genes involved in the KEGG ECM-receptor interaction pathway after knockdown of candidates demonstrates that many genes involved in this pathway are downregulated after candidate knockdown. (D) LAMA5, ITGB1 and ITGA3 expression upon depletion of PRPF4B, BUD31 and BPTF (mean + stdev of 3 biological replicates, one-way ANOVA, *P<0.05, **P<0.01, ***P<0.001). (E) Effect of PRPF4B, BUD31 and BPTF depletion on levels of different ECM and focal adhesion components in MDA-MB-231 cells analyzed with western blot. (F) Effect of indicated gene depletion on focal adhesion and actin cytoskeleton organization. Hs578T cells were fixed and stained against the actin cytoskeleton, p-FAK (Y397). Scale bar is 50 µm.

Figure 7. ***PRPF4B is essential for breast cancer metastasis formation in vivo.*** (A) mRNA and (B) protein expression of PRPF4B in shPRPF4B MDA-MB-417.5 cell line compared to shCtrl#1 and shCrtl#2 MDA-MB-417.5 cell lines (one-way ANOVA, *P<0.05, **P<0.01). (C) Immunofluorescent staining for PRPF4B (green) and DNA (DAPI; blue) in shCtrl#1, shCtrl#2 and shPRPF4B MDA-MB-417.5 cell lines. (D) Tumor growth of shPRPF4B, shCtrl#1 and shCtrl#2 MDA-MB-417.5 cells engrafted in the mammary fat pad of Rag2–/– Il2rg–/– mice (n=7-8 animals per group). (E) Dissemination of shPRPF4B, shCtrl#1 and shCtrl#2 MDA-MB-417.5 cells in the lung, liver, spleen, kidney, heart, uterus and axillar lymph node as determined by ex vivo bioluminescent analysis (total flux (p/s)) of different target tissues. (F) Number of lung metastases for shPRPF4B, shCtrl#1 and shCtrl#2 injected mice as determined by macroscopic evaluation of lungs injected with ink. (G) Images of ink injected lungs (left) and bioluminescent signal in the lungs (right) of all individual mice. Blanc indicates mice that did not receive MDA-MD-417.5 cells. Groups were compared by Kruskal-Wallis test with Dunn's post correction test (luminescence) or ANOVA (metastasis count). * $p < 0.05$, ** $p < 0.01$.

FIGURE 1

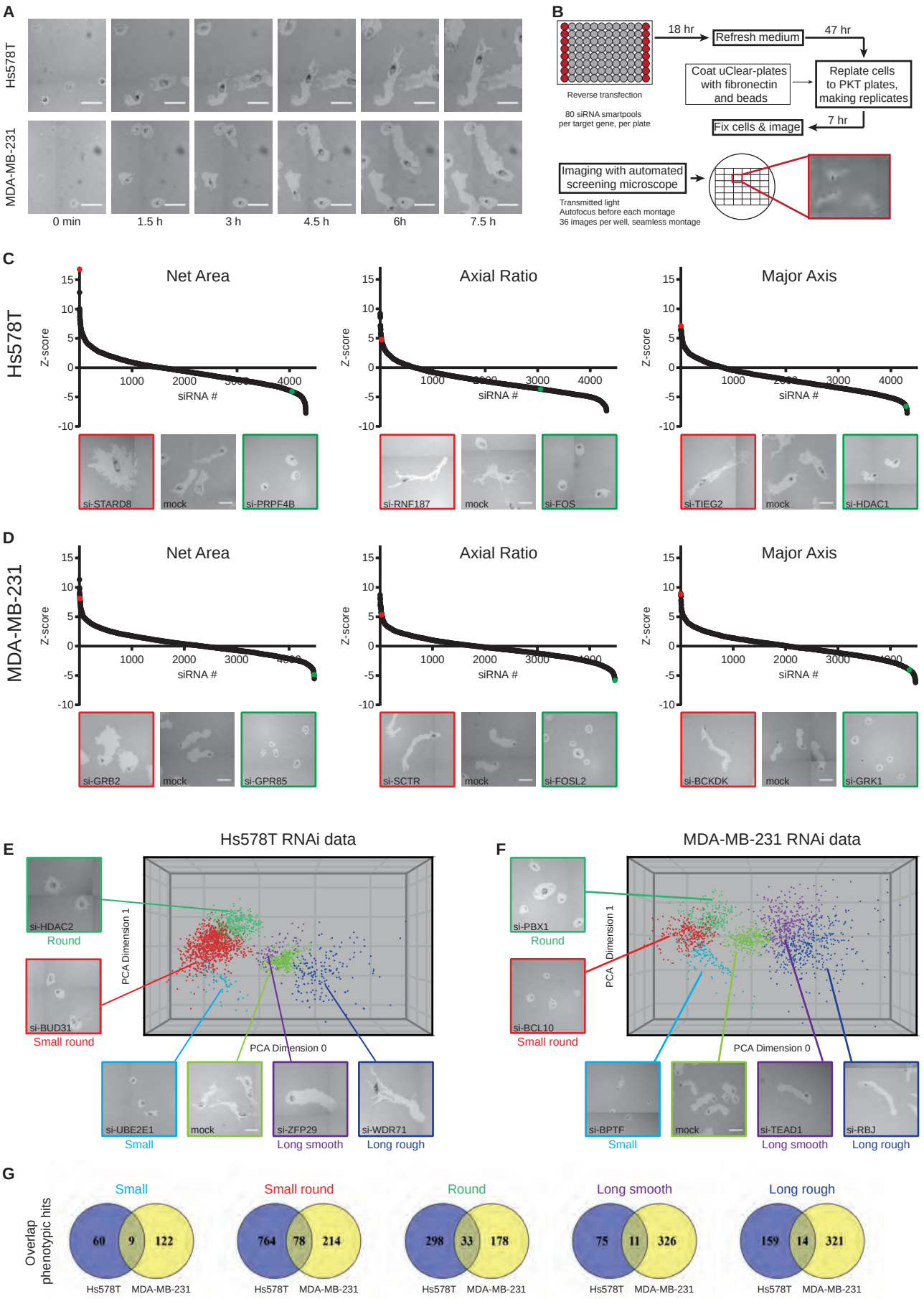

A
Hs578T
MDA-MB-231
0 min
1.5 h
3 h
4.5 h
6h
7.5 h
B
18 hr
Refresh medium
47 hr
Coat uClear-plates with fibronectin and beads
Replate cells to PKT plates, making replicates
7 hr
Fix cells & image
Reverse transfection
80 siRNA smartpools per target gene, per plate
Imaging with automated screening microscope
Transmitted light
Autofocus before each montage
36 images per well, seamless montage
C
Hs578T
Net Area
Axial Ratio
Major Axis
Z-score
siRNA #
si-STARD8
mock
si-PRPF4B
si-RNF187
si-FOS
si-TIEG2
si-HDAC1
D
MDA-MB-231
si-GRB2
si-GPR85
si-SCTR
si-FOSL2
si-BCKDK
si-GRK1
E
Hs578T RNAi data
si-HDAC2
Round
si-BUD31
Small round
PCA Dimension 1
PCA Dimension 0
si-UBE2E1
Small
si-ZFP29
Long smooth
si-WDR71
Long rough
F
MDA-MB-231 RNAi data
si-PBX1
si-BCL10
si-BPTF
si-TEAD1
si-RBJ
G
Overlap phenotypic hits
Small
60
9
122
Small round
764
78
214
Round
298
33
178
Long smooth
75
11
326
Long rough
159
14
321
Hs578T
MDA-MB-231

FIGURE 2

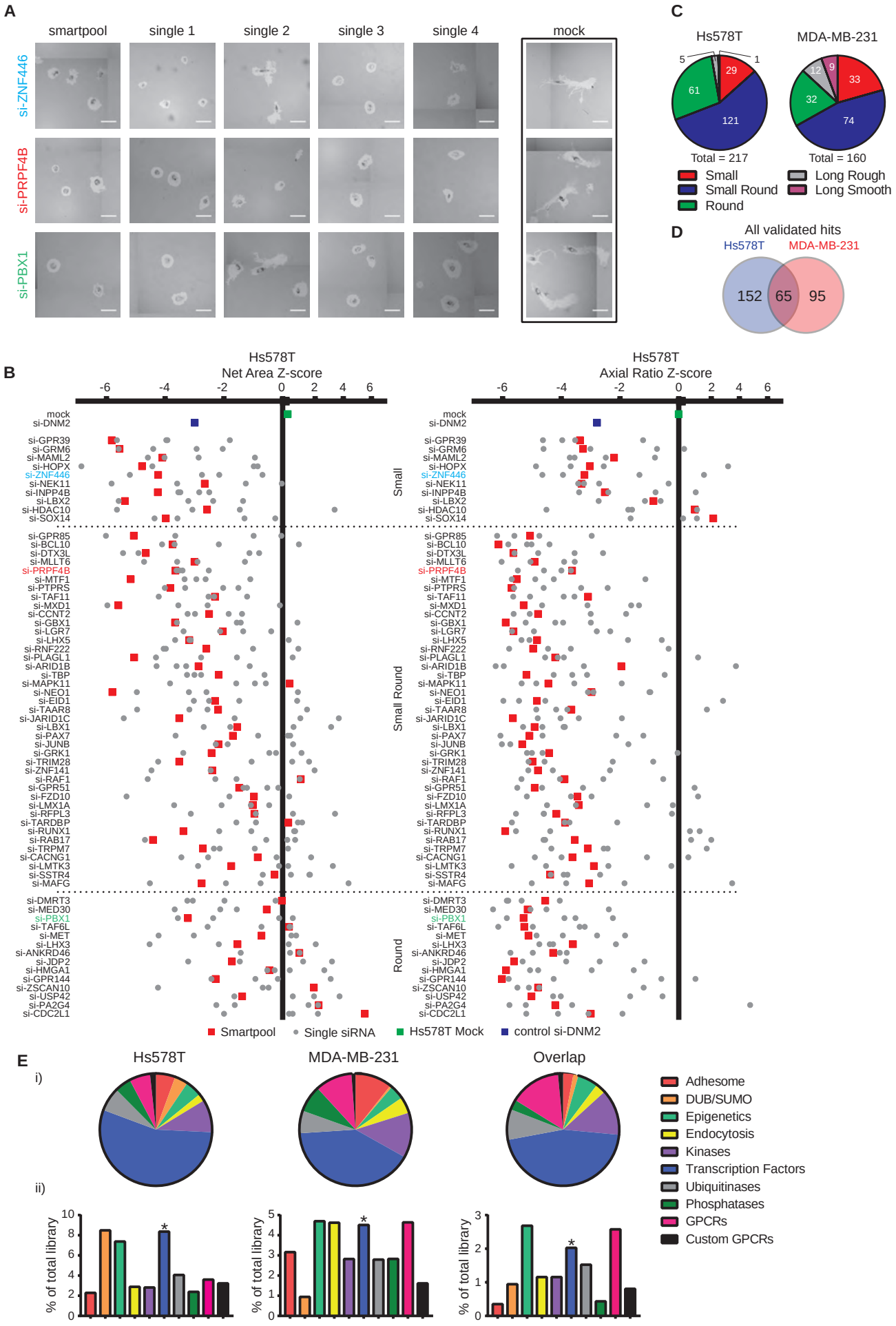

A
smartpool
single 1
single 2
single 3
single 4
mock
si-ZNF446
si-PRPF4B
si-PBX1
C
Hs578T
MDA-MB-231
Total = 217
Total = 160
Small
Small Round
Round
Long Rough
Long Smooth
D
All validated hits
Hs578T
MDA-MB-231
152
65
95
B
Hs578T
Net Area Z-score
Hs578T
Axial Ratio Z-score
Small
Small Round
Round
Smartpool
Single siRNA
Hs578T Mock
control si-DNM2
E
i)
Hs578T
MDA-MB-231
Overlap
Adhesome
DUB/SUMO
Epigenetics
Endocytosis
Kinases
Transcription Factors
Ubiquitinases
Phosphatases
GPCRs
Custom GPCRs
ii)
% of total library

FIGURE 3

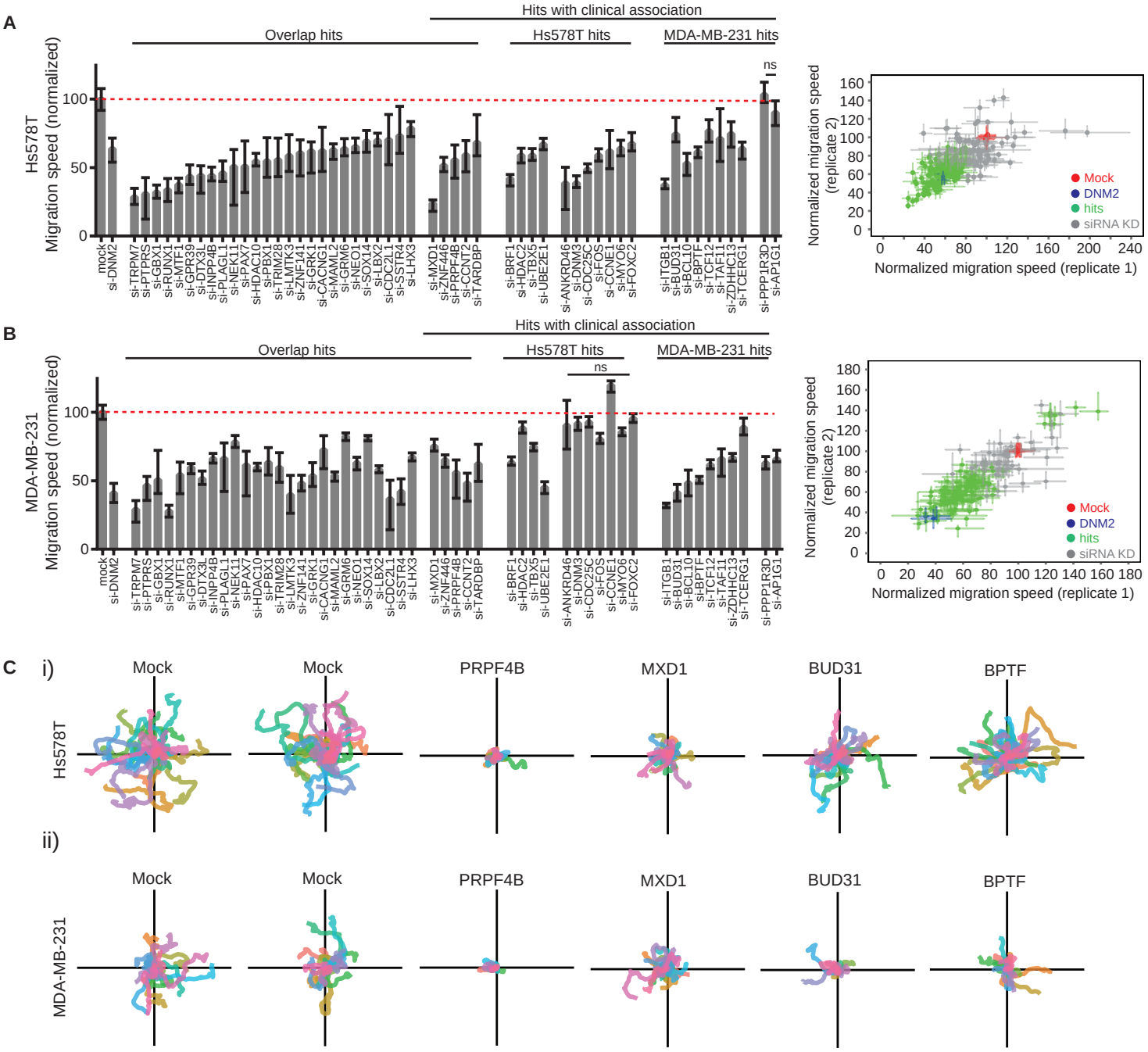

A
Hits with clinical association
Overlap hits
Hs578T hits
MDA-MB-231 hits
Hs578T
Migration speed (normalized)
ns
Normalized migration speed (replicate 2)
Normalized migration speed (replicate 1)
Mock
DNM2
hits
siRNA KD
B
MDA-MB-231
C
i)
ii)
Mock
PRPF4B
MXD1
BUD31
BPTF

FIGURE 4

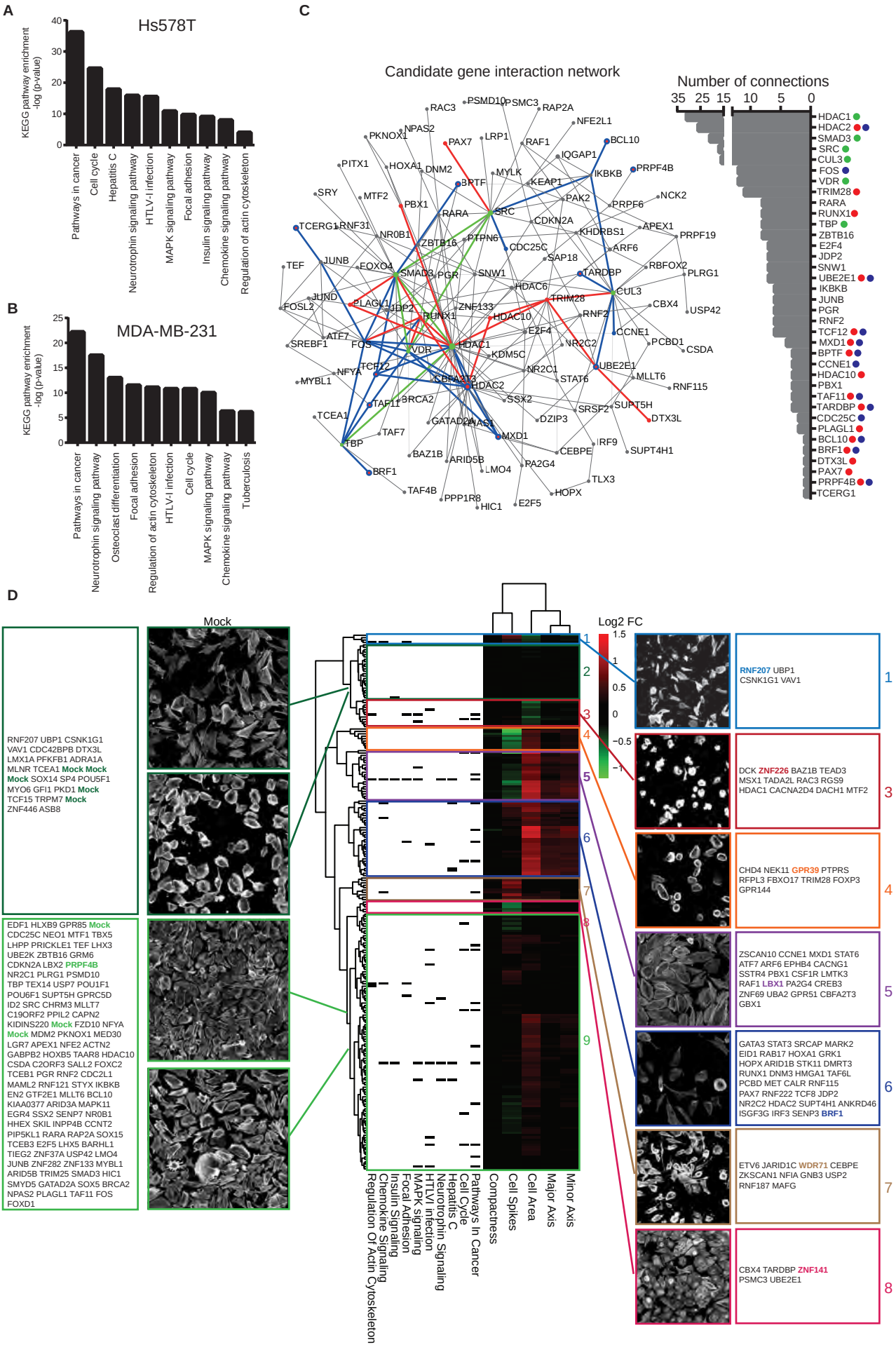

A
Hs578T
KEGG pathway enrichment -log (p-value)
Pathways in cancer
Cell cycle
Hepatitis C
Neurotrophin signaling pathway
HTLV-I infection
MAPK signaling pathway
Focal adhesion
Insulin signaling pathway
Chemokine signaling pathway
Regulation of actin cytoskeleton
B
MDA-MB-231
KEGG pathway enrichment -log (p-value)
Pathways in cancer
Neurotrophin signaling pathway
Osteoclast differentiation
Focal adhesion
Regulation of actin cytoskeleton
HTLV-I infection
Cell cycle
MAPK signaling pathway
Chemokine signaling pathway
Tuberculosis
C
Candidate gene interaction network
Number of connections
D
Mock
Log2 FC
Compactness
Cell Spikes
Cell Area
Major Axis
Minor Axis
Pathways In Cancer
Cell Cycle
Hepatitis C
Neurotrophin Signaling
HTLV-I infection
MAPK signaling
Focal Adhesion
Insulin Signaling
Chemokine Signaling
Regulation Of Actin Cytoskeleton
RNF207 UBP1 CSNK1G1 VAV1
DCK ZNF226 BAZ1B TEAD3 MSX1 TADA2L RAC3 RGS9 HDAC1 CACNA2D4 DACH1 MTF2
CHD4 NEK11 GPR39 PTPRS RFPL3 FBXO17 TRIM28 FOXP3 GPR144
ZSCAN10 CCNE1 MXD1 STAT6 ATF7 ARF6 EPHB4 CACNG1 SSTR4 PBX1 CSF1R LMTK3 RAF1 LBX1 PA2G4 CREB3 ZNF69 UBA2 GPR51 CBFA2T3 GBX1
GATA3 STAT3 SRCAP MARK2 EID1 RAB17 HOXA1 GRK1 HOPX ARID1B STK11 DMRT3 RUNX1 DNM3 HMGA1 TAF6L PCBD MET CALR RNF115 PAX7 RNF222 TCF8 JDP2 NR2C2 HDAC2 SUPT4H1 ANKRD46 ISGF3G IRF3 SENP3 BRF1
ETV6 JARID1C WDR71 CEBPE ZKSCAN1 NFIA GNB3 USP2 RNF187 MAFG
CBX4 TARDBP ZNF141 PSMC3 UBE2E1
RNF207 UBP1 CSNK1G1 VAV1 EDF1 HLXB9 GPR85 Mock CDC25C NEO1 MTF1 TBX5 LHPP PRICKLE1 TEF LHX3 UBE2K ZBTB16 GRM6 CDKN2A LBX2 PRPF4B NR2C1 PLRG1 PSMD10 TBP TEX14 USP7 POU1F1 POU6F1 SUPT5H GPRC5D ID2 SRC CHRM3 MLLT7 C19ORF2 PPIL2 CAPN2 KIDINS220 Mock FZD10 NFYA Mock MDM2 PKNOX1 MED30 LGR7 APEX1 NFE2 ACTN2 GABPB2 HOXB5 TAAR8 HDAC10 CSDA C2ORF3 SALL2 FOXC2 TCEB1 PGR RNF2 CDC2L1 MAML2 RNF121 STYX IKBKB EN2 GTF2E1 MLLT6 BCL10 KIAA0377 ARID3A MAPK11 EGR4 SSX2 SENP7 NR0B1 HHEX SKIL INPP4B CCNT2 PIP5KL1 RARA RAP2A SOX15 TCEB3 E2F5 LHX5 BARHL1 TIEG2 ZNF37A USP42 LMO4 JUNB ZNF282 ZNF133 MYBL1 ARID5B TRIM25 SMAD3 HIC1 SMYD5 GATAD2A SOX5 BRCA2 NPAS2 PLAGL1 TAF11 FOS FOXD1

FIGURE 5

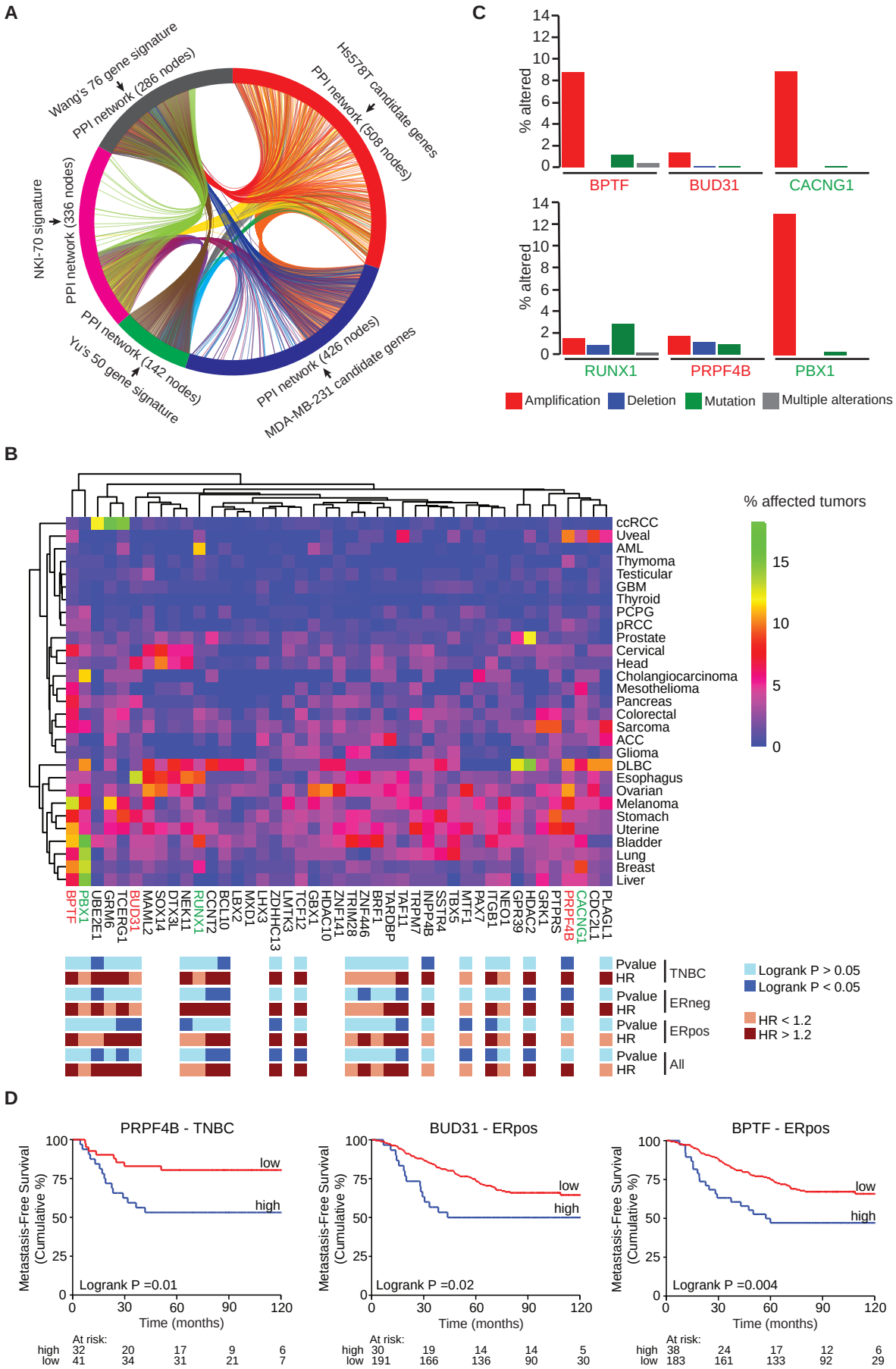

A
Wang's 76 gene signature
PPI network (286 nodes)
Hs578T candidate genes
PPI network (508 nodes)
NKI-70 signature
PPI network (336 nodes)
PPI network (142 nodes)
Yu's 50 gene signature
PPI network (426 nodes)
MDA-MB-231 candidate genes
C
% altered
BPTF
BUD31
CACNG1
RUNX1
PRPF4B
PBX1
Amplification
Deletion
Mutation
Multiple alterations
B
% affected tumors
ccRCC
Uveal
AML
Thymoma
Testicular
GBM
Thyroid
PCPG
pRCC
Prostate
Cervical
Head
Cholangiocarcinoma
Mesothelioma
Pancreas
Colorectal
Sarcoma
ACC
Glioma
DLBC
Esophagus
Ovarian
Melanoma
Stomach
Uterine
Bladder
Lung
Breast
Liver
Pvalue
HR
TNBC
ERneg
ERpos
All
Logrank P > 0.05
Logrank P < 0.05
HR < 1.2
HR > 1.2
D
PRPF4B - TNBC
BUD31 - ERpos
BPTF - ERpos
Metastasis-Free Survival (Cumulative %)
Time (months)
low
high
Logrank P =0.01
Logrank P =0.02
Logrank P =0.004
At risk:
high 32 20 17 9 6
low 41 34 31 21 7
high 30 19 14 14 5
low 191 166 136 90 30
high 38 24 17 12 6
low 183 161 133 92 29

FIGURE 6

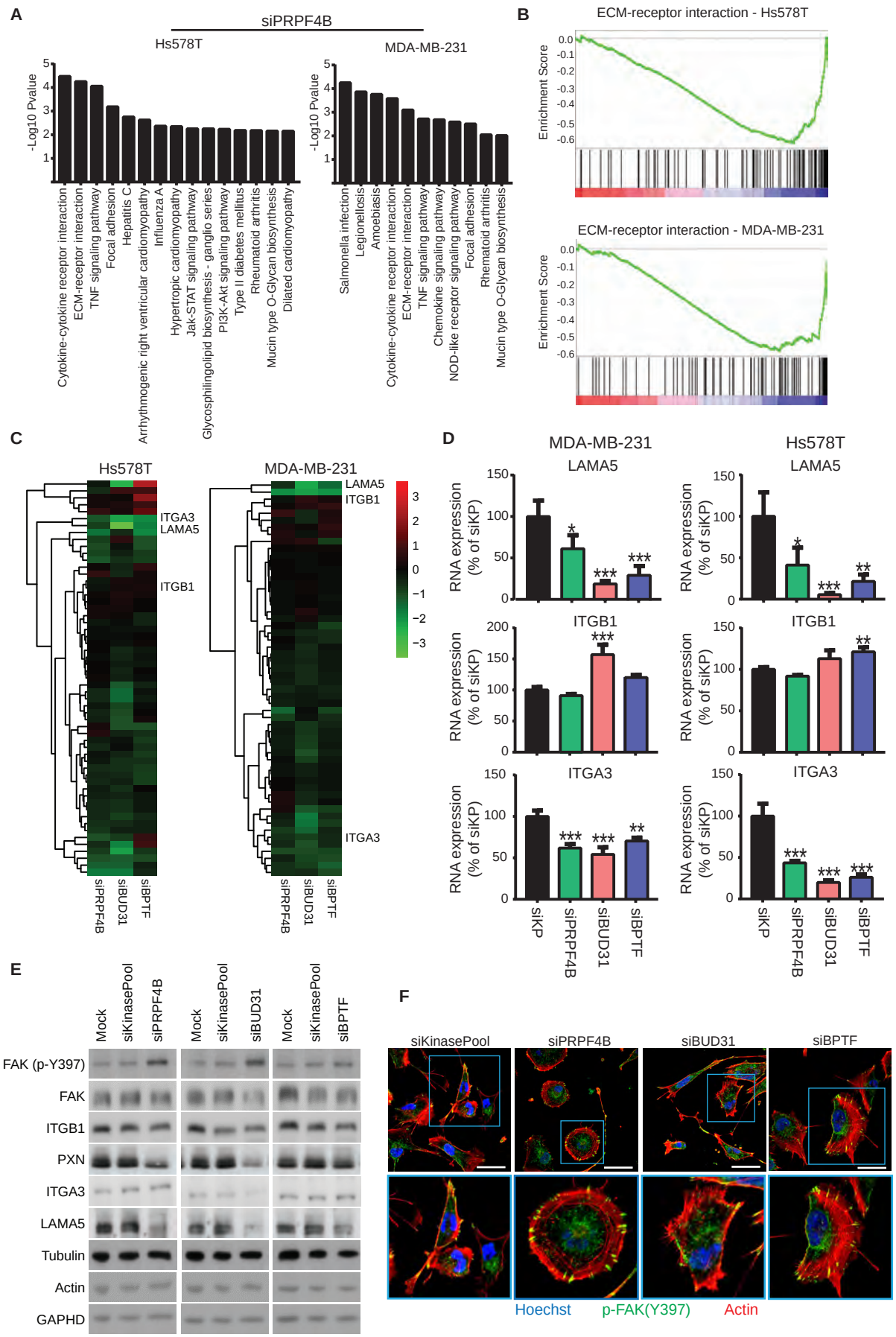
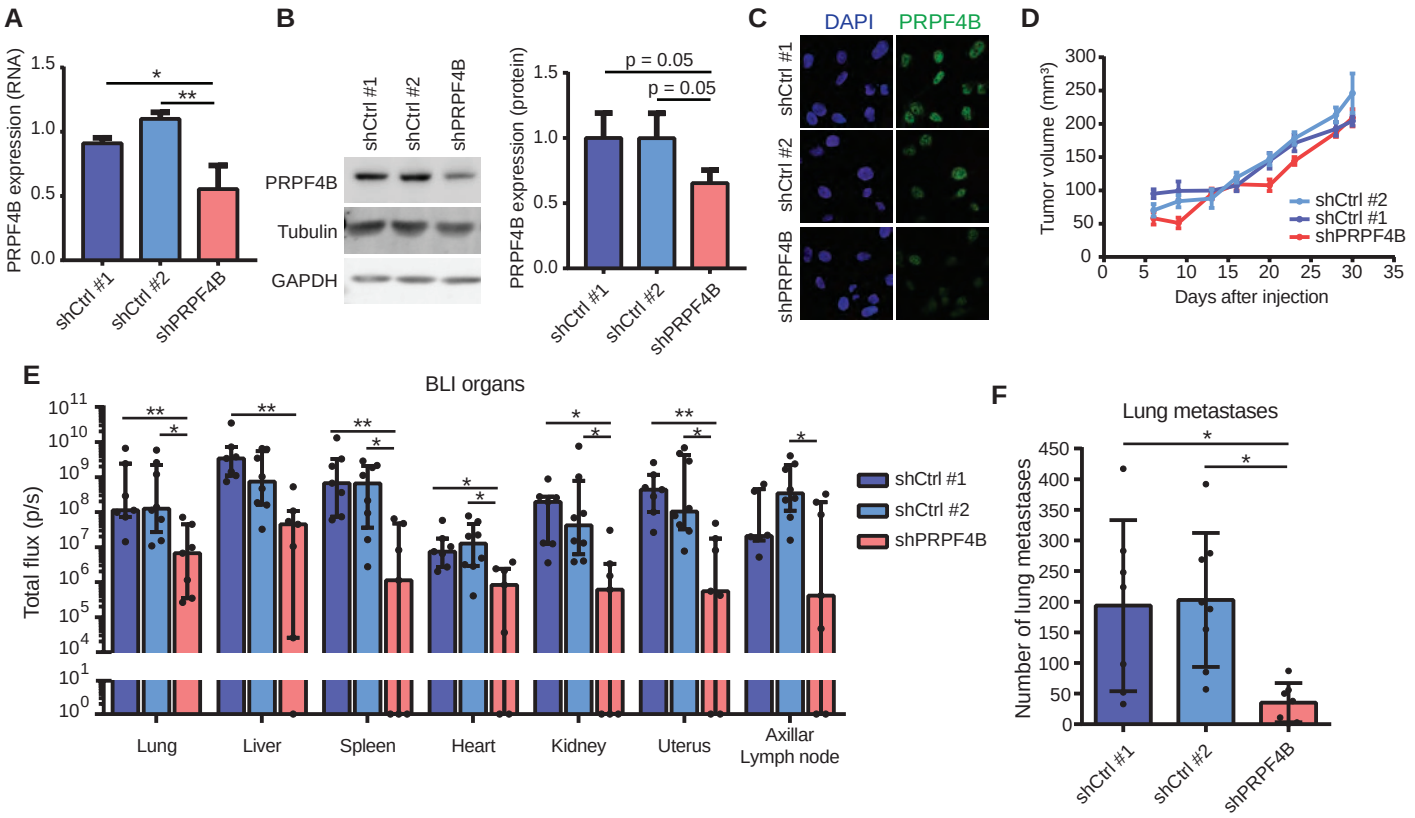
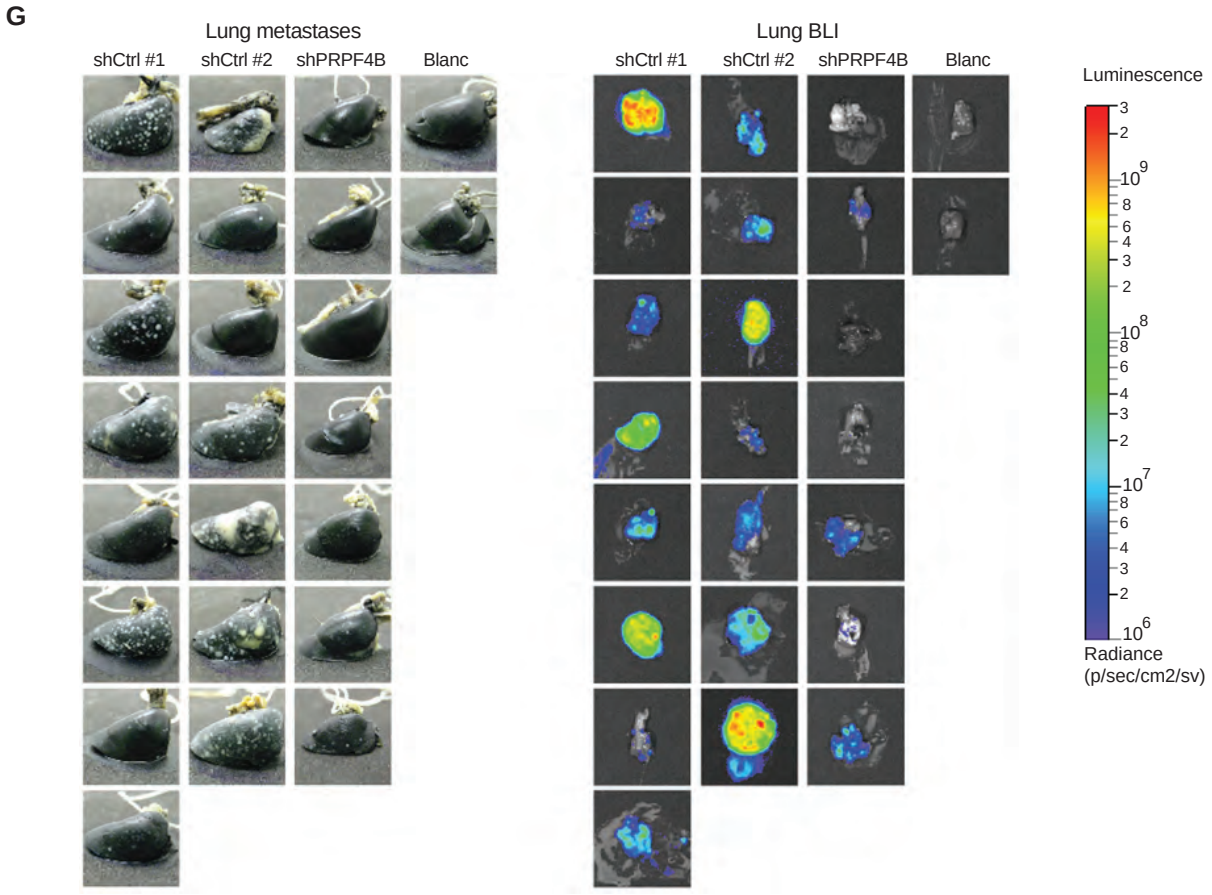

A
siPRPF4B
Hs578T
MDA-MB-231
-Log10 Pvalue
Cytokine-cytokine receptor interaction
ECM-receptor interaction
TNF signaling pathway
Focal adhesion
Hepatitis C
Arrhythmogenic right ventricular cardiomyopathy
Influenza A
Hypertrophic cardiomyopathy
Jak-STAT signaling pathway
Glycosphilingolipid biosynthesis - ganglio series
PI3K-Akt signaling pathway
Type II diabetes mellitus
Rheumatoid arthritis
Mucin type O-Glycan biosynthesis
Dilated cardiomyopathy
Salmonella infection
Legionellosis
Amoebiasis
Chemokine signaling pathway
NOD-like receptor signaling pathway
B
ECM-receptor interaction - Hs578T
ECM-receptor interaction - MDA-MB-231
Enrichment Score
C
Hs578T
MDA-MB-231
ITGA3
LAMA5
ITGB1
siPRPF4B
siBUD31
siBPTF
D
MDA-MB-231
Hs578T
LAMA5
ITGB1
ITGA3
RNA expression (% of siKP)
siKP
siPRPF4B
siBUD31
siBPTF
E
Mock
siKinasePool
siPRPF4B
siBUD31
siBPTF
FAK (p-Y397)
FAK
ITGB1
PXN
ITGA3
LAMA5
Tubulin
Actin
GAPHD
F
siKinasePool
siPRPF4B
siBUD31
siBPTF
Hoechst
p-FAK(Y397)
Actin

FIGURE 7


A
PRPF4B expression (RNA)
1.5
1.0
0.5
0.0
*
**
shCtrl #1
shCtrl #2
shPRPF4B
B
shCtrl #1
shCtrl #2
shPRPF4B
PRPF4B
Tubulin
GAPDH
PRPF4B expression (protein)
p = 0.05
p = 0.05
C
DAPI
PRPF4B
shCtrl #1
shCtrl #2
shPRPF4B
D
Tumor volume (mm³)
300
250
200
150
100
50
0
0
5
10
15
20
25
30
35
Days after injection
shCtrl #2
shCtrl #1
shPRPF4B
E
BLI organs
Total flux (p/s)
Lung
Liver
Spleen
Heart
Kidney
Uterus
Axillar Lymph node
shCtrl #1
shCtrl #2
shPRPF4B
F
Lung metastases
Number of lung metastases
450
400
350
300
250
200
150
100
50
0
shCtrl #1
shCtrl #2
shPRPF4B


G
Lung metastases
shCtrl #1
shCtrl #2
shPRPF4B
Blanc
Lung BLI
shCtrl #1
shCtrl #2
shPRPF4B
Blanc
Luminescence
Radiance
(p/sec/cm2/sv)