

RESEARCH ARTICLE

Open Access

Integrative genomic and functional profiling of the pancreatic cancer genome

A Hunter Shain¹, Keyan Salari^{1,2}, Craig P Giacomini¹ and Jonathan R Pollack^{1*}

Abstract

Background: Pancreatic cancer is a deadly disease with a five-year survival of less than 5%. A better understanding of the underlying biology may suggest novel therapeutic targets. Recent surveys of the pancreatic cancer genome have uncovered numerous new alterations; yet systematic functional characterization of candidate cancer genes has lagged behind. To address this challenge, here we have devised a highly-parallel RNA interference-based functional screen to evaluate many genomically-nominated candidate pancreatic cancer genes simultaneously.

Results: For 185 candidate pancreatic cancer genes, selected from recurrently altered genomic loci, we performed a pooled shRNA library screen of cell growth/viability across 10 different cell lines. Knockdown-associated effects on cell growth were assessed by enrichment or depletion of shRNA hairpins, by hybridization to barcode microarrays. A novel analytical approach (CORrelated Phenotypes for On-Target Effects; COPOTE) was used to discern probable on-target knockdown, based on identifying different shRNAs targeting the same gene and displaying concordant phenotypes across cell lines. Knockdown data were integrated with genomic architecture and gene-expression profiles, and selected findings validated using individual shRNAs and/or independent siRNAs. The pooled shRNA library design delivered reproducible data. In all, COPOTE analysis identified 52 probable on-target gene-knockdowns. Knockdown of known oncogenes (KRAS, MYC, SMURF1 and CCNE1) and a tumor suppressor (CDKN2A) showed the expected contrasting effects on cell growth. In addition, the screen corroborated purported roles of PLEKHG2 and MED29 as 19q13 amplicon drivers. Most notably, the analysis also revealed novel possible oncogenic functions of nucleoporin NUP153 (ostensibly by modulating TGF β signaling) and Kruppel-like transcription factor KLF5 in pancreatic cancer.

Conclusions: By integrating physical and functional genomic data, we were able to simultaneously evaluate many candidate pancreatic cancer genes. Our findings uncover new facets of pancreatic cancer biology, with possible therapeutic implications. More broadly, our study provides a general strategy for the efficient characterization of candidate genes emerging from cancer genome studies.

Keywords: Pancreatic cancer, Functional genomics, RNAi screen, shRNA screen, NUP153, KLF5

Background

Pancreatic ductal adenocarcinoma (hereafter, pancreatic cancer) is the fourth leading cause of cancer death in the United States [1,2]. The five-year survival rate is a dismal 5%, as effective treatment regimens are limited [3]. A better understanding of the underlying disease biology is needed to develop new and successful treatment strategies to manage this deadly disease.

Several key molecular genetic alterations in pancreatic cancer have been identified [4,5]. Activating mutations of *KRAS* occur in 95% of cases. The *CDKN2A* locus, encoding p16INK4A and p14ARF, which respectively intersect the Rb and p53 pathways, is homozygously deleted in 80% of tumors. *TP53* is itself inactivated, usually through point mutation, in 55% of cases. *SMAD4*, a central mediator of TGF β signaling, is deleted in approximately 50% of cases. Furthermore, *TGFBR2*, its upstream receptor, is deleted in 20% of tumors, underscoring a central importance of this signaling pathway in pancreatic cancer. *MYC* is amplified in approximately 30% of cases. Recently, deletions and mutations in five different subunits of the SWI/SNF chromatin

* Correspondence: pollack1@stanford.edu

¹Departments of Pathology, Stanford University School of Medicine, 269 Campus Drive, CCSR-3245A, Stanford, CA 94305-5176, USA
Full list of author information is available at the end of the article

remodeling complex have been found to occur in about a third of cases [6]. However, despite what is already known, recent surveys of the pancreatic cancer genome have identified scores of additional candidate cancer genes that merit further investigation [7,8].

With the advent of DNA microarrays and “next-generation” DNA sequencing, the field of genomics has transformed our ability to study diseases like cancer on an “omic” scale. Over the past decade, these technologies have spurred structural studies producing a compendium of cancer alterations, including DNA mutations, deletions, amplifications, and rearrangements. Yet, because of the sheer volume of data, such studies have far outpaced our ability to functionally evaluate candidate cancer genes [9].

The development of RNA interference (RNAi) techniques has accelerated our capacity to study knockdown phenotypes and infer the function and mechanism of disease genes [10]. While traditionally used to characterize single genes at a time, several groups have adapted the technology to use small interfering RNA (siRNA) or short hairpin RNA (shRNA) libraries for high-throughput screens [11], including in pancreatic cancer [12-17]. These large-scale, highly parallel efforts provide the potential to functionally annotate genes on an “omic” scale.

Here, we describe a high-throughput functional interrogation of the pancreatic cancer genome using an shRNA-based screen. We simultaneously evaluate 185 candidate pancreatic cancer genes, nominated from genomic profiles, across 10 genetically diverse cell lines. After integrating the functional and genomic data, we further characterize nine top candidates, both uncovering new pancreatic cancer biology and validating an integrative approach for the functional annotation of cancer genomes.

Methods

Cell lines

Cancer cell lines were obtained directly from the American Type Culture Collection, and grown in RPMI-1640 high-glucose media (Invitrogen) supplemented with 10% fetal bovine serum (Hyclone). HPDE cells [18] were obtained from Dr. Ming Tsao (University of Toronto), and grown in keratinocyte serum-free media (supplemented as directed with EGF and bovine pituitary extract; Invitrogen).

Pooled shRNA lentiviral library screen

The shRNA screen, schematically depicted in Figure 1 (with summary information in Table 1), was adapted from published protocols [19,20]. Potential advantages of a pooled (compared to well-based) screen include economies of scale and discernment of subtle fitness effects by competitive growth over many days. The 185 targeted genes were selected based on the identification of recurrent structural abnormalities (focal DNA amplifications, deletions, and/or mutations) in pancreatic cancer genomes

[6,7]. These abnormalities are listed for each gene in Additional file 1. GIPZ lentiviral shRNAmir constructs targeting these genes (average 3, range 1–7 shRNA/gene) were obtained from Open Biosystems/Thermo Scientific; catalog numbers are listed in Additional file 2. The 558 pGIPZ shRNAmir plasmid DNAs were combined at equimolar concentration into a single pool. The shRNA DNA pool was then used to transfect 293T cells, together with a trans lentiviral packaging mix (Open Biosystems). Pooled shRNA lentiviral supernatant was collected 48 hrs later, and frozen in aliquots to improve screen reproducibility.

The lentiviral library was then used to infect target cell lines at low multiplicity of infection (average 0.3 integrants/cell; determined by flow cytometry of GFP expression from the GIPZ vector), so that most cells contained a single shRNA knocking down the expression of a single gene. Additionally, enough cells were infected to provide an average representation of approximately 1,000 lentiviral integrations (range 150–2000) for each of the 558 shRNAs in the library, mitigating potential artifacts from specific integration sites or from multiple integrations [20,21]. To infect target cell lines, lentivirus was diluted in serum/antibiotic-free media containing 10 µg/ml polybrene (determined to optimally enhance infection with minimal cell toxicity). Cells were spun at 30°C for 1 hr at 2,400 rpm, allowed to recover for 4 hrs (37°C, 5% CO₂), spun again for an additional 1 hr, and then the media replaced with complete RPMI-1640 growth media. All target cell line infections were carried out minimally in triplicate (range 3–7 replicates).

Two days post infection, a fraction of the infected cells was harvested for an initial time point (T = 0), and puromycin selection (using cell line-specific levels previously determined by killing curves) was then initiated for the remaining cells. Cells were cultured for an additional 4 weeks in the presence of selective media. Care was taken not to allow the cells to become too confluent or to split too thinly. Cells were periodically harvested, including the last time point (T = 4 weeks), which was the standard comparison point for the screen data presented.

Genomic DNA was isolated from harvested cells, sheared ten times through a 25 gauge needle, and subsequently used as template for PCR amplification of library shRNA hairpins. Sufficient genomic DNA template was included in the PCR reaction to ensure an approximate 1,000-fold average representation of each library shRNA hairpin (assuming 5 pg of DNA/cell), thereby maintaining the initial ~1,000-fold representation of integrations. PCR primers common to the shRNAmir backbone vector and bounding the half hairpin target sequence were as follows: Forward 5'-TAGTGAAGCCACAGATGTA-3'; Reverse 5'-ATGTATCAAAGAGATAGCAAGGTATTTCAG-3'.

To deconvolute shRNA representation in the T = 4 weeks vs. T = 0 cell pools, gel-purified PCR products were

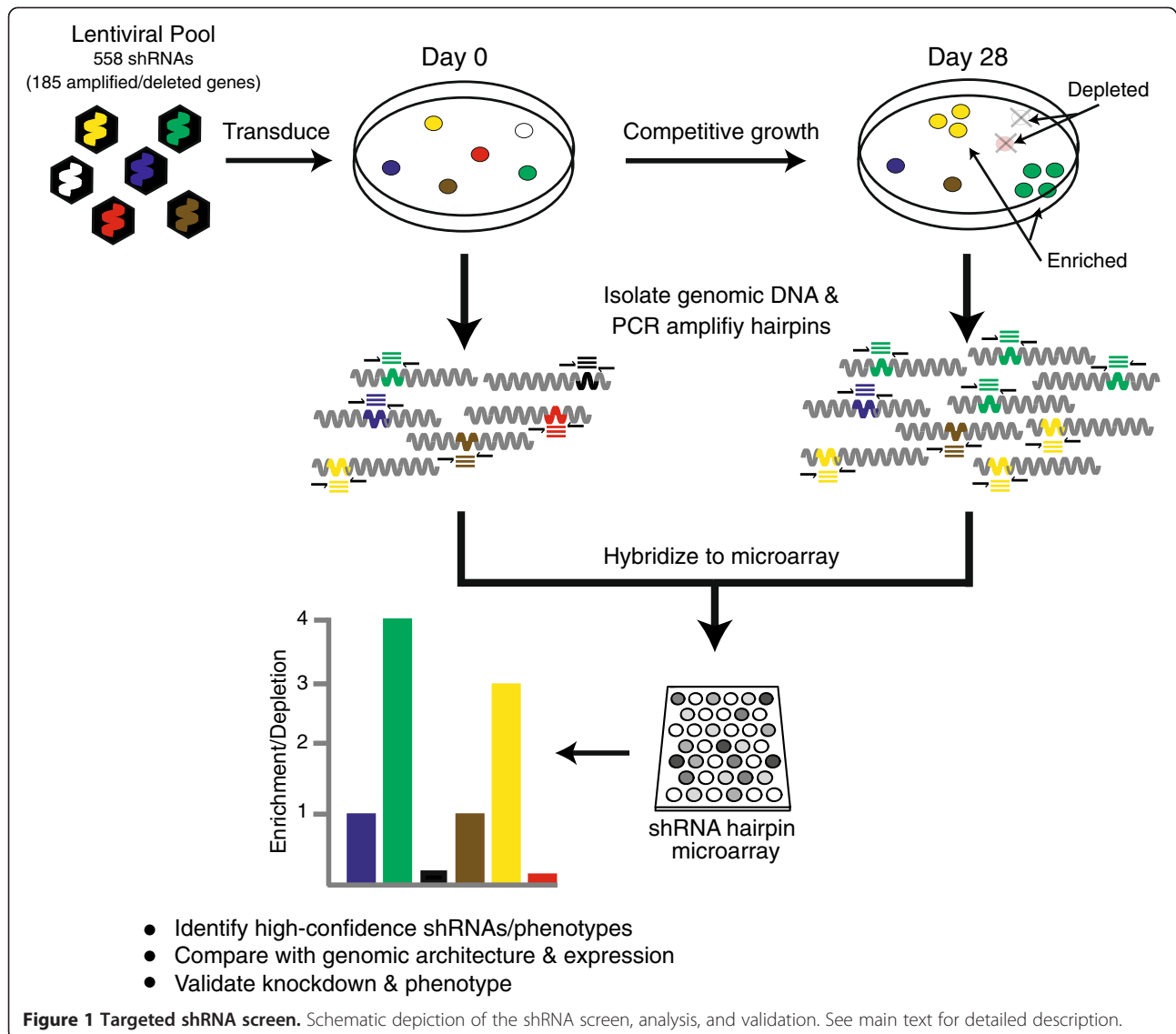


Table 1 Cell lines included in shRNA screen

Cell line	Description	Number of replicate infections	Multiplicity of infection	shRNA fold-representation
Aspc1	Pancreatic cancer	3	0.3	1,000
BXPC3	Pancreatic cancer	3	0.3	1,500
Capan1	Pancreatic cancer	7	0.3	400
HPAC	Pancreatic cancer	3	0.3	1,500
Panc1	Pancreatic cancer	6	0.3	150
PL5	Pancreatic cancer	3	0.3	2,000
SU86.86	Pancreatic cancer	3	0.3	1,000
HPDE	Immortalized, non-tumorigenic pancreatic ductal epithelial cells	3	0.3	1,000
MDA157	Breast cancer	3	0.3	600
SKBR3	Breast cancer	3	0.3	1,000

differentially labeled and co-hybridized (using manufacturer protocols) to a custom Agilent microarray designed (using Agilent eArray software) to include probes complementary to shRNA hairpin sequences. Microarrays were imaged and fluorescence intensities extracted. For each microarray, \log_{10} background-subtracted fluorescence ratios were globally normalized (which assumes overall equal library shRNA depletion and enrichment). Unless otherwise specified, screen data reported represent the average ratios from replicate cell line infections.

Screen data analysis and validation

High-confidence on-target shRNAs/phenotypes were identified using a new analytical approach (CORrelated Phenotypes for On-Target Effects; COPOTE), based on finding two or more different shRNAs targeting the same gene and displaying concordant enrichment/depletion profiles across the cell lines (Pearson correlation). Custom Perl scripts were used to calculate the Pearson correlation coefficient between shRNAs targeting the same genes, and also to permute cell line identities to generate the randomized null distribution of correlations. To define meaningful correlations, we determined False Discovery Rates (FDRs) as the ratio of false positives (from the null distribution) to observed positives at or above any given correlation cutoff.

For validation, Q-RT-PCR was performed using Assay-on-Demand TaqMan probes and reagents (Applied Biosystems). Catalog numbers for probes include: NUP153 (Hs01018919_m1) and KLF5 (Hs00156145_m1). Western blots were done on whole cell lysates, using the following primary antibodies: Myc (Santa Cruz sc-40), pan-Ras (Millipore 05-516), NUP153 (Acris BM5527), GAPDH (Santa Cruz sc-25778), SMURF1 (Santa Cruz H-60). Densitometry calculations were carried out using publicly available ImageJ software. For siRNA knockdown, ON-TARGETplus SMARTpool siRNAs were obtained from Thermo Scientific: KRAS (L-005069-00-0005), SMURF1 (L-007191-00-0005), NUP153 (L-005283-00-0005), KLF5 (L-013571-00-0005), non-targeting control (D-001810-10-20). Cell growth/viability assays were done using a modified WST-1 protocol as previously described [6]. In each assay, a full time course was performed, though usually only day 5 is shown for brevity.

Array-based comparative genomic hybridization (aCGH) data

Findings from the screen were interpreted in the context of previously published aCGH data [6] (GSE26089). Briefly, that data set comprises Agilent 244 K CGH array profiles from 70 pancreatic cancers (48 primary tumor xenografts and 22 cancer cell lines). Tumor/normal fluorescence ratios were normalized and mapped onto the genome (build 18) using Agilent software.

Results and discussion

Targeted shRNA screen

In pancreatic (as with other) cancer genomes, loci that are recurrently amplified or deleted are likely enriched for known or novel cancer genes [22]. Typically, each such locus is studied individually to discover the driver gene(s). In an effort to accelerate this process, we developed a pooled shRNA screening strategy to simultaneously evaluate 185 candidate pancreatic cancer genes that together represent 104 different loci of recurrent DNA amplification or deletion; thus the genes selected included both candidate oncogenes and tumor suppressors (see Additional file 1). Most of the 185 candidates were focally amplified or deleted in a subset of cancers, and some were also reported to carry mutations (Additional file 1) [6,7]. A few known cancer genes (e.g. *KRAS*, *MYC*, and *CDKN2A*) were also included in the screen as positive controls.

The pooled shRNA screen was carried out as a competitive growth/viability assay. The general workflow, adapted from Schlabach *et al.* [19], is depicted in Figure 1 and described in more detail in Methods. In brief, pancreatic cancer cell lines were infected with a pooled shRNA lentiviral library comprising 558 shRNAs targeting the 185 genes (on average 3 hairpins/gene) (Additional file 1). Infections were done at low multiplicity of infection, ensuring that most cells harbored a single shRNA knocking down the expression of a single gene. The pooled, infected cells were then cultured for four weeks, after which depleted or enriched shRNAs (i.e., those targeting genes conferring positive or negative growth advantage, respectively) were identified by PCR amplification of shRNA hairpins and comparative hybridization ($T = 4$ weeks *vs.* $T = 0$) to a custom hairpin microarray.

The shRNA library screen was carried out on ten different cell lines (Table 1). These included seven genetically-diverse pancreatic cancer cell lines that together harbor the vast majority of copy number alterations from which the 185 genes were selected. We also included a single immortalized, non-tumorigenic human pancreatic ductal epithelial (HPDE) cell line [18] to facilitate discovery of tumor suppressor genes (as enriched shRNAs targeting growth suppressive genes). Lastly, we also screened two breast cancer cell lines to help distinguish genome-specific from generic essential genes (i.e., an "out-group"). The raw and cell line-averaged screening data are available in Additional files 2 and 3.

Screen analysis and validation

The pooled shRNA library screen yielded high quality and reproducible data, as supported by multiple lines of evidence. First, replicate screens of each cell line showed good correlation; the average pair-wise Pearson correlation coefficients (R-values) for cell line replicates ranged from 0.3-0.9 (mean 0.5). The two cell lines (Capan1 and Panc1)

with the lowest R-values were those with the least shRNA library representation; however, this was mitigated by more screen replicates (Table 1). Notably, in an unsupervised analysis of the shRNA depletion/enrichment data, cell line replicates most often clustered together (Figure 2A). Second, time course samplings, done for a subset of the cell lines, demonstrated consistently increasing shRNA depletion (or enrichment) over time (Figure 2B). Third, we noted that distinct shRNAs targeting the same gene often exhibited correlated growth phenotypes (depletion/enrichment profiles) across the panel of cell line screened (Figure 2C). We capitalized on this last observation to define “high-confidence” on-target knockdown phenotypes, an approach we termed COPOTE (CORrelated Phenotypes for On-Target Effects).

When analyzing hundreds of shRNAs, it is possible that two different shRNAs targeting the same gene have a similar growth phenotype (across the panel of cell lines) just by random chance. Therefore, to correct for multiple hypothesis testing, we compared the distribution of observed Pearson correlations to that from randomly permuted data (Figure 2D). Notably, we observed a significant rightward shift in the observed distribution of correlations, indicating an enrichment of shRNAs targeting the same gene (and showing similar phenotype) above that expected by chance; false discovery rates (FDR; q values) are reported in Additional file 4. Of the 185 genes included in the screen (and of the 157 genes (85%) represented by two or more shRNAs and therefore evaluable), we identified 52 genes with “high confidence” knockdown phenotypes, defined by having at least two different shRNAs exhibited significantly correlated depletion/enrichment profiles ($R > 0.7$; $q < 0.135$) across the ten cell lines.

We note that other approaches have been used to leverage information from multiple shRNAs targeting the same gene, including redundant siRNA activity (RSA) analysis [23], RNAi gene enrichment ranking (RIGER) [24], and the Gene Activity Ranking Profile (GARP) score [17]. All of these methods consider either the top-most depleted shRNA(s) or all shRNAs for a given gene. Our approach is fundamentally different in that we consider all shRNA pairs, therefore focusing primarily on the evidence for correlated (implying on-target) phenotype rather than on the strength of shRNA depletion.

Focusing primarily on the 52 high-confidence genes, we next validated knockdown for a subset of biologically-interesting genes. We included some known pancreatic cancer genes, but mainly focused on novel genes with biological functions plausibly related to pancreatic cancer, and also demonstrating logical connections to the underlying genomic data, e.g. shRNA depletion for amplified candidate oncogenes. Validation was done by transducing individual (rather than the pool of) shRNAs, or by transfecting independent siRNAs (which we found to be more time and cost

effective), and then verifying target knockdown by RT-PCR and/or Western blot (summarized in Table 2). For known pancreatic cancer oncogenes and tumor suppressor genes, finding the expected growth phenotypes (as described below) provided additional screen validation.

Observations on known cancer genes

Among the high-confidence shRNAs/phenotypes, several shRNAs targeting known oncogenes showed the anticipated pattern of depletion (i.e. knockdown reduces growth fitness) (Figure 3). For example, three different knockdown-validated [26] shRNAs targeting the *KRAS* oncogene were generally depleted across the panel of cell lines, though with some cell lines (e.g. Panc1 and HPAC) showing substantially more depletion than others (e.g. PL5) (Figure 3A). On-target knockdown of *KRAS* and consequent reduced cell viability were verified using independent *KRAS*-targeting siRNAs (Figure 3A). In pancreatic cancer, *KRAS* frequently harbors activating point mutations and may be amplified and/or overexpressed. Notably, in the cell line panel there was a good correlation of *KRAS*-shRNA screen depletion levels with both *KRAS* mutation status and *KRAS* transcript levels ($R = 0.8$). This pattern of dependency is consistent with classic oncogene addiction [27].

The shRNAs targeting the *MYC* oncogene also exhibited variable depletion across the panel of cell lines. However, unlike with *KRAS*, the cell lines with lower *MYC* transcript levels were associated with substantially higher depletion levels (Figure 3B). The basis for this observation is unclear, and we caution that the RNAi knockdown and phenotype requires validation. Nonetheless, we speculate that there might exist a threshold below which *MYC* levels are insufficient to support cell growth, and that in cell lines with high *MYC* levels incomplete shRNA-mediated knockdown is insufficient to pass that threshold.

We recently reported *SMURF1*, encoding an E3 ubiquitin ligase, to be amplified in a subset of pancreatic cancers, where it drives cell invasion but not growth in AsPC1 cells [33]. Here, in the context of the pooled shRNA screen, we reproduced that result. The single shRNA targeting *SMURF1* was neither depleted nor enriched in *SMURF1*-amplified AsPC1 cells (Figure 3C), consistent with its role in cell invasion but not growth. Interestingly, in non-amplified cell lines the *SMURF1* shRNA was generally depleted, suggesting a possible additional role in cell growth specific to a non-amplified context, a knockdown threshold effect as speculated above for *MYC* (endogenous *SMURF1* expression levels are substantially elevated with amplification [33]), or an off-target effect. The findings for *SMURF1* also suggest an added benefit of screening additional phenotypes, e.g. cell invasion.

We also observed the expected pattern of shRNA enrichment for shRNAs targeting known tumor suppressor

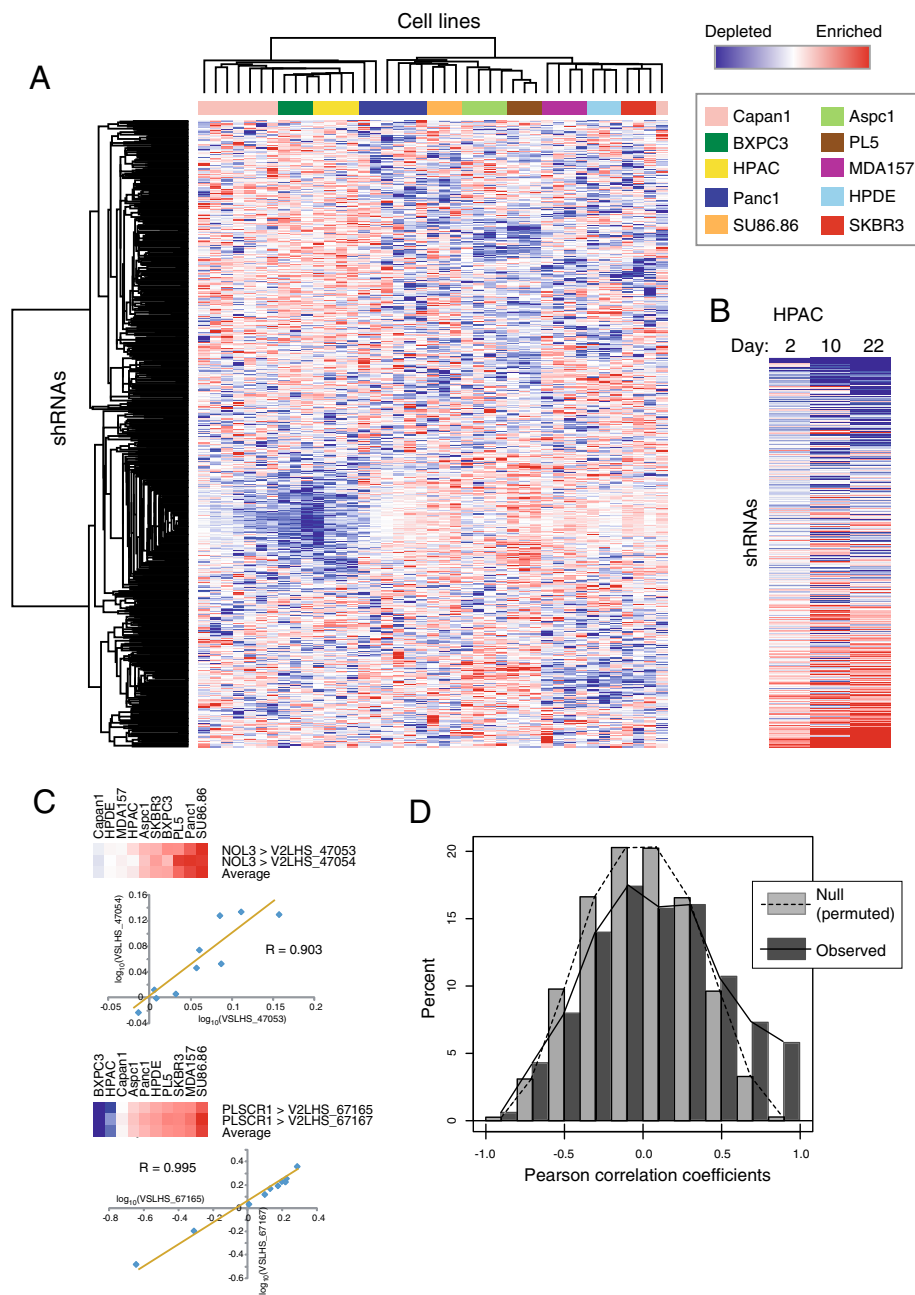


Figure 2 Performance of shRNA screen. (A) Heatmap representation of unsupervised hierarchical cluster analysis of screen data. Columns represent screened cell lines (key provided), and rows represent the 558 different shRNAs (here, mean-centered). The color scale indicates fold-depletion (blue) or fold-enrichment (red) of shRNAs in T = 4 week vs. T = 0 comparisons. Note that replicate screens for each cell line tend to cluster together, which demonstrates reproducibility across all steps of the screen (infection, cell line passaging, PCR amplification, and microarray hybridization). The basis for the single outlier (Capan1, far right) is unclear. (B) Time course analysis of shRNA depletion/enrichment, shown for a single cell line (HPAC). Note that across the time series, library shRNAs show expected increased depletion (blue, top of heat map) or enrichment (red; bottom of heat map), consistent with competitive growth selection. (C) Depletion/enrichment status of selected genes is shown. Knockdown of the same gene by different shRNAs often produced concordant phenotypes; e.g., Pearson correlation ($R = 0.903$) for two hairpins targeting *NOL3*, and $R = 0.995$ for two hairpins targeting *PLSCR1*. (D) Frequency plot of observed Pearson correlations (R values; correlating depletion/enrichment vectors across the cell line panel) between hairpin pairs targeting the same gene (dark gray bars) compared to a null distribution of Pearson correlations generated from randomly permuted data (light gray bars). Note, the rightward shift of observed Pearson correlations, above that expected by chance, reflects enrichment of on-target shRNAs/phenotypes.

Table 2 Summary of validation for target knockdown and growth phenotype

Gene	shRNA knockdown validated by RT-PCR	shRNA knockdown validated by western blot	shRNA knockdown validated by cancer genome anatomy Project [25]	siRNA knockdown validated by RT-PCR	siRNA knockdown validated by western blot	siRNA knockdown growth phenotype validated
<i>KRAS</i>		X [26]		X	X	X
<i>SMURF1</i>		X			X	X
<i>CDKN2A</i>			X			
<i>CCNE1</i>			X			
<i>NUP153</i>	X				X	X
<i>KLF5</i>		X		X	X	X

genes, best exemplified by *CDKN2A*. A single shRNA targeting *CDKN2A* was enriched in a subset of the screened cell lines (Figure 3D). Notably, the cell lines that did not show substantial enrichment were in fact those with homozygous deletion of *CDKN2A* (where knockdown of a non-existent gene would not be expected to promote growth). These data suggest a utility of the shRNA knockdown screen approach in identifying not only oncogenes (by reduced growth fitness), but also potentially novel tumor suppressor genes (by enhanced growth fitness).

While the focus of our screen was pancreatic cancer, the inclusion of a breast cancer out-group also provided an opportunity to identify breast cancer selective dependencies. Among the high-confidence genes/phenotype, most notably shRNAs targeting the cell-cycle regulator *CCNE1* showed selective depletion in the two breast cancer cell lines (Figure 3E). *CCNE1* has been reported to be the driver of 19q12 amplification in breast cancer [34,35]. Indeed, Natajaran *et al.* reported MDA157 breast cancer cells to be sensitive to *CCNE1* knockdown [34], which we replicated here (Figure 3E). Our findings underscore a likely selective dependency and possible point of therapeutic attack in breast cancer.

Cytoband 19q13 amplicon

While *KRAS* (12p12) and *MYC* (8q24) are among the most commonly amplified oncogenes, the 19q13 region also shows broad amplification in the majority of pancreatic cancers, and high-amplitude, focal amplifications in approximately 20% of cases (Figure 4A). The recurrence and focality strongly suggest the presence of an important oncogene within this amplicon. *AKT2*, which functions in the PI3 kinase pathway, resides near this chromosomal region, and would be a natural candidate for the primary driver of this amplification. However, as we previously noted [6], the smallest common region of amplification occurs proximal to *AKT2*, as *AKT2* is excluded from the amplicon in several pancreatic cancers (Figure 4A). These data imply that *AKT2* is not the principal driver of 19q13 amplification. To evaluate this amplicon, our screen included shRNAs targeting each of the 12

genes residing within the smallest common region of amplification.

From our screen, three shRNAs targeting genes residing at 19q13 were preferentially depleted in those cell lines harboring 19q13 amplification (Figure 4B). Among these, two high-confidence (correlated) shRNAs targeted *PLEKHG2*, encoding a plekstrin homology domain protein, and a single shRNA (and the only one included in the screen) targeted *MED29*, a mediator of RNA polymerase II transcription complex subunit. These data agree well with findings described by Kuuselo *et al.* [36]. Those investigators performed a focused siRNA-based analysis of genes within the 19q13 amplicon, reporting that knockdown of *PLEKHG2* and *MED29* reduced cell viability in Panc1 (19q-amplified) but not Miapaca2 (not amplified) cells. Thus, our data corroborate and extend (using independent assays and additional cell lines) prior work and highlight a still under-appreciated role of *PLEKHG2* and *MED29* (rather than *AKT2*) as likely oncogenes driving 19q13 amplification in pancreatic cancer.

NUP153/FAST1 TGFβ signaling axis

One novel observation derived from our screen centers on *NUP153*, which encodes a nuclear pore complex protein (or nucleoporin). *NUP153* was included in the screen because we identified it to be focally amplified in a single pancreatic cancer cell line, PL5 (Figure 5A). The *NUP153* nucleoporin regulates the distribution of specific proteins between the nucleus and the cytoplasm, interestingly including the transducer of TGFβ signaling, *SMAD2* [37]. In particular, *NUP153* stoichiometrically competes with *FAST1*, a *SMAD2* nuclear retention factor, to shuttle *SMAD2* out of the nucleus, thus dampening TGFβ signaling. Further implicating this axis in pancreatic cancer, we also noted recurrent focal deletions of *FAST1* in other pancreatic cancer cases (Figure 5B).

In our screen, two different high-confidence shRNAs targeting *NUP153* showed varying levels of depletion across the cell line panel. Notably, cell lines with an intact upstream TGFβ signaling pathway (determined by absence of *SMAD4* and *TGFBR2* mutation/deletion,

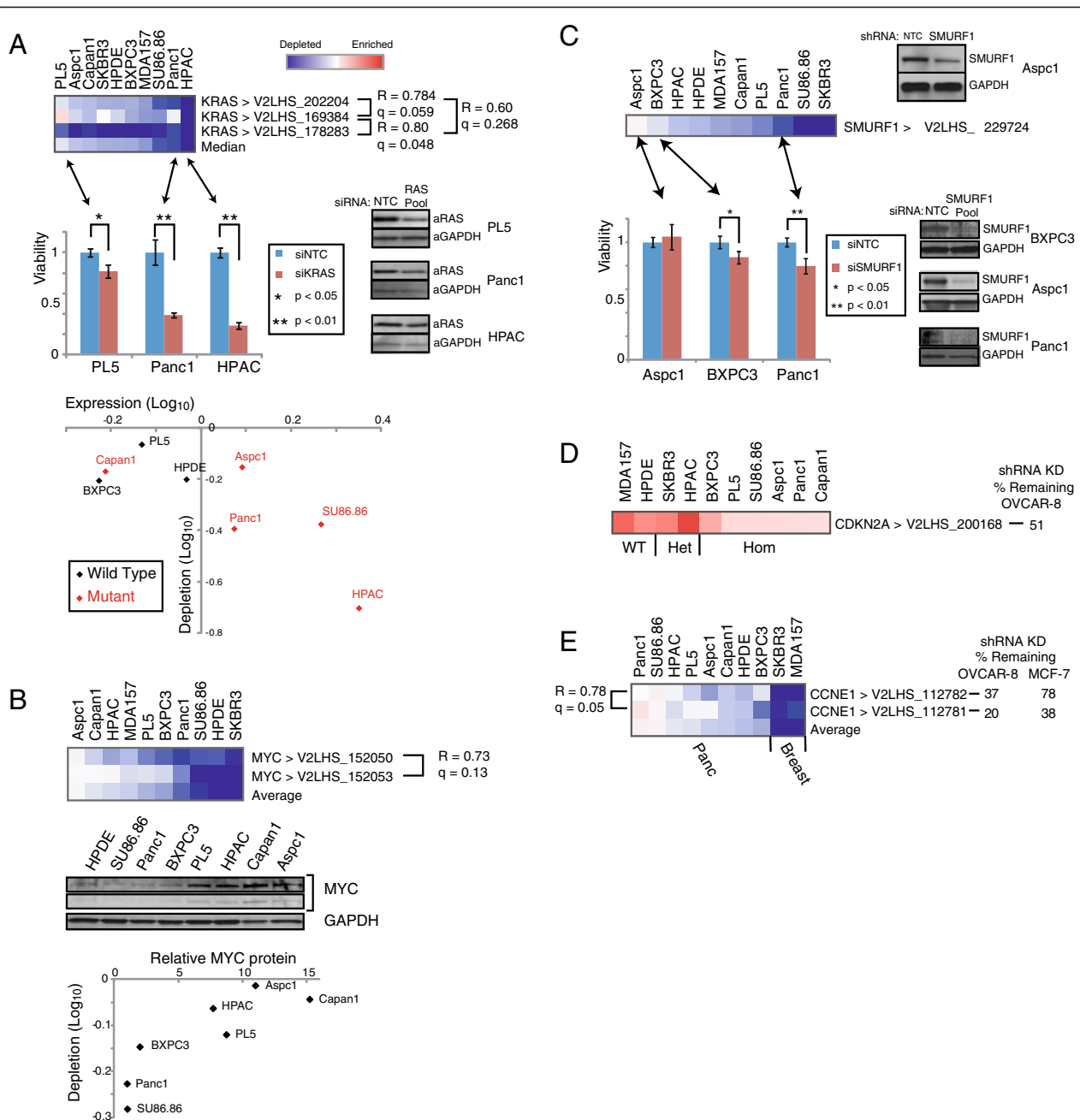
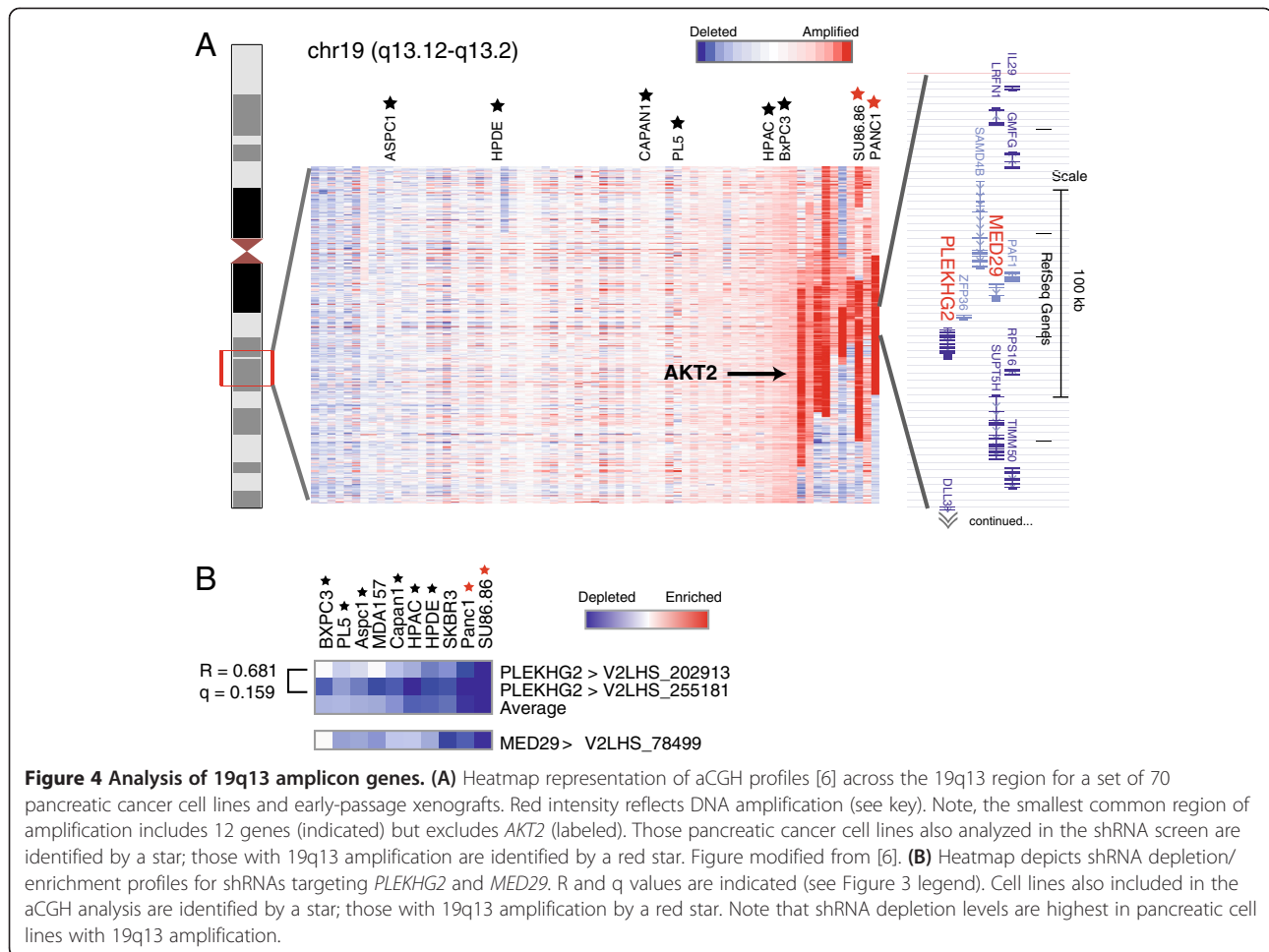


Figure 3 Evaluation of known cancer genes. (A) KRAS. Above left, heatmap depicts shRNA depletion (blue) or enrichment (red) for each KRAS-targeting hairpin across the cell line panel (ordered by median shRNA depletion levels). The corresponding Pearson correlation (R) and false discovery rate (q) values are indicated. For select cell lines, the effect of KRAS knockdown (Western blot, middle right) on cell growth/viability (WST-1 assay, middle left) was independently verified by siRNA transfection. Indicated P-values determined by Student's t-test (KRAS-siRNA vs. non-targeting control). Below, depletion of KRAS-targeting shRNAs is most evident in cell lines with KRAS activating mutation (red data points) and/or elevated KRAS transcript levels. KRAS mutational status was sequenced for PL5 [28] and previously reported for the other lines [18,29-32]. **(B) MYC.** Above, heatmap shows shRNA depletion/enrichment profile, as described above. Below, levels of shRNA depletion are inversely correlated with MYC protein levels; MYC protein levels quantified by Western blot (middle). **(C) SMURF1.** Above left, heatmap depicts shRNA depletion/enrichment profile; knockdown validated by Western blot (middle). For select cell lines, the effect of SMURF1 knockdown (Western blot, below right) on cell growth/viability (below left) was independently verified by siRNA transfection. **(D) CDKN2A.** Heatmap depicts shRNA depletion/enrichment profile; CDKN2A deletion status of cell lines is indicated. Note, the shRNA used here was pre-validated [25] by Q-RT-PCR in OVCA8 cell line (% residual transcript indicated). **(E) CCNE1.** Heatmap depicts shRNA depletion/enrichment profile; shRNAs pre-validated (in OVCA8 and MCF7 cells) as above.



and/or TGFβ-induced growth suppression) generally exhibited higher levels of *NUP153*-shRNA depletion, compared to cell lines with a compromised TGFβ signaling pathway (Figure 5C) (Additional file 5). Our findings therefore suggest that altered NUP153-FAST1 shuttling of SMADs may provide an alternate means by which pancreatic cancers disrupt TGFβ signaling. The role of an altered NUP153/FAST1 TGFβ shuttling axis in pancreatic carcinogenesis warrants further investigation.

KLF5 dependency

Another compelling finding relates to *KLF5*, the intestinal kruppel-like transcription factor. *KLF5* was included in our screen because we identified it to be focally amplified in a pancreatic cancer cell line, HPAC (Figure 6A). Interestingly, *KLF5* has been paradoxically reported to function as both a tumor suppressor [39,40] and an oncogene [41,42] in different tumor types (reviewed in [43]). Our finding of focal *KLF5* amplification in pancreatic cancer suggests a possible oncogenic role.

In our screen, three different high-confidence shRNAs targeting *KLF5* showed depletion in most cell lines

(Figure 6B). Though interestingly, a single cell line, Panc1, with extremely low levels of *KLF5* transcript showed substantial *KLF5*-shRNA enrichment (Figure 6B-D). We verified on-target *KLF5* knockdown and the contrasting knockdown growth-phenotypes (in Panc1 compared to two other lines) by siRNA transfection (Figure 6B, C). The basis for shRNA enrichment in the sole Panc1 line is unclear. We speculate that *KLF5* might be epigenetically-silenced in Panc1 (based on barely-detectable expression) because in that cell line it plays a growth suppressive role; thus further knockdown would promote cell growth. In contrast, in every other cell line in the panel, *KLF5* transcript levels are appreciable, and knockdown reduces cell growth (Figure 6D, E). Taken together, our findings support a context-dependency of *KLF5* function, whilst the preponderance of data (focal amplification and growth dependency) supports a predominantly oncogenic role in pancreatic cancer.

Of note, a recent exome sequencing study [7] reported a single heterozygous mutation (H389N) in the DNA-binding domain of *KLF5* in the pancreatic cell line Panc5.04. The authors interpreted the mutation to be

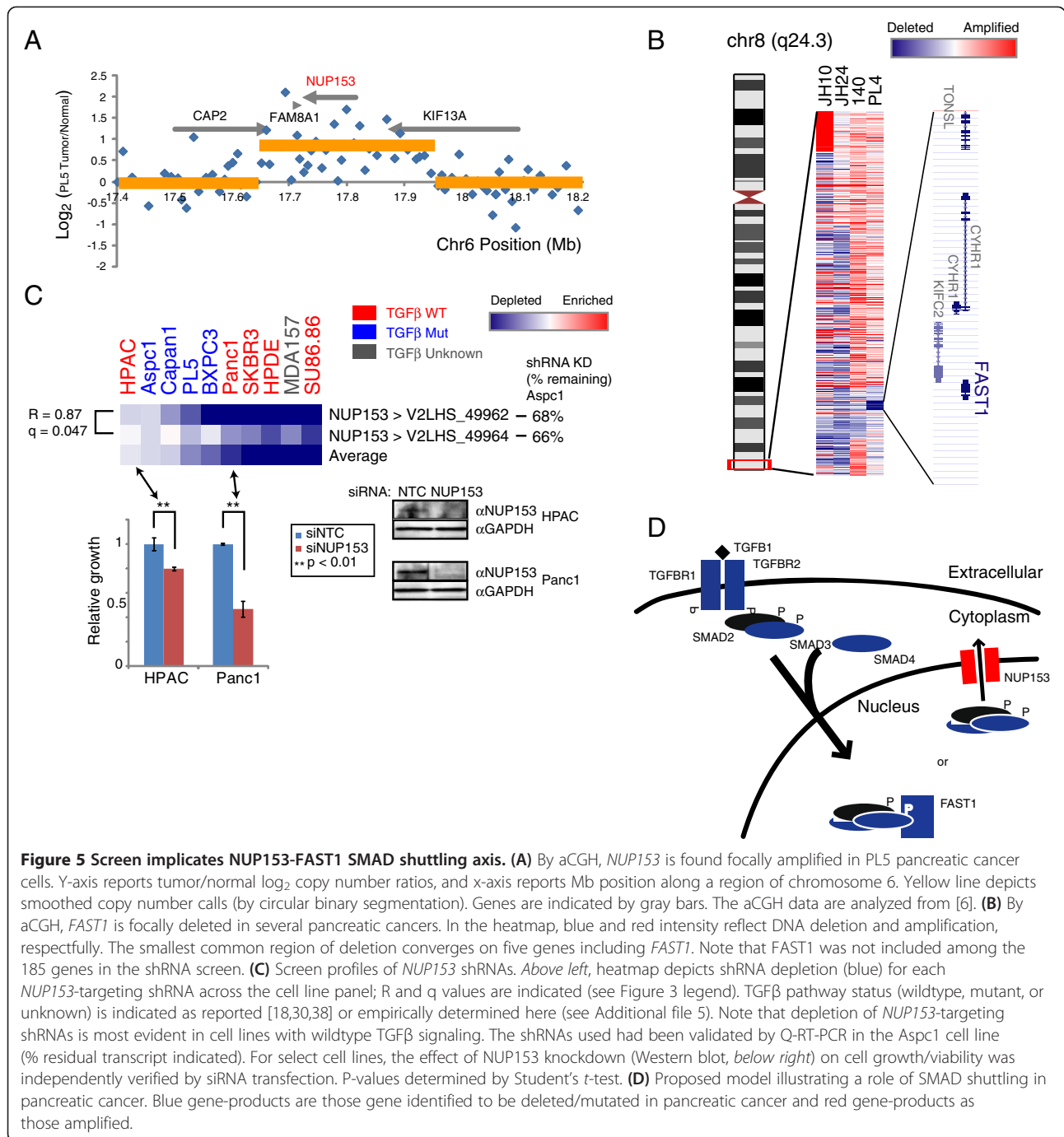
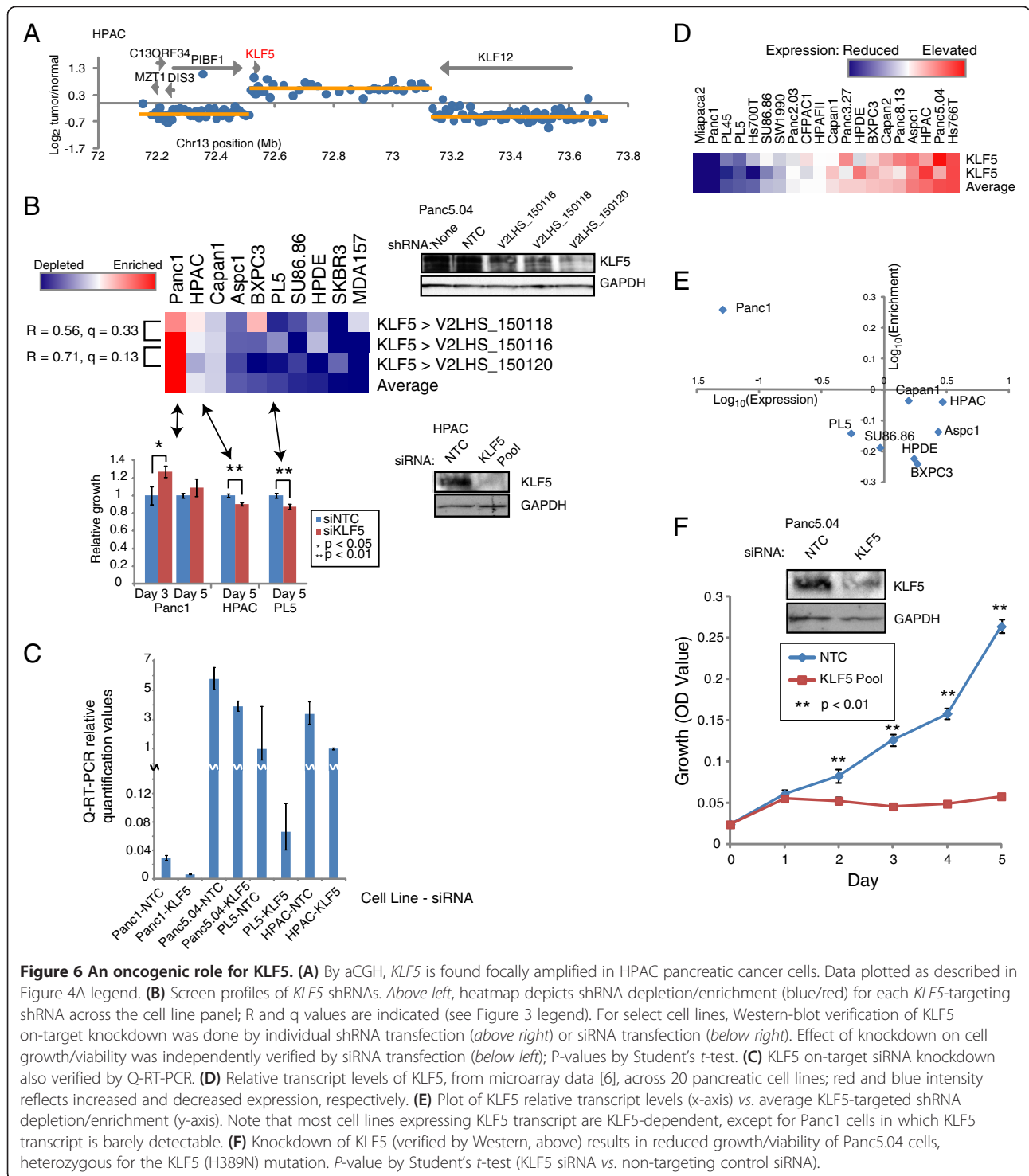


Figure 5 Screen implicates NUP153-FAST1 SMAD shuttling axis. (A) By aCGH, *NUP153* is found focally amplified in PL5 pancreatic cancer cells. Y-axis reports tumor/normal log₂ copy number ratios, and x-axis reports Mb position along a region of chromosome 6. Yellow line depicts smoothed copy number calls (by circular binary segmentation). Genes are indicated by gray bars. The aCGH data are analyzed from [6]. **(B)** By aCGH, *FAST1* is focally deleted in several pancreatic cancers. In the heatmap, blue and red intensity reflect DNA deletion and amplification, respectively. The smallest common region of deletion converges on five genes including *FAST1*. Note that *FAST1* was not included among the 185 genes in the shRNA screen. **(C)** Screen profiles of *NUP153* shRNAs. Above left, heatmap depicts shRNA depletion (blue) for each *NUP153*-targeting shRNA across the cell line panel; R and q values are indicated (see Figure 3 legend). TGFβ pathway status (wildtype, mutant, or unknown) is indicated as reported [18,30,38] or empirically determined here (see Additional file 5). Note that depletion of *NUP153*-targeting shRNAs is most evident in cell lines with wildtype TGFβ signaling. The shRNAs used had been validated by Q-RT-PCR in the Aspc1 cell line (% residual transcript indicated). For select cell lines, the effect of *NUP153* knockdown (Western blot, below right) on cell growth/viability was independently verified by siRNA transfection. P-values determined by Student's *t*-test. **(D)** Proposed model illustrating a role of SMAD shuttling in pancreatic cancer. Blue gene-products are those gene identified to be deleted/mutated in pancreatic cancer and red gene-products as those amplified.

functionally inactivating, suggesting a likely tumor suppressive role. Although Panc5.04 was not included in our screening panel, we sought to further characterize *KLF5* in that line. By analysis of microarray data [6], *KLF5* showed relatively high transcript levels in Panc5.04 compared to other pancreatic cancer cell lines (Figure 6D). Further, from our prior transcriptome sequencing (RNA-seq) data [6], 64% of *KLF5* reads in Panc5.04 mapped to the wildtype allele (78 WT reads vs. 45 mutant reads), thus

excluding epigenetic silencing of the wildtype allele. Notably, siRNA mediated knockdown of *KLF5* in Panc5.04 cells resulted in marked growth inhibition (Figure 6F). This finding suggests that the single *KLF5* mutation in Panc5.04 is most likely a passenger mutation, and provides additional support for *KLF5* being predominantly oncogenic in pancreatic cancer. Future studies should define the transcriptional targets and mechanisms underlying *KLF5* dependency in pancreatic cancer.



Conclusions

In summary, we have detailed a proof-of-principle approach for the highly-parallel functional evaluation of candidate cancer genes, here for pancreatic cancer. We have simultaneously evaluated 185 candidate pancreatic cancer

genes, selected as those recurrently and focally amplified or deleted, by a pooled shRNA library screen on 10 genetically-diverse cell lines. We have also described a novel approach, COPOTE, to enrich for on-target shRNAs and knockdown phenotypes, based on identifying shRNAs

targeting the same gene that exhibit correlated phenotype. Our screen has uncovered novel pancreatic cancer genes and pathways, most notably highlighting potential roles of a putative NUP153-FAST1 SMAD shuttling axis controlling TGF β signaling and an oncogenic function of the KLF5 transcription factor, both meriting further study. Future screens might include more shRNAs per gene (plausibly decreasing the likelihood of false negatives), more cell lines, and additional phenotypic assays. Nonetheless, our study here supports the general feasibility of a highly-parallel functional analysis of candidate cancer genes, addressing a fundamental bottleneck in the annotation of cancer genomes.

Note added in proof

During review of our manuscript, Shao et al. [44] described an RNAi-screen analysis approach (ATARiS) that is conceptually similar to ours (COPOTE).

Additional files

Additional file 1: Genes included in the shRNA screen.

Additional file 2: Depletion/enrichment (\log_{10} ratio) of each shRNA for each cell line (cell line replicates averaged).

Additional file 3: Depletion/enrichment (\log_{10} ratio) of each shRNA for each cell line replicate screened.

Additional file 4: Correlation coefficients (R) and false discovery rates (q) for all hairpins targeting the same gene.

Additional file 5: TGF β pathway status, empirically determined by TGF β -responsive growth inhibition, in pancreatic cancer cell lines (A) PLS (unresponsive/mutant pathway) and (B) HPAC (responsive/wildtype pathway).

Abbreviations

aCGH: Array-based comparative genomic hybridization; COPOTE: COrelated phenotypes for on-target effects; FDR: False discovery rate; KD: Knockdown; PCR: Polymerase chain reaction; RNAi: RNA interference; RT-PCR: Reverse transcription PCR; Q-RT-PCR: Quantitative RT-PCR; shRNA: Short hairpin RNA; siRNA: Short interfering RNA.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AHS and JRP conceived the study, designed experiments, and wrote the manuscript. AHS and KS developed and optimized the screen. AHS performed the screen and all subsequent analysis and follow up experiments. CPG characterized the genetic context of various cell lines. All authors read and approved the manuscript.

Acknowledgements

We wish to thank members of the Pollack lab for helpful discussions. This study was funded in part by grants from the NCI: R01CA112016 (J.R.P.), and from the Stanford Digestive Disease Center (J.R.P.). A.H.S. was supported by fellowships from the NSF, Stanford Graduate Fellowship program, and Cancer Biology program. K.S. was supported by fellowships from the Stanford Medical Scientist Training Program and the Paul & Daisy Soros Foundation. C.P.G. was supported by a fellowship from the Stanford Genome Training Program.

Author details

¹Departments of Pathology, Stanford University School of Medicine, 269 Campus Drive, CCSR-3245A, Stanford, CA 94305-5176, USA. ²Department of Genetics, Stanford University School of Medicine, 269 Campus Drive, CCSR-3245A, Stanford, CA 94305-5176, USA.

Received: 19 December 2012 Accepted: 12 September 2013

Published: 16 September 2013

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ: Cancer statistics, 2009. *CA Cancer J Clin* 2009, **59**(4):225–249.
2. Parkin DM, Bray FI, Devesa SS: Cancer burden in the year 2000. The global picture. *Eur J Cancer* 2001, **37**(8):S4–S66.
3. Hidalgo M: Pancreatic cancer. *N Engl J Med* 2010, **362**(17):1605–1617.
4. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA: Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev* 2006, **20**(10):1218–1249.
5. Koorstra JB, Hustinx SR, Offerhaus GJ, Maitra A: Pancreatic carcinogenesis. *Pancreatol* 2008, **8**(2):110–125.
6. Shain AH, Giacomini CP, Matsukuma K, Karikari CA, Bashyam MD, Hidalgo M, Maitra A, Pollack JR: Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc Natl Acad Sci USA* 2012, **109**(5):E252–E259.
7. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al: Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008, **321**(5897):1801–1806.
8. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, et al: Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 2012, **491**(7424):399–405.
9. Eifert C, Powers RS: From cancer genomes to oncogenic drivers, tumour dependencies and therapeutic targets. *Nat Rev Cancer* 2012, **12**(8):572–578.
10. Hannon GJ, Rossi JJ: Unlocking the potential of the human genome with RNA interference. *Nature* 2004, **431**(7006):371–378.
11. Chin L, Hahn WC, Getz G, Meyerson M: Making sense of cancer genomic data. *Genes Dev* 2011, **25**(6):534–555.
12. Giroux V, Iovanna J, Dagorn JC: Probing the human kinome for kinases involved in pancreatic cancer cell survival and gemcitabine resistance. *Faseb J* 2006, **20**(12):1982–1991.
13. Azorsa DO, Gonzales IM, Basu GD, Choudhary A, Arora S, Bisanz KM, Kiefer JA, Henderson MC, Trent JM, Von Hoff DD, et al: Synthetic lethal RNAi screening identifies sensitizing targets for gemcitabine therapy in pancreatic cancer. *J Transl Med* 2009, **7**:43.
14. Henderson MC, Gonzales IM, Arora S, Choudhary A, Trent JM, Von Hoff DD, Mousset S, Azorsa DO: High-throughput RNAi screening identifies a role for TNK1 in growth and survival of pancreatic cancer cells. *Mol Cancer Res* 2011, **9**(6):724–732.
15. Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA, East A, Ali LD, Lizotte PH, Wong TC, et al: Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci USA* 2011, **108**(30):12372–12377.
16. Xie L, Kassner M, Munoz RM, Que QQ, Kiefer J, Zhao Y, Mousset S, Yin HH, Von Hoff DD, Han H: Kinome-wide siRNA screening identifies molecular targets mediating the sensitivity of pancreatic cancer cells to Aurora kinase inhibitors. *Biochem Pharmacol* 2012, **83**(4):452–461.
17. Marcotte R, Brown KR, Suarez F, Sayad A, Karamboulas K, Krzyzanowski PM, Sircoulomb F, Medrano M, Fedysyn Y, Koh JLY, et al: Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discovery* 2012, **2**(2):172–189.
18. Ouyang H, Mou L, Luk C, Liu N, Karaskova J, Squire J, Tsao MS: Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am J Pathol* 2000, **157**(5):1623–1631.
19. Schlachet MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, Zhao Z, Smogorzewska A, Sowa ME, Ang XL, et al: Cancer proliferation gene discovery through functional genomics. *Science* 2008, **319**(5863):620–624.
20. Sims D, Mendes-Pereira AM, Frankum J, Burgess D, Cerone MA, Lombardelli C, Mitsopoulos C, Hakas J, Murugaesu N, Isacke CM, et al: High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. *Genome Biol* 2011, **12**(10):R104.

21. Strezoska Z, Licon A, Haimes J, Spayd KJ, Patel KM, Sullivan K, Jastrzebski K, Simpson KJ, Leake D, van Brabant SA, et al: **Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens.** *PLoS One* 2012, **7**(8):e42341.
22. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**(7239):719–724.
23. Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al: **A probability-based approach for the analysis of large-scale RNAi screens.** *Nat Methods* 2007, **4**(10):847–849.
24. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhim R, Weir BA, et al: **Highly parallel identification of essential genes in cancer cells.** *Proc Natl Acad Sci USA* 2008, **105**(51):20380–20385.
25. *Cancer Genome Anatomy Project - shRNA validation project.* <http://cgap.nci.nih.gov/RNAi/shRNAValidation>.
26. Eskiocak U, Kim SB, Ly P, Roig AI, Biglione S, Komurov K, Cornelius C, Wright WE, White MA, Shay JW: **Functional parsing of driver mutations in the colorectal cancer genome reveals numerous suppressors of anchorage-independent growth.** *Cancer Res* 2011, **71**(13):4359–4365.
27. Jain M, Arvanitis C, Chu K, Dewey W, Leonhardt E, Trinh M, Sundberg CD, Bishop JM, Felsner DW: **Sustained loss of a neoplastic phenotype by brief inactivation of MYC.** *Science* 2002, **297**(5578):102–104.
28. Giacomini CP, Sun S, Varma S, Shain AH, Giacomini MM, Balagtas J, Sweeney RT, Lai E, Del Vecchio CA, Forster AD, et al: **Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types.** *PLoS Genet* 2013, **9**(4):e1003464.
29. Jin K, Park S, Ewton DZ, Friedman E: **The survival kinase Mirk/Dyrk1B is a downstream effector of oncogenic K-ras in pancreatic cancer.** *Cancer Res* 2007, **67**(15):7247–7255.
30. Moore PS, Sipos B, Orlandini S, Sorio C, Real FX, Lemoine NR, Gress T, Bassi C, Kloppel G, Kalthoff H, et al: **Genetic profile of 22 pancreatic carcinoma cell lines. Analysis of K-ras, p53, p16 and DPC4/Smad4.** *Virchows Arch* 2001, **439**(6):798–802.
31. Rejiba S, Wack S, Aprahamian M, Hajri A: **K-ras oncogene silencing strategy reduces tumor growth and enhances gemcitabine chemotherapy efficacy for pancreatic cancer treatment.** *Cancer Sci* 2007, **98**(7):1128–1136.
32. Wang Z, Kong D, Banerjee S, Li Y, Adsay NV, Abbruzzese J, Sarkar FH: **Down-regulation of platelet-derived growth factor-D inhibits cell growth and angiogenesis through inactivation of Notch-1 and nuclear factor-kappaB signaling.** *Cancer Res* 2007, **67**(23):11377–11385.
33. Kwei KA, Shain AH, Bair R, Montgomery K, Karikari CA, van de Rijn M, Hidalgo M, Maitra A, Bashyam MD, Pollack JR: **SMURF1 amplification promotes invasiveness in pancreatic cancer.** *PLoS One* 2011, **6**(8):e23924.
34. Natrajan R, Mackay A, Wilkerson PM, Lambros MB, Wetterskog D, Arnedos M, Shiu KK, Geyer FC, Langerod A, Kreike B, et al: **Functional characterisation of the 19q12 amplicon in grade 3 breast cancers.** *Breast Cancer Res* 2012, **14**(2):R53.
35. Scaltriti M, Eichhorn PJ, Cortes J, Prudkin L, Aura C, Jimenez J, Chandarlapaty S, Serra V, Prat A, Ibrahim YH, et al: **Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients.** *Proc Natl Acad Sci U S A* 2011, **108**(9):3761–3766.
36. Kuuselo R, Savinainen K, Azorsa DO, Basu GD, Karhu R, Tuzmen S, Mousset S, Kallioniemi A: **Intersex-like (IXL) is a cell survival regulator in pancreatic cancer with 19q13 amplification.** *Cancer Res* 2007, **67**(5):1943–1949.
37. Xu L, Kang Y, Col S, Massague J: **Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus.** *Mol Cell* 2002, **10**(2):271–282.
38. Sun C, Yamato T, Furukawa T, Ohnishi Y, Kijima H, Horii A: **Characterization of the mutations of the K-ras, p53, p16, and SMAD4 genes in 15 human pancreatic cancer cell lines.** *Oncol Rep* 2001, **8**(1):89–92.
39. Chen C, Bhalala HV, Qiao H, Dong JT: **A possible tumor suppressor role of the KLF5 transcription factor in human breast cancer.** *Oncogene* 2002, **21**(43):6567–6572.
40. Chen C, Bhalala HV, Vessella RL, Dong JT: **KLF5 is frequently deleted and down-regulated but rarely mutated in prostate cancer.** *Prostate* 2003, **55**(2):81–88.
41. Nandan MO, McConnell BB, Ghaleb AM, Bialkowska AB, Sheng H, Shao J, Babbitt BA, Robine S, Yang VW: **Kruppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis.** *Gastroenterology* 2008, **134**(1):120–130.
42. Nandan MO, Yoon HS, Zhao W, Ouko LA, Chanchevalap S, Yang VW: **Kruppel-like factor 5 mediates the transforming activity of oncogenic H-Ras.** *Oncogene* 2004, **23**(19):3404–3413.
43. McConnell BB, Ghaleb AM, Nandan MO, Yang VW: **The diverse functions of Kruppel-like factors 4 and 5 in epithelial biology and pathobiology.** *Bioessays* 2007, **29**(6):549–557.
44. Shao DD, Tsherniak A, Gopal S, Weir BA, Tamayo P, Stransky N, Schumacher SE, Zack TI, Beroukhim R, Garraway LA, et al: **ATARIS: computational quantification of gene suppression phenotypes from multisample RNAi screens.** *Genome Res* 2013, **23**(4):665–678.

doi:10.1186/1471-2164-14-624

Cite this article as: Shain et al.: Integrative genomic and functional profiling of the pancreatic cancer genome. *BMC Genomics* 2013 **14**:624.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

