PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: Med/04

DISSECTION OF THE MOLECULAR PATHWAYS INVOLVED IN PANCREATIC CANCER INITIATION AN PROGRESSION WITH A NOVEL IN VIVO APPROACH

Alessandro Carugo

IEO, Milan

Matricola n. R09845

Supervisor: Dr. Luisa Lanfrancone

IEO, Milan

Added Supervisor: Prof. Dr. Giulio Draetta

MD Anderson Cancer Center, Houston

Anno accademico 2014-2015

"Questo lavoro è dedicato a Claudia e Angelo:
Lei lo ha reso un poema cavalleresco,
anche se non si parla d'amore;
Lui ne ha fatto un un romanzo di formazione,
si parla della mia"

"*Do not go where the path may lead,*
go instead where there is no path
and leave a trail"

Ralph Waldo Emerson

Outline

Abstract

We have seen great advances in our knowledge of the genetic regulation of various cancers in recent years, thanks in large part to large-scale genome sequencing efforts. As we catalogue and characterize the genomic aberrations associated with cancers with increasing detail and accuracy, we are faced with the challenge of having to cull bystanders from biologically active drivers and establish relevant disease context in which these drivers are rate-limiting. To address this challenge, we have adapted a loss-of-function screening approach to function in the context of an intact tumor microenvironment using patient-derived xenografts that more faithfully recapitulate the human disease compared to established cell lines. Due to the relevant genetic heterogeneity between human tumors with the same clinico-pathological indications, we have integrated independent screening approaches in a flexible platform for the interrogation of patient-derived samples as well as genetically-defined mouse models in exactly the same experimental conditions. The goal of this platform is to identify context-specific genetic vulnerabilities and translate these findings into drug discovery opportunities. As proof of concept for this approach, we describe the development of an *in vivo* loss-of-function screen to systematically interrogate epigenetic dependencies in pancreatic ductal adenocarcinoma (PDAC). In addition to the well-known genetic alterations (Kras, TP53, CDKN2A/p16, SMAD4), some epigenetic mechanisms demonstrated to play a central role in PDAC evolution and progression. The screening system utilizes tumor cells isolated from low-passage PDAC xenografted tissue and a lentiviral library of pooled shRNAs targeting 236 potentially ”druggable” epigenetic regulators. The custom-designed shRNA library (10 shRNAs per gene) was engineered with unique molecular barcodes that allow quantitation of each clone by massively parallel sequencing. Hairpins are clustered according to their depletion or enrichment in comparison

to a control population before transplantation. To date, we have completed a total of 5 *in vivo* screens using diverse PDAC target cell models that have informed on novel epigenetic dependencies. So far, the main limitation for the systematic exploitation of *in vivo* loss-of-function screens to identify specific patient vulnerabilities come from the limited number of human cells contributing to tumor establishment in a transplantation setting. The frequency of these tumors initiating cells (TICs) is commonly estimated by time-consuming limiting dilution assays and may consistently vary between different tumor origins. With this in mind, we have integrated in our platform a system based on scrambled barcoded libraries that allows to directly assess the required coverage of screening libraries in each model and adjust the shRNA screens for this factor. Our coverage study demonstrated to be a powerful tool to identify the minimal number of cells/barcode required to sustain a complex library in transplantation assay and at the same time a step forward to personalize the *in vivo* screening patient by patient. We optimized a comprehensive data analytics pipeline and developed a high-throughput validation scheme to triage "hits" that emerge from each screen. The most potent "hits" have been enrolled in both functional and clinico-pathological validation studies to determine the highest priority targets for this devastating disease. Significantly, different components of the COMPASS histone H3 Lys4 (H3K4) methyltransferase complexes were identified as candidates in our screens. COMPASS and COMPASS-like complexes are characterized by unique subunits composition, whose identities provide insight into the different biological functions of these complexes. The methyltransferase unit of the COMPASS complexes is directly involved into the methylation of Lys4 on histone H3, a commonly accepted sign of open-chromatin and active transcription. Chromosomal translocations involving MLL gene are frequent events characterizing the Mixed Lineage Leukemia. In this disease, it has been shown that fusion events with a variety of different partners compromise the MLL methyltransferase activity. However, multiple members of the MLL family could be deregulated via different oncogenic mechanisms in PDAC, as the genetic alteration in MLL2 (amplification) and MLL3 (mutation)

suggested. Our platform represents an ideal starting point to understand the COMPASS functionalities. So, a deeper understanding of genes and pathways regulated by each MLL subunit in the context of PDAC is critical to better elucidate the molecular dynamics of this disease and identify additional key points of vulnerability. Our study identified the core different subunits of the COMPASS complexes (WDR5-ASH2L-RBBP5) as broad relevant players in sustain PDAC progression, while the dependency on the MLL subunits appears to be more context-dependent and potentially consequent to specific genetic alterations. Mechanistically, WDR5 functions to sustain proper execution of DNA replication in PDAC cells, as previously suggested by replication stress studies involving MLL1, a critical ATR substrate, and c-Myc, also found to interact with WDR5. By showing that ATR inhibition mimicked the effects of WDR5 suppression, we open up the possibility of testing inhibitors currently in development for activity in this disease. These findings are proposing a new layer of complexities in trapping the COMPASS complexes during tumor development and unmasking unexplored directions for new therapeutical opportunities.

Acronyms and Abbreviations

4-OHT = 4-Hydroxytamoxifen;
ACA = Anti-centromere antibodies;
Alb = Albumin;
BrdU = 5-Bromo-2-Deoxyuridine;
Cag = Gag promoter;
Cas9 = CRISPR-associated protein 9;
CDF = Cumulative distribution function;
CDKi = CDK inhibitors;
CFA = Colony formation assay;
CRISPR = Clustered regularly interspaced short palindromic repeats;
CSC = Cancer stem cell;
DAPI = 4',6-diamidin-2-fenilindolo;
DOX = doxycycline;
dsRNA = double strand RNA;
ELDA = Extreme limiting dilution analysis;
FC = Fold-change;
FDR = False discovery rate;
Flpo = Flippase;
Gag = Group-specific antigen;
GEMM = Genetic engineered mouse model;
GFP = Green fluorescent protein;
GSC = Glioblastoma stem cell;
GSEA = Gene set enrichment analysis;
H3K27 = Histone 3 Lysine 27;
H3K4 = Histone 3 Lysine 4;
H3K4me3 = Trimethylation of the Histone 3 Lysine 4;
HMT = Histone methyltransferase;
HU = Hydroxyurea;
IC = Initiation complex;
ICGC = International cancer genome consortium;
IEG = Immediate early gene;
IRES = Internal ribosome entry site;
KINK = $Kras^{G12D\_LSL/+}$; $Ink4a^{L/L}$ background;
KP53 = $Kras^{G12D\_LSL/+}$; $Trp53^{L/L}$ background;
Krt19 = Cytokeratin 19 promoter;
LOF = Loss-of-function;
LSM = Lentiviral-based somatic mosaic system;
miRNA = micro RNA;
MOI = Multiplicity of infection;
MRI = Magnetic resonance imaging;
mRNA = messenger RNA;
MT = Methyltransferase;
NGS = Next-generation sequencing;
NlsCre = Cre recombinase gene;
NM = Number of molecules;
NSG = NOD/SCID gamma;
NT = Non targeting;
ORF = open reading frame;
OTE = Off-target effect;
PCC = Premature chromosome condensation;
PCR = Polymerase chain reaction;
PDAC = Pancreatic ductal adenocarcinoma;
PDX = Patient-derived xenograft;
$Puro^{R}$ = Puromycin resistance cassette;
R26 = Rosa26 gene locus;
RC = replication complex;

Ref Cells = Reference cell population in studies with barcoded libraries;
RFP = Red fluorescent protein;
RISC = RNA-induced silencing complex;
RNAi = RNA interference;
SD = Standard deviation;
shRNA = short hairpin RNA;
siRNA = small interfering RNA;
stRNA = small temporal RNA;
Tet = Tetracycline;
TF = Transcription factor;
TIC = Tumor-initiating cell;
TMA = Tissue microarray;
TPC = Tumor perpetuating cell;
TSG = Tumor suppressor gene;
TSS = Transcriptional start site;
TU = Transducing units;
Tx = Tumor;
UTR = untranslated region;
Veh = Vehicle;
WRAD = WDR5-RBBP5-ASH2L-DPY30 complex;
WT = Wild-type;
Xenograft I = Primary xenograft;
Xenograft II = Secondary xenograft;
γH2AX = Phospho-H2AX.

Introduction

1. Pancreatic cancer

Pancreatic ductal adenocarcinoma (PDAC) is the most common cancer of the pancreas and contributes to 6.9% of all cancer deaths in the US, with an estimated 48,960 new cases and 40,560 deaths in 2015 alone. Present estimates predict that >1.5% of the US population will be diagnosed with PDAC in their lifetime (NCI/SEER 2015). Moreover, there has been an approximate 14% increase in PDAC incidence compared to 5 years ago (estimation based on SEER 2015 data). In keeping with this upward trend, it is projected that PDAC will be the second leading cause of cancer death in the next 15 years, trailing only lung cancer[1]. Unfortunately, the overall 5-year survival rate for PDAC patients is a dreadful ~6%. Major risk factors for PDAC include smoking, diabetes, chronic pancreatitis as well as family history. Early PDAC is often asymptomatic and there are no serological biomarkers in the clinic that are specific for detecting precursor lesions in the general population, although promising biomarkers are emerging.[2] It is estimated that it may take ~17 years for a single tumor-initiating cell to develop into metastases, suggesting that there should be an emphasis on biomarker discovery for early diagnosis[3].

The standard of care for advanced PDAC has been either gemcitabine or its combination with other agents (i.e. nab-paclitaxel, erlotinib), which only provides very limited benefit[4-7]. FOLFIRINOX was approved recently for advanced PDAC and results in real responses, but only in a subset of patients[8]. Approximately half of patients who progress on one of the front-line regimens receive second-line therapy. For patients who progressed on gemcitabine-based therapy, a fluropyrimidine plus oxaliplatin regimen is standard, whereas a gemcitabine-based regimen is often chosen for patients who progressed on FOLFIRINOX[9,10]. Even with aggressive treatment, median survival remains less than 1 year. No standard of care treatment has been identified for patients who progress after second-line treatment[7,8].

Chemotherapy has proven disappointing in the treatment of PDAC, and a multi-pronged approach is being taken to identify new, more effective therapeutics. This includes more potent chemotherapeutic drugs with improved delivery for solid tumors, as well as targeted therapies. Currently, the only approved targeted therapy for PDAC is the EGFR inhibitor erlotinib, which showed a modest survival benefit in combination with gemcitabine compared to gemcitabine alone[11]. Given that up to 90% of PDAC are estimated to have overexpression of EGFR, drugs targeting EGFR continue to be investigated for this disease[12,13].

The genomic landscape of PDAC is dominated by oncogenic mutation of *KRAS*, present in up to 90% or more of cases[14,15]. Ample functional evidence from animal models suggests that mutant Kras is critical in both PDAC initiation and in maintenance[16,17]. The characteristic activating point mutation in this small GTPase has been profoundly challenging to inhibit from a pharmacological perspective, and drugs intended to impede membrane localization and activation by blocking prenylation have also failed[18,19,20]. Activating mutation in KRAS results in constitutively active RAF/MEK/ERK and PI3K/AKT signaling, which drives uncontrolled cell growth. As such, pharmacological targeting of these pathways is of significant interest. So far, no single agent has yielded a clinical benefit, including the MEK1/2 inhibitor selumetinib, which is now approved for the treatment of metastatic melanoma. Based on clinical trial results, it is widely held that simultaneous targeting of multiple signaling cascades will be required to effectively block the KRAS signaling network. Unfortunately, cumulative drug toxicities may diminish the clinical utility of such an approach[21] .

*A*ctivating KRAS mutations is an early event in PDAC pathogenesis that is present in pancreatic intraepithelial neoplasias (PanINs), the precursor lesion for PDAC[15.] (**Fig.1**) Additional genetic inactivation of the tumor suppressors, CDKN2A, TP53, and SMAD4 (each present in >50% of PDAC) is associated with transition from PanINs to PDAC. Moreover, the presence of Smad4 and p53 mutations in primary tumors correlates with increased

propensity for metastatic dissemination. The pioneering development of a conditional mutant Kras allele by Jacks and Tuveson enabled the first studies that demonstrated that mutant Kras drives PanIN formation in association with local inflammation[21,22]. Moreover, it has been shown that inactivation of CDKN2A, P53, and SMAD4 synergizes with mutant Kras to promote PDAC progression[22-25].

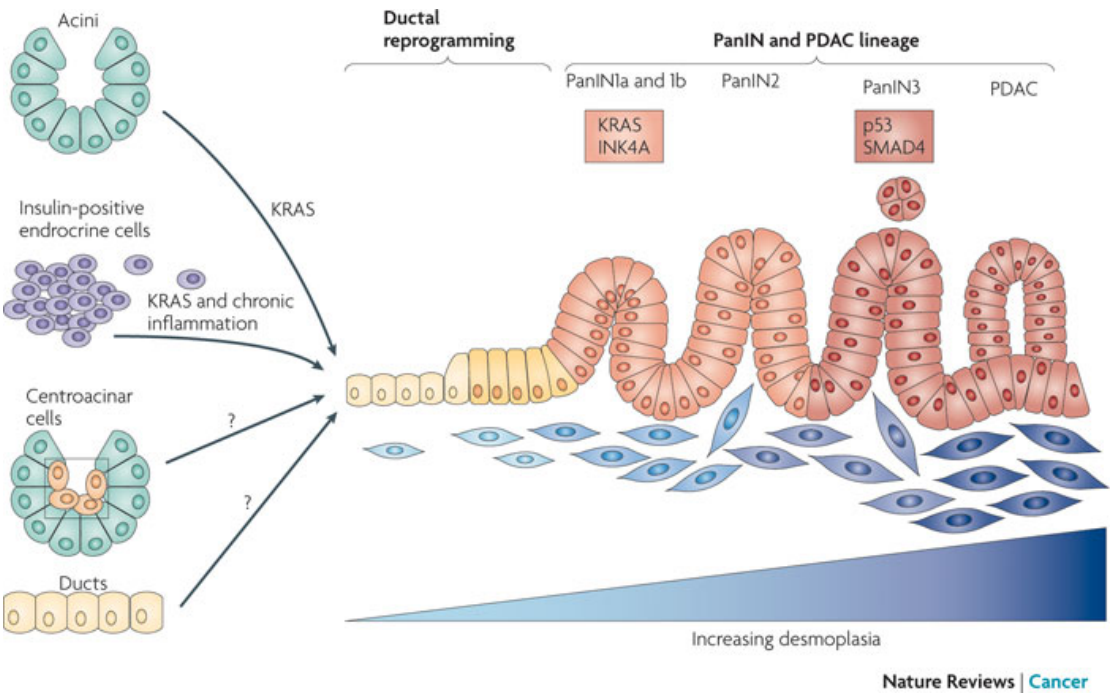


Figure 1. KRAS is a master regulator of pancreatic ductal adenocarcinoma initiation and progression. Constitutively active KRAS (caused by $Kras^{G12D}$ or $Kras^{G12V}$ mutations) is sufficient to initiate the development of pancreatic intraepithelial neoplasia (PanIN) and pancreatic ductal adenocarcinoma (PDAC). PanINs are classified into three stages of increasing cellular atypia and, in humans, have been found to possess increasing numbers of mutations (common mutations are indicated in boxes). Changes in the epithelium are matched by desmoplastic changes in the stroma. In mouse models, the human PanIN spectrum followed by progression to PDAC has been recapitulated by activating mutant KRAS in embryonic pancreatic progenitors. Eliminating tumour suppressors commonly inactivated in the human disease dramatically decreases PDAC latency (a limited set of examples is indicated). Mouse models in which KRAS is activated specifically in some adult cell types have shown that both acini and insulin-positive cells can give rise to PanINs and, in some cases, PDAC depending on tissue damage and tumour suppressor inactivation. For these cell types, reprogramming into a

A series of cancer genomics studies has expanded the PDAC mutation landscape, identifying numerous less frequent genetic lesions. It is now clear that, in addition to the signature set of mutations described above, PDAC tumors harbor recurrent genetic alterations in chromatin regulators (e.g. ARID1A, ARID1B, KDM6A, MLL3), WNT pathway components (RNF43), DNA damage genes (ATM, BRCA2), and other signaling pathways (e.g. MAP2K4)[26,27,28]. However, none of these alterations are present in more than ~15% of PDAC.

In addition, studies of PDAC patients at autopsy have revealed that there is extensive intratumoral genomic heterogeneity with only KRAS mutation being consistently detected across different tumor clones. This strikingly high intratumoral heterogeneity has important implications for the deployment of targeted therapeutics against these newly identified pathways. Thus, agents targeting the founder mutations in PDAC (i.e. KRAS) may be predicted to have the highest potential for clinical impact. However, this does not preclude the possibility of targeting lower frequency mutations that may render subclasses of PDAC cells sensitive to particular therapies. For example, approximately 5% of PDAC harbor bi-allelic mutations of *BRCA2*, *PALB2* or other genes in the Fanconi anemia DNA repair pathway, and these tumors are highly susceptible to double-strand break-inducing drugs, such as platinum compounds[29,30,31]. As noted above, parsing the genome of PDAC from emerging large scale datasets like TCGA and ICGC reveals the existence of numerous such potential vulnerability nodes in PDAC, the vast majority of which remain untested and unexploited from a therapeutic perspective.

2. Epigenetics and pancreatic cancer

All of the cells in an organism carry the same genome, but it is the regulated expression of specific genes that allows cells to achieve diverse, stable phenotypes. The term "epigenetics" describes heritable molecular modifications within cells that maintain the required information to preserve a cellular phenotype across generations. A number of different changes are used by cells to "remember" their transcriptional programming. These include direct modifications of DNA, such as methylation, hydroxylation, formylation, and carboxylation; variations in nucleosome occupancy and positioning; histone variants; and histone post-translational modifications, including methylation, acetylation, phosphorylation, ubiquitylation, sumoylation, and at least 11 other modifications[32,33].

While cancer has historically been considered a disease driven by genomic alterations, it is now clear that epigenetic modifications of chromatin also play a key role in tumor pathogenesis. Several lines of evidence underscore the importance of epigenetic control in tumors. First, epigenetic silencing of tumor suppressor genes by promoter CpG island hypermethylation has been documented and demonstrated to be mutually exclusive from mutational inactivation of the same gene[34,35,36]. Second, epigenetic writers and readers have both been shown to be required for tumor development in various mouse models[37,38,39]. Third, evidence exists that tumor cells may acquire oncogene addiction to epigenetic alterations, particularly DNA methylation[40]. Finally, whole-genome sequencing of a number of different cancers has catalogued recurrent somatic mutations in a vast array of epigenetic regulators[41].

The COMPASS complex (complex of proteins associated with Set1) was first identified in yeast as a complex with specific methyltransferase (MT) activity for lysine 4 of histone 3 (H3K4)[42]. Subsequently, six homologous complexes, including 2 COMPASS and 4 COMPASS-like complexes, were identified in humans that shared core components but form complexes with different histone methyltransferases (HMTs; Set1A, Set1B, MLL1, MLL2, MLL3, or

MLL4). All six HMTs have H3K4 methylation activity in common due to a conserved SET domain, resulting in "activating" marks on chromatin at actively transcribed regions[43]. Despite this common function, the complexes, as well as the subunits within the complexes, have a range of both redundant and non-redundant roles, including methylation-independent roles, in cell development and differentiation[44].

Recently, exome sequencing of PDAC patient samples by the ICGC has revealed loss-of-function somatic mutations in three members of the chromatin-modifying COMPASS-like complex family, MLL3, MLL4 and UTX, at a cumulative frequency exceeding 20% of cases and a recent work by a collaborator at MD Anderson has produced in vivo data from a mouse model of PDAC supporting tumor suppressor activity by MLL3 in the context of mutant Kras[45]. Moreover, numerous roles for COMPASS-like proteins have been detailed in the literature for many cancers, with the oncogenic effects of MLL1 in hematologic malignancies being particularly well characterized[46].

The core COMPASS-like subunits (WDR5-RBBP5-ASH2L-DPY30), also known as WRAD, complex with MLL2, MLL3, and MLL4 as well as MLL1. MLL1 and MLL2 more closely resemble Drosophila Trithorax protein and can complex with the tumor suppressor Menin[47]. Conversely, MLL3 and MLL4 bind with the H3K7 demethylase UTX (among others), and these complexes have been shown to act as tumor suppressors in both hematological malignancies and solid tumors[48,49,50]. Despite their different activities in the cell, all MLL proteins can complex with the core WRAD proteins, and all have H3K4 MT activity[51,52].

MLL and the WRAD complex have been implicated in cell cycle control and execution, also modulating their expression[53,54,55]. Most recently, it has been reported that MLL1 association with the WRAD complex is required for cells to progress through S phase, and that MLL1 is a substrate of ATR during the S-phase checkpoint[56].

3. RNA interference (RNAi) discovery and biological implications

The biological process of RNA interference (RNAi) was first discovered in *Caenorhabditis elegans* as a response to double-stranded RNA (dsRNA), which resulted in sequence-specific gene silencing. Starting from the initial findings in worms, Fire and colleagues proposed to use antisense RNA as an approach to inhibit gene expression, demostarting that the double strand DNA (dsRNA) was at least ten-fold more potent as a silencer than sense or antisense RNAs alone[57,58]. This discovery promoted the hypothesis that a number of previously characterized mechanisms for silencing gene expression could be originated by a common biological mechanism. Biochemical and genetic experiments were able to prove that RNAi, co-suppression and virus-induced gene silencing are mechanistically similar, and that RNAi gene silencing is shared by many eukaryotic organisms[59,60].

Genetic studies in *C. elegans* and plants and biochemical studies in *Drosophila* generated our present understanding of the mechanisms underlying dsRNA gene silencing. Basically, injection of dsRNA into *Drosophila* embryos induced sequence-specific post-transcriptional silencing[61]. Then, it was also demonstrated that *Drosophila* embryo extracts might also be competent for RNA interference[62]. Reduction in luciferase synthesis from a synthetic mRNA was observed in cell-free lysates upon incubation with dsRNA and suggested that dsRNA might drive the assembly of a nuclease complex that targets the homologous RNA for degradation to induce gene silencing.

This nuclease complex, now known as RISC (RNA-induced silencing complex), was extracted from *Drosophila* S2 cells in which RNAi response was simulated by treatment with dsRNA[63]. Additional studies, proving the existence of “RNA guides”, corroborated the original RNAi hypothesis that some dsRNA byproducts could drive the identification of substrates for RNAi. In this direction, antisense RNAs with homology to genes being targeted by co-suppression were identified[8]. More than that, a ~25-nucleotide RNA was discovered only in plants experiencing specific gene suppression and homologous RNA products were observed

during virus-induced gene silencing. Similar small RNAs were produced from dsRNAs and found in complex together with nuclease activity of RISC in *Drosophila* embryos[63,64,65].

A model for RNAi was proposed, in which silencing initiates upon recognition of dsRNA by machinery that converts the dsRNAs in ~21–25-nucleotide RNAs[66]. These small interfering RNAs (siRNAs) drive the identification of the homologous substrates being assembled with an effector RISC machinary. The wide biological nature of dsRNA-induced silencing suggested that the central RNAi machinery potentially modified itself to answer to specific biological requests in different organisms.

However, in all the organisms the dsRNA has to be cleaved to generate siRNAs. Experiments oriented to the discovery of the enzyme involved into the initial RNAi step highlighted a new RNase III ribonuclease family with unprecedented specificity for dsRNA. RNase III enzymes can be divided into three classes based on structural domains: i) RNase III costituted by a single catalytic domain and a dsRNA-binding domain; ii) Drosha family nucleases contain dual catalytic domains; and iii) a third family also contains dual catalytic domains and additional helicase and PAZ motifs[67,68,69]. The third family was named Dicer and was demonstrated to be evolutionarily conserved[69,70]. The tridimensional structure of the RNase III catalytic domain sustained a model in agreement with generation of ~22-nucleotide RNAs by Dicer cleavage[70].

RNAi is also constituted by RISC, the ribonuclease complex that recognizes and destroys specific target mRNAs. Initally, the siRNA, which presumably identifies substrates through base pairing, was identified as critical subunit of the RISC[71]. Cleavage is apparently endonucleolytic, and occurs only in the region homologous to the siRNA. siRNAs are double-stranded duplexes with two-nucleotide overhangs and phosphate termini, and this configuration is functionally important for incorporation into RISC complexes[72-74]. (**Fig.2**)

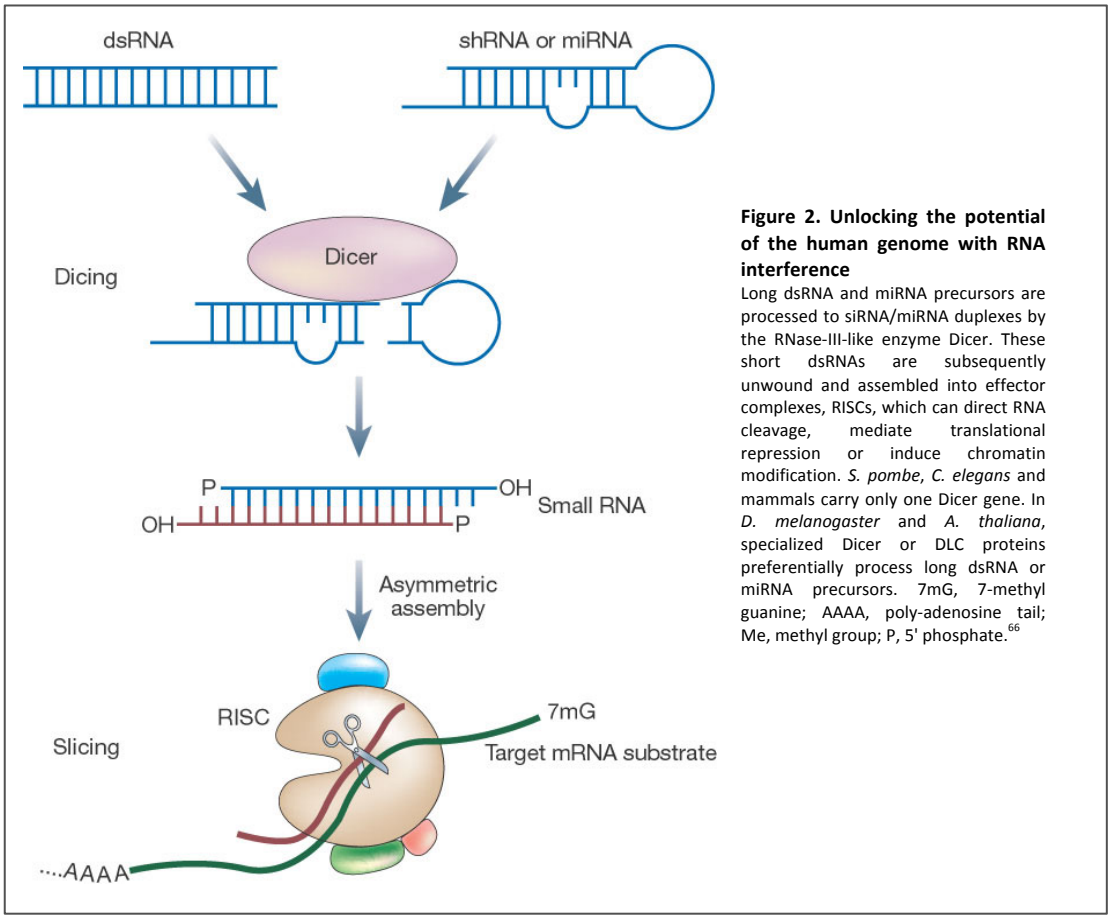


Figure 2. Unlocking the potential of the human genome with RNA interference

Long dsRNA and miRNA precursors are processed to siRNA/miRNA duplexes by the RNase-III-like enzyme Dicer. These short dsRNAs are subsequently unwound and assembled into effector complexes, RISCs, which can direct RNA cleavage, mediate translational repression or induce chromatin modification. *S. pombe*, *C. elegans* and mammals carry only one Dicer gene. In *D. melanogaster* and *A. thaliana*, specialized Dicer or DLC proteins preferentially process long dsRNA or miRNA precursors. 7mG, 7-methyl guanine; AAAA, poly-adenosine tail; Me, methyl group; P, 5' phosphate.[66]

Because base pairing interactions drive mRNA identification, the RNAi machinery can be both dynamic and specific. Thus, the RNAi machinary has been adapted to numerous cellular functions and organisms[75]. The normal gene regulation of endogenous levels could be influenced by the RNAi mechanisms, as suggested through the analysis of plants and animals containing dysfunctional RNAi components, even though this findings should not be interpreted as proof that RNAi pathways independently regulate endogenous gene expression. A possible mechanism emerged from the study of *C. elegans* containing mutations in their single Dicer gene, *DCR-1*[76]. Interestingly, Dicer mutants displayed alterations in developmental timing similar to those observed in *let-7* and *lin-4* mutants. These loci encode small RNAs, which are synthesized as precursors and post-transcriptionally processed to a ~21-nucleotide mature form. Additional studies have confirmed that these RNAs are processed by Dicer[77]. The small temporal RNAs (stRNAs) encoded by *let-7* and *lin-4* are negative regulators of specific protein-coding genes, but regulate expression at the translational level[77,78]. This generated the hypotesis that stRNAs

and dsRNA could share only the processing enzyme Dicer. However, Mello and colleagues discovered another piece of the puzzle, the requirement for Argonaute family proteins in both stRNA biogenesis and stRNA-mediated suppression. These findings opened the doors for a new model in which the effector complexes constituted by siRNAs and stRNAs are strictly related, but regulate expression by distinct mechanisms[77].

A model that emerged to define RNA control of gene expression proposed that a similar RISC complex is formed containing either siRNAs or stRNAs. Another model sustained that siRNAs and stRNAs may recognize related but distinct complexes. In agreement with this model siRNAs or exogenous hairpin RNAs fail to repress gene silencing when modified by single mismatches with their substrates. In this direction, RISC complex could be considered as a dynamic platform to structure different regulatory modules in a flexible way. The core complex should be required to recognize the small RNA cleaved by Dicer and adoperate this as a guide to search for its homologous substrate. As a consequence and in relation with the kind of signal, different downstream functions are associated with the core: in RNAi, nucleases should be joined to the RISC, whereas in stRNA-mediated regulation, translational repressors would join the complex. Inclusion of chromatin remodelling factors and other adaptations could follow the initial effectors to accomplish the transcriptional silencing[79,80].

3.1 RNAi as a solution for mammalian genetics

RNAi demostrated to be a powerful tool for investigating gene function. For a long time exploitment of RNAi to interogate mammalian systems seemed to not be feasible, at least until the first prove that the technology could work was obtained, thanks to the demonstartion of RNAi in mouse embryos[81,82,83]. However, as mammalian somatic cells, but not some embryonic cells, display nonspecific responses to dsRNA, RNAi appeared to be of limited utility. Then, Tuschl and colleagues demostrated that siRNAs themselves could be applied to cause effective silencing in many mammalian cells[84]. These small RNAs, chemically synthesized mimicking the Dicer products, are presumably assembled in the RISC complex

and target specific substrates for degradation. Fortunately, the siRNAs are too small to activate nonspecific dsRNA responses such as the RNA-dependent protein kinase (PKR)[85]. Using a variety of standard transfection methods siRNAs can be easily introduced into mammalian cells. The strength and timing of the interfering response is affected by several parameters: the silencing response is primarly influenced by the overall efficiency of transfection in the overall population. In individual cell, silencing is intrinsically related to the amount of siRNA that is nternalized and on the potency of each siRNA against its target. A weak siRNA can also be used to silence its target delivering a sufficient quantities of the siRNA in the host cell. However, numerous undesired effects could be associated with the use of large amount of reagent.

One limitation releated to the siRNA transient effects in mammals is the apparently lack of the mechanisms that sustain the silencing amplification present in worms and plants. More recently, different groups have proven that short hairpin RNAs (shRNAs) can be applied to manipulate gene expression in experimental settings[86]. The shRNAs are not 100% homogeneous in size and design, with stems ranging from 19 to 29 nucleotides in length, and with various degrees of structural similarity with natural miRNAs. These interfering RNAs can be expressed *in vivo* from RNA polymerase II (Pol II) or III (Pol III) to induce stable suppression in mammalian cells. Because these shRNAs are encoded by DNA vectors, they can be delivered to cells exploiting the different tools that have been developed for delivery of DNA exogenous constructs[87]. These include standard transient transfection systems, stable transfection and infection using viruses ranging from retroviruses to adenoviruses. Expression can also be regulated by either constitutive or inducible promoter systems.

The easy way of stable RNAi systems generation increased the request and applications beyond the utility of transient siRNAs. Specific phenotypes can now be evaluated over long time spans. Stable infected cells can be used either for *in vitro* or *in vivo* purposes, for example investigating the cooperative role of the microenvironment with the tumor cells in xenograft models. RNAi can also be applied to rapidly produce hypomorphic alleles in

transgenic mouse models. In addition, shRNAs could be coupled with existing high-efficiency gene delivery vehicles to develop RNAi-based therapeutics. At the end, the stringent specificity of RNAi technologies supported the possibility to direct the silencing effect against a mutant allele, such as an activated oncogene, without affecting the normal allele. In search for rules that could generate more effective and specific siRNAs, different groups tested large numbers of siRNAs, sorting them into classes depending on their potency, and then looking for features that characterized effective on-target siRNAs from ineffective ones[88]. siRNAs in which the helix at the 5' end of the antisense strand has a lower stability than the 3' end of the siRNA are generally more effective silencers than those with the opposite arrangement. Biochemical studies also showed unequal incorporation of the two RNA strands of the siRNA into the RISC complex. So, the the effectiveness of an RNAi response leaded by a siRNA is strongly dependent on its tridimensional structure and associated with the step of RISC assembly, in which the asymmetry in the dsRNA must be sensed and a single RNA strand must be selected for incorporation into the enzymatic complex.

3.2 RNAi as a tool for genetic screens

The success in applying RNAi for interrogate single gene-associated phenotypes has promoted numerous efforts to translate this approach on a large scale for genetic investigation. With the human, mouse and rat genomes sequenced, RNAi demostrated to be the required mechanism by which this enormous amount of genetic information can be translated into functional annotations.

3.2.1 RNAi libraries

siRNAs libraries can be constructed in fundamentally different ways, including chemical synthesis or enzymatic digestion of long dsRNAs[90,91]. On the other side, libraries can be synthesized by constructing shRNA lentiviral vectors that target each gene of interest[92].

In the same way as siRNAs, demonstration in the feasibility of the approach came from small-scale studies. Generating a library for targeting the family of de-ubiquinating enzymes, it was possible to identify a new unprecedented function of the CYLD gene in suppressing the activity of NF-κB[93]. This result unequivocally confirmed that unbiased genetic screens can generate not only new biological informations but also tangible advances for the treatment of specific disease. Several groups have produced arrayed libraries from chemically synthesized oligonucleotides that cover about 10,000 different human genes each[69,94]. Another group has expoited the polymerase chain reaction (PCR) to produce a library that encode shRNAs, and several other groups have developed new methods for constructing random shRNA libraries based on complementary DNA or genomic DNA[95-98]. Random libraries are relatively cheap to be produced and can gurantee the coverage of an individual gene with many different shRNAs. Nonetheless, these libraries have the limitation of the number of genes that can be properly represented. On the other side, libraries synthesized using chemical oligonucleotides are more expensive, but allow the introduciton of powerful informatic methods to improve the shRNA design and exhibit tolerance in rapid modification of the shRNA structure. More than that, synthetic libraries can be used either as mixtures or as individual arrays, in the same way as siRNA libraries. Independent transfection and phenotypic optimization of target cells must be carried out for large-scale screening using siRNA libraries. siRNA libraries may be used to the deep range of screening applications that were generated by the pharmaceutical companies as cell-based assays for drug discovery. These methods consist of fluorescent reporter assays for various phenotypes and screens by throught automated microscopy technologies. In the same way RNAi effectors can be synthesized as microarrays and evaluated for their functions as a consequence of *in situ* transfection[99]. shRNA libraries can be also applied in a similar way, as demonstrated by the development of arrayed library for genes that are specifically involved in functions of the proteasome[100]. Pools of shRNAs must be transduced into a selective cell line with a high representation and a low multiplicity of infection (MOI = 0.1–1.0). The high representation

gurantees that each shRNA in the library has a reasonable possibility to function, and the low MOI gurantees that the cells are transduced with only one shRNA, avoiding cooperative effects among shRNAs. RNAi screening approaches generated the possible to discover new gene functions or networks involved in a wide range of biological processes, including mechanisms critical for signal transduction, cell viability, cell or organelle morphology, organelle or protein localization and/or function, drug resistance, and responses of host cells to pathogens. shRNA libraries are available as pool from several commercial providers, and recently the DECIPHER open source RNAi screening project proposed free of charge access to lentiviral shRNA pools (http://www.decipherproject.net).

3.2.2 Off-target effects and false postives/negatives

The initial excitement about RNAi screens was diminished by the demonstration that RNAi approaches, like all screening technologies, are connected with false discovery rate (false-positive and false-negative results). In RNAi tools, the most relevant caveat is false positives that are generated by sequence-specific off-target effects (OTEs)[101,102]. In addition, meta-analyses to facilitate the estimation of false discovery rates were helped by the deveopment of RNAi screen data sets. These databases have also identified 'frequent hitters', genes that frequently score as positive hits across different screens, such as genes involved in essential processes that might exert relevant but relatively broad as demonstrated for components of the ribosome or proteasome. However, they have increased the information connected to the specificity and relevance of primary screen hits. Sequence-specific OTEs happen when RNAi reagents recognize RNAs different from their intended target due to partial complementarity. Applying sequence alignment is also possible to identify the subsets of OTEs that are associated to the extended regions of complementarity between RNAi reagents and genes other than the target, such as regions common to the target gene and its paralogues. New methodologies for RNAi re-annotation have been developed (ex. UP-TORR and GenomeRNAi) in order to facilitate the identification of RNAi triggers that no longer

meet quality standards[103,104]. To further limit OTEs in RNAi screening approaches, new and improved bioinformatic tools have recently been proposed and successfully applied.

3.2.3 RNAi screening hits validation

The experimental approach more frequently used to validate RNAi screen hits is the evaluation of multiple RNAi reagents for each gene, as different triggers have different seed regions. In siRNA screens seven or more independent RNAi reagents per gene are arrayied. In pooled shRNA screens, more than 15 constructs per gene are normally applied[105,106]. The general startegy is oriented to increase as much as possible the number of independent RNAi triggers per gene that reproduce the same phenotype, in order to maximize the chance that the gene is a true hit in a screen. However, the experimental approach for demonstrating the specificity of an RNAi reagent is to highlight that the displayed phenotype can be rescued with the expression of an RNAi-resistant version of the targeted gene. Unfortunately, this rescue experiment could not be routinely performed, especially because the interpretation of rescue experiments is complicated when rescue constructs are expressed at non-physiological levels[101]. Applying a comparative approach is also a potential strategy to overcome species-specific limitations, such as incomplete genome coverage of screening reagents or sequence-specific OTEs. The hits that emerged associated with cellular processes and mechanisms in the data set of both species have a higher chance of being true positive hits if gene ontology terms are consistently enriched. In addition, genes that score as positive hits in both species can also be considered high confidence hits, because they have been validated by different screen reagents, methodologies and organisms.

3.2.4 Positive screens

siRNAs/shRNAs that are associated with a cellular phenotype that can be selected for (e.g. increased survival, invasion or migration abilities) or isolated through can be considered for positive screens[107-112]. Using pooled libraries in positive selection approaches, the shRNA of interest can be sequenced starting from the genomic DNA isolated from the isolated cell

population. On the other way, deep sequencing can be applied to identify shRNAs that are enriched in comparison with the reference population. So far, several strategies have been proposed to identify the most promising hits sequencing the full hairpin, half of the hairpin or the molecular barcode associated with the hairpin in some libraries[113,114].

3.2.5 Negative screens

Dropout of siRNA/shRNA sequences in the experimental population compared to a control population characterize the negative screens. The phenotype associated with the shRNA/siRNA of interest could result in cell death and the shRNA/siRNA would therefore be significantly reduced or completely absent in respect of the control population[115,116]. So far, shRNA pooled negative screens can only be performed applying microarrays or deep sequencing detection strategies. Negative screens for synthetic lethality emerged from classical genetic screens in yeast performed with the purpose of identifying two mutations that result in lethality only when present in combination[117]. Synthetic lethal screens obtained in the last years lot of attention from the onco-biology, because they can be used to discover the secondary genetic events (second hits) that can selectively kill a cell with a primary oncogenic mutation. Rrecently, RNAi screening approaches have identified synthetic lethal interactions with the RAS oncogene and the KRAS oncogene[118,119]. A slightly different approach, known as chemical synthetic lethality screen, is focused on the identification of genetic elements that cooperate with the effect of a chemical drug[100]. This strategy is extremely informative to elucidate genetic networks that can have direct clinical applications. In this direction, synthetic lethal RNAi screens have been exploited to identify genes that are able to increase the sensitivity to a variety of chemotherapeutic drugs when are targeted[120-124]. The presence of residual gene activity upon RNAi knock-down more closely recapitulate the physiological situation that could be obtained with pharmacological inhibitors than the knock-out system which completely eliminates any gene activity[100].

So far, hundreds of large-scale cell-based RNAi screens have been performed in *Drosophila melanogaster*, mouse and human cells. Recently, a further evolution of pooled RNAi screens highlighted the use of molecular 'barcodes' associated with individual hairpins to facilitate the tracking of individual shRNAs behaviour in complex populations.

3.3 In vivo genetic screens using RNAi

RNAi screening approaches are now applicable to animal models due to the increasing throughput and decreased costs associated with next-generation sequencing (NGS). So, individual shRNAs can be used as biological probe to interrogate specific cellular phenotypes but also to understand *in vitro* or *in vivo* molecular details. In this direction, loss-of-function (LOF) screens with pooled shRNA libraries have been successfully proposed in animal models of different cancer types (ex. Leukemia, Breast, Glioblastoma)[125-130]. Screens permormed in lymphomas have highlighted that several genes are involved in functions that are not critical in tissue culture conditions but are essential during cancer formation due to *in vivo* cooperation with the host microenvironment[127]. This finding proves the unique feature of *in vivo* RNAi screens and highlights the ability of these approches in investigating the tumor cells' vulnerability in more physiological contexts compared to *in vitro* systems. An *in vivo* RNAi screening experiment is based on the silencing of a panel of genes in a cell population constituted by tumor-initiating cells, the transplantation of the infected cells into recipient hosts and the tumor formation as an endpoint (**Fig.3**). In these approaches, tumor-initiating cells (TICs) infected with pooled shRNA libraries sustain *in vivo* tumor proliferation and the initial equal representation of single-shRNA carrying cells is drifted toward tumor cells that displayed a functional advantage due to the biological effect of the shRNA product. The TICs can be generally defined as a population of cells with the unique feature of generating a tumor in recipient animals, and so this definition includes the so-called 'cancer stem cells'.

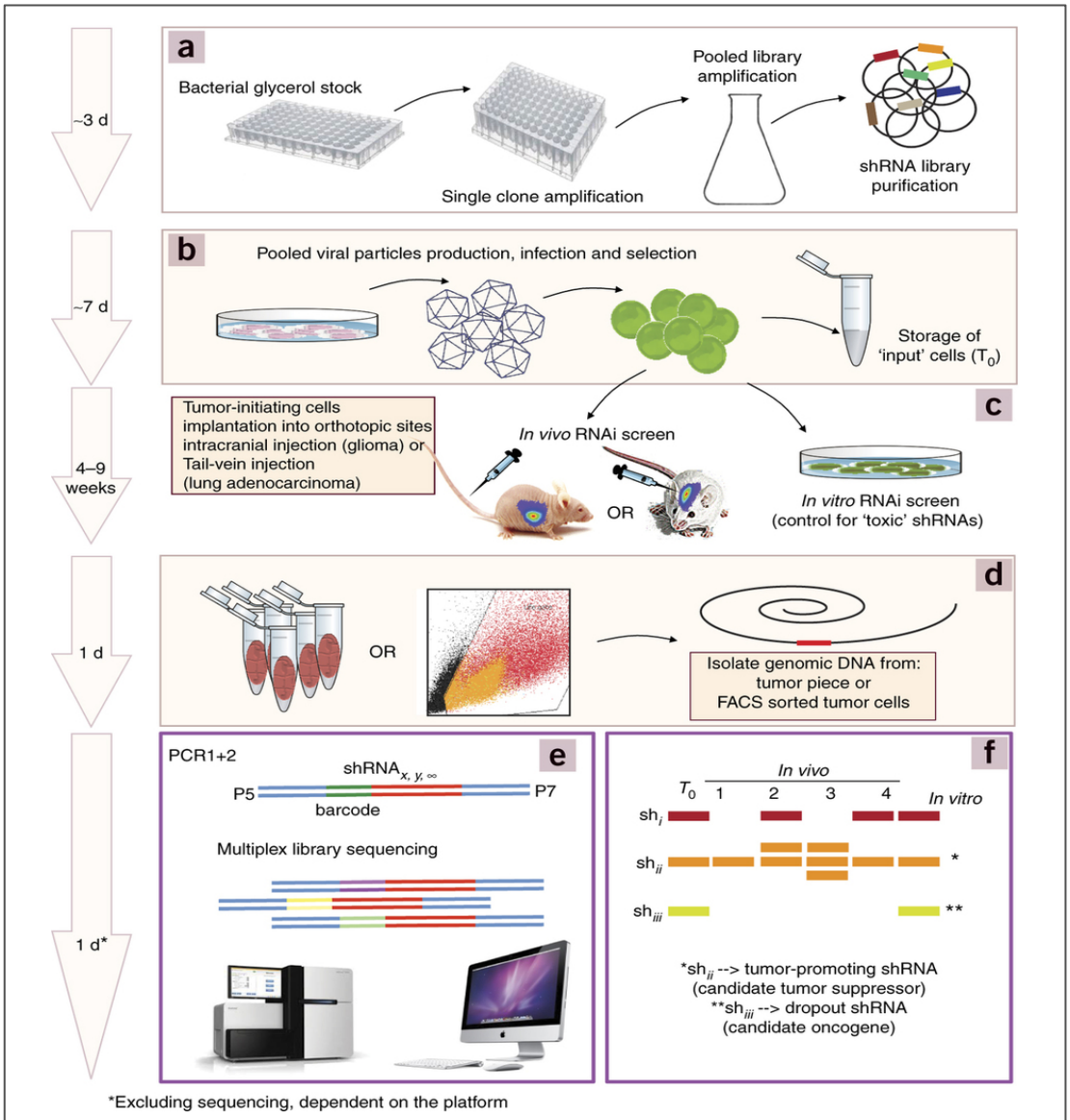


Figure 3. Outline of an *in vivo* shRNA screen.
(a–d) shRNA library assembly, transfection and transduction **(a–c)**, is followed by implantation of shRNA modified cells into recipient animals **(d)**. **(d–f)** Tumor-enriched shRNAs are amplified from tissue or FACS-purified cells by PCR **(d,e)** and counted to identify enrichments and dropouts **(f)**.[131]

Basically, shRNA-drived reduction of oncogenes level determine the depletion of the cells carrying that specific shRNA. On the other side, shRNAs enriched in the final tumor population could be considered hairpins that target putative tumor suppressor genes (TSGs). Neutral shRNAs can be defined as hairpins that are not changing their representation in comparison with the starting population and the number is largely influenced by the duration of the experiment[131].

3.3.1 in vivo RNAi screen procedures

Six main procedural steps should be taken into account for performing *in vivo* RNAi screens[131]:

- *Production of shRNAs*. Individual shRNAs are primarily expanded as bacterial cultures and then pooled together when the bacterial preparations are confluent. This approach gurantees a fexible shRNA selection and permits individual validation of successful amplification of bacterial cultures. At the end, this step attenuates the chance of individual shRNA over-growth, which may be responsabile for critical over- or under-representation of specific shRNAs in the pooled library.
- *Transfection*. The pool of produced shRNAs is co-transfected into HEK293T cells together with second- or third-generation lentiviral packaging constructs. TICs of the model of interst are infected at a multiplicity of infection (MOI). MOI must be <1 in order to gurantee that only one hairpin is integrated in one cell but not more. Then, TICs are selected using puromycin (for pLKO.1 lentiviral vectors) or fluorescent reporters (ex. GFP, RFP) to remove the uninfected portion of the cell population.
- *TIC implantation*. Upon selection, TICs are transplanted into recipient hosts in a number that is calculated taking into account the technical limits of the required surgical procedure, the physiological requirement for the selected tumor to ricapitulate the tumor of origin and the size of the selected library. At this step, a representative portion of the TICs is collected as refrence cells to control for hairpin representation before transplantation. If possible, *in vitro* screen should be performed in parallel with the *in vivo* one, in order to allow the infected TICs to proliferate under specific conditions of culturing.
- *DNA extraction*. At the appearance of tumor-related symptoms or after *in vivo* imaging with luciferase systems, animals are sacrified and the tumor is *in toto* collected. shRNAs, integrated in the genome of the host cells, can be retrieved from

genomic DNA or directly extracted from tumor biopsies or from viable tumor cells that have been FACS-purified to exclude host-derived cells.

- *Sequencing*. Extracted genomic DNA from tumor cells, reference cells and tissue is then subjected to PCR amplification cycles. Generally, a first-step PCR is applied to amplify the hairpins in the samples and a secondary-step PCR is performed to introduce the adapters for deep sequencing. At the end, PCR libraries are quantified, mixed and sequenced.
- *Data analysis*. PCR product sequences are aligned to the reference list of the hairpins present in the library, and each identified hairpin is counted and assigned to the sample of origin (if multiplexed). Each hairpin is normalized in aech sample by the sequencing depth (i.e., total counts per sample) and by the relative abundance in the reference cells. A final list of tumor enriched and reduced genes is then generated for secondary validation.

3.3.2 In vivo RNAi screen in human samples

The *in vivo* screens performed in mouse model systems opened the intriguing possibility of expanding the same functional genomic studies to directly inform on human sample vulnerabilities. The first attempt in this direction was proposed by Possemato and colleagues, establishing an *in vivo* RNAi dropout screen on a breast cancer cell line[132]. They were able to adapt the required steps to the human engraftment conditions and identified a new serine biosynthesis pathway dependency for this disease. Following on this breakthrough, Passik and colleagues extended the approach to melanoma samples shedding light on the usefulness of DNA repair inhibition for the treatment of this disease and highlighting different molecular points of vulnerability[133].

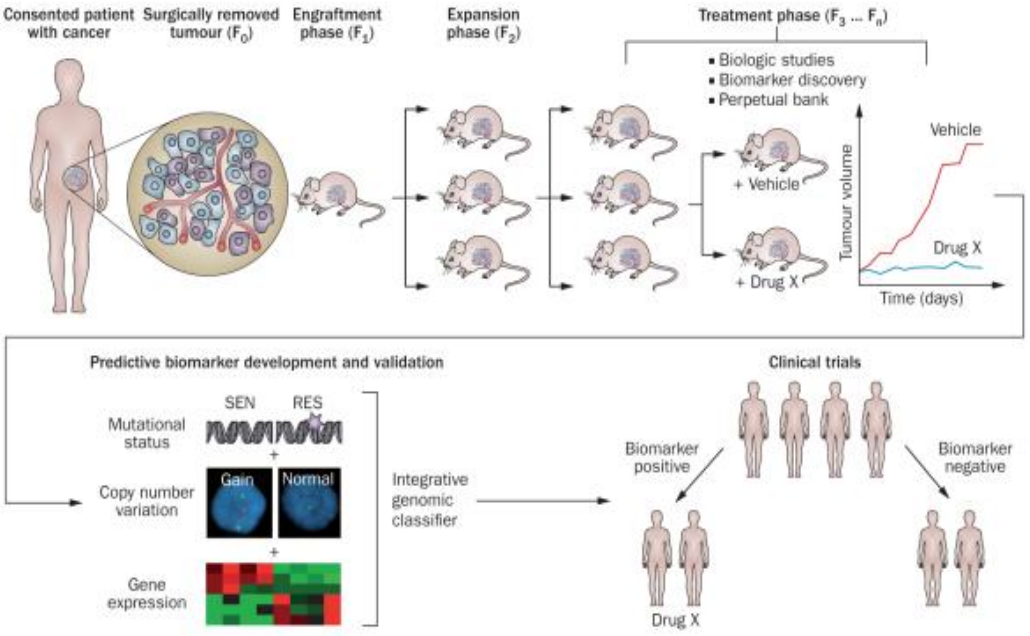
However, the difficulties intrinsically connected with the nature of the xenograft engraftment models limited the extensive application of the *in vivo* RNAi screening technologies and step back the approach more as a biological tool than a medical resource.

4. Patient-derived xenograft (PDX) models in cancer discovery

One of the most frequent reason for the high failure rate of new drugs in oncology is the lack of pre-clinical models that are able to fully mimicking the patient tumor heterogeneity.[134] Even though the application of cancer cell-line culture techniques pushed an acceleration and expansion of cancer biology discovery, in reality the situation is that the chance of translating these findings into clinical outcomes has been limited by the same models that generated such relevant improvements. Several different explanations have been proposed for this failure, for example the fact that cell lines, even when amplifieded *in vivo*, are obtained from cancer cells that have adapted to proliferate without a physiological tumor microenvironment, determining genetic profiles that are differents from the ones imposed by the genetic pressure in patients[135]. In the same way, there are numerous evidences that the genetic drift is more consistent between a primary tumor and the corresponding cell line, than a direct patient-derived xenograft, even after several generations[135]. So, even though the chance of successfully transplant patient-derived tumors has been established some decades ago, these models for preclinical studies are just recently sistematically characterized and applied for drug discovery purposes in oncology[136–142].

The procedure for establishment and amplification of patient-derived xenografts (PDXs, **Fig.4**) has been deeply investigated and presented by multiple groups[140–143]. The approach is very simple and based on the collection of fresh surgical tissue, fractionation into ~3 mm^3 pieces and subcutaneous or orthotopic transplantation into the flank of an immunodeficient mouse. The primary mouse generation transplanted with the patient-derived material is termed F0, while the subsequent generations are numbered consecutively (F1, F2, F3 and so on), although some other groups have used the names G0, G1 and so on[143]. The time span required for tumor tumor formation in the primay generation is highly variable considering tumor types, implantation site and recipient mouse strain, but generally it takes between 2 months and 4 months, even though a failure of engraftment could not be considered

definitive until 6 months from transplantation[140]. As a general approach, starting from the third generation (F3 or G3) the cohort can be expanded for drug treatment. However, the only real parameter that should be considered to evaluate the reliability of the PDX is the degree of divergency between the patient's tumor and the corresponding xenograft in terms of genetics and histology, two factors that are underestimated when results of therapeutic studies are presented. (**Fig.4**)



Figuere 4. Establishing and testing of PDX models. Excess tumor specimens not needed for clinical diagnosis are obtained from the consented patients (F_0). Non-necrotic areas of these tumors are sectioned into ~3 mm^3 pieces and, after processing, implanted subcutaneously into anaesthetized 5-week to 6-week-old female athymic nude mice. During the engraftment phase, tumors are allowed to establish and grow and then are harvested upon reaching a size of 1,500 mm^3 (F_1). Similar protocols are employed for subsequent expansion cohort (F_2) and treatment cohort (F_3 … F_n). Typically, biological assays are performed on tumors in early generations ($\leq F_5$); these biological assays include drug efficacy studies, rational combination studies and the development of predictive biomarkers for novel targeted therapies. If the developed biomarkers achieved accurate prediction in a validation set of PDX models (or 'xenopatients'), they might be translated into early phase clinical trials as tools for patient selection strategies. Abbreviations: PDX, patient-derived xenografts; RES, resistant; SEN, sensitive.[143]

4.1 Engraftment in host recipients

The comparison among recipient strains or hosts, such as athymic nude mice, NOD/SCID mice or NSG (NOD/SCID/IL2Rγnull) mice was not sistematically performed taking into account the engraftment time span, engraftment rate, genetic landscape, or histological profiles. The majority of studies highlighted engraftment rates of 75% or above using athymic nu/nu mice. The NOD/SCID mice are more often applied in F1[145,146]. Then, the

development of the NSG mice has increased the chance of reaching even higher engraftment rates (close to 95–100%). The NSG mice are characterized by a further inhibition of the innate immunity by arresting the maturation of natural killer (NK) cells[147].

4.2 Phenocopy the tumor of origin

The main advantages poited out for the PDX models is the ability of phenocopying the original tumor architecture and histological features, even though there are still debates about timing and extent the human-derived microenvironment is preserved[148]. A new promosing approach to bypass the disappearance of human-derived microvasculature applied tissue microarrays obtained from 150 PDX models, which were evaluated for a set of vasculature-associated genes demonstrating features of an angiogenic profile that could be exploited for the identification of new therapies oriented towards the tumor microenvironment[148,149]. Similar approaches, such as gene-set enrichment analysis (GSEA) of angiogenic and metastatic profiles, can be applied to circumvent the caveats releated to the ability of human microenvironment recapitulation in PDX models[150].

4.3 Transcriptional and mutational stability

A relevant point associated to the PDX model stability is the extent of changes that the methodologies of engraftment and amplification introduce to the genetic profile of the tumors. Comprehensive genome-wide gene-expression analysis studies have highlighted that PDX maximize the preservation of the key genes and global pathway activity in primary tumors[135,152]. In this direction, studies performed in PDX models of non-small-cell lung cancer (NSCLC) using unsupervised hierarchical clustering of genome-wide gene-expression profiles demonstrated that 9 out of the 17 primary tumors clustered together with the derived PDX models, with Pearson's correlation coefficients ranging from 0.78 to 0.95. Notably, 10 of this 17 primary–PDX tumor pairs displayed correlation coefficients >0.90

highlighting a high degree of similarity between the primary cancer and the corresponding PDX model[151]. In the same way, 10 out of 12 primary pancreatic cancer PDX (F0 versus F3) tumors showed the same *KRAS* mutational and *SMAD4* expression status[144]. More than that, some of the pancreatic cancer PDX models have been also selected to integrate the tumor DNA content from the patient's material for the pancreatic cancer genome sequencing project[152] (**Fig.5**). Comparative studies of matched patient–PDX models applying genome-wide gene expression methodologies in colorectal cancer (CRC) and pancreatic ductal adenocarcinoma (PDAC) were also recently presented. The third-generation PDX model for colon cancer and the fifth-generation PDX model for pancreatic cancer displayed high correlation of global gene expression when compared with their matched primary tumors.

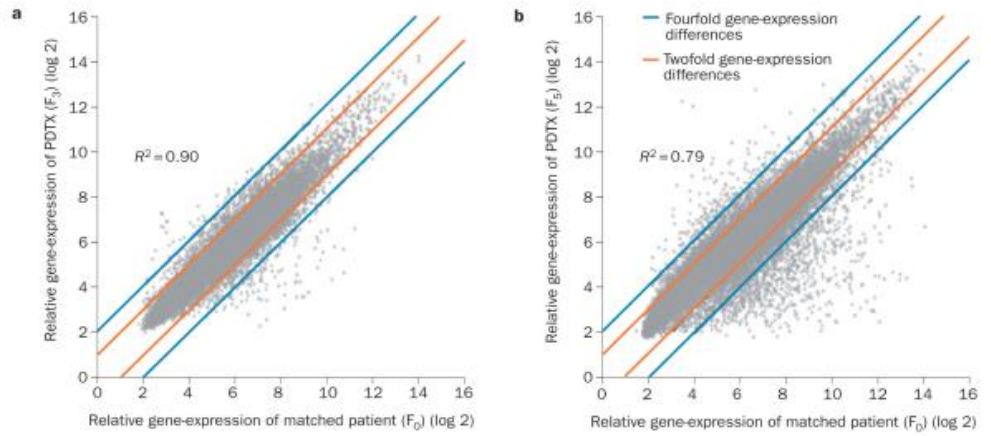


Figure 5. Comparison of genome-wide gene-expression profiles between primary patient tumors and PDX tumors. **(a)** Matched patient primary CRC tumor (F_0) and PDX (F_3). Genome-wide gene-expression profiles of a patient with CRC and their matched PDX were profiled with Affymetrix® HuGene 1.0 ST arrays. **(b)** Matched patient primary PDA tumor (F_0) and PDTX (F_5). Genome-wide gene-expression profiles of a patient with PDA and their matched PDX were profiled with Affymetrix® HG-U133 Plus 2.0 arrays. High correlations were observed in both PDX models and their matched primary tumors. Abbreviations: CRC, colorectal cancer; PDA, pancreatic ductal adenocarcinoma; PDX, patient-derived tumor xenografts.[143]

4.4 PDX as a platform for translational cancer research (co-clinical trials)

It is pretty well established that one of the main limitations in oncology drug discovery processes is the low rate of new compounds that are able to reach clinical approval.[153] The reason for these unsuccesses is in part due to the fact that the conventional preclinical models used to valiadate new drugs lack in predictive value[154]. So, rhe reason for diffusion of PDX models is intensically connected to the fact that these models are better predictors of

the patient's clinical outcome. Several studies have tested the response rate of drugs used as standard of care in medical oncology in PDX models, particularly in colorectal cancer, NSCLC, SCCHN, human breast cancer, and renal cell cancer (RCC). These reports demonstrated that the response rates in PDX models match with good approximation with those observed in the corresponding patients, both for targeted agents and for classic cytotoxics. More recently, a prospective study in PDAC highlighted a new potential role of PDX models as screening platforms for clinical trials. This study clarified that the combination of *nab*-paclitaxel and gemcitabine is effective in PDX models of PDAC, a result perfectly in agreement with the efficacy of this combination in PDAC patients. This regimen has recently confirmed to be able to determine a survival improvement for patients with advanced PDAC, as demonstrated in a randomized phase III study, and is now under evaluation to become a standard of care for this dramatic disease[7]. In the same way, absence of antitumor efficacy in PDX models reflects negative results in clinical treatments. These findings were confirmed in PDAC with agents such as the SRC inhibitor saracatinib and the mTOR inhibitor sirolimus, for which lack of efficacy in PDX preclinical studies predicted failure of the same strategy in the clinic[156,156]. Taking all these data together, it could be easily understood why PDX models have now obtained a prominent role in the preclinical phase of new anticancer drug development. One crucial advantage exerted by PDX models in large preclinical studies is connected to the fact they contribute to clinical indications prioritization, as well as identification of new biomarkers for potential drug efficacy[157,158]. Preclinical studies in PDX models can also help in optimizing the clinical trial design. This potential has been illustrated in studies involving cancer stem cell (CSC) drugs such as inhibitors of the Sonic Hedgehog, Nodal/Activin, TGFβ, and Notch pathways[159-162]. In experiments with PDX models, these agents failed to cooperate with classical chemotherapy in tumor regression but significantly delayed tumor growth and decreased tumor initiation and relapse. So, the application of PDX models demonstrated to be critical to determine and evaluate the effect of pharmacologic compounds on CSCs. In addition, these findings could

be crucial for clinical trial design that are also oriented to consider the treatment of minimal residual disease and to decide the appropriate setting in which to apply this approach. So, PDX models may exert a fundamental role in drug efficacy studies and they could help in selecting the populations of patients most likely to be sensitive to a new drug, as well as to prioritize the development of new biomarkers.

4.5 Tumor-Initiating Cell (TIC) frequency

The finding that phenotypically distinct cancer cell subpopulations are able to sustain tumor growth in serial transplants, whereas the majority of tumor cells appear to be bystanders in the process, has contributed to the arising of the cancer stem cell (CSC) concept in tumorigenesis both in hematological and solid tumor malignancies[163-167].

Starting from the first experimental prove for the existence of a solid tumor-initiating cell in human breast cancers, the classification models corresponding to CSC have remained complicated and the final link between CSC, normal stem cell populations, and the 'cell of origin' in cancer was not fully clarified[163,164]. So, even though a CSC involved in tumor's origin can probably originate from normal stem cells as a result of sequencial mutations that confer oncogenic properties, the cell of origin in cancer and the predominant CSC present in an evolved tumor may possess differing features releated to the point the tumor is considerd[164]. To be precise, the CSC properties during disease development can change and the identity can be different from from the cell of origin.

To cirumvent semantic misuderstandings connected to CSC concept, a different name chosen for these cells is tumor perpetuating cells (TPCs), as these cells are characterized by their ability to fully recapitulate tumors as demonstrated by serial transplantation of small numbers of tumor cells with defined properties[163,164].

It is now fully accepted that a relevant heterogeneity is associated with CSC sub-populations. Serial transplantations in limiting dilution are performed to quantify the tumor-initiating cell

frequency, estimated to be between 1:100–1:20000 for the majority of solid tumors[167-171]. Melanoma appears to represent an exception, because one in four cells is able to reform tumors withwith the same features of the original tumor[172]. The concept that CSC must be infrequent was originated by the results highlighting that normal stem cells generally represent a small fraction of normal tissues. The relative infrequency of CSC in the initial studies of AML also enforced this assumption[173,174]. However, the melanoma case suggests that CSC can also be relatively frequent, depending on the indication, patient and/or stage of disease.

Concers related to the CSC concept have been emerged as soon as the xenograft tumor models were proposed, due to the fact that tumorigenicity can be obtained using a plethora of distinct markers previously identified as specific for CSC (ex. CD44, CD133, and ALDH1A1). These critics are primarly focused on experimental functional prove of tumor cell heterogeneity and paucity of tumor-initiating cells in solid tumors[175-179]. Basically, the features and frequencies of tumor-initiating cells in mouse and rat tumors have been highlighted applying autologous transplantation[175-178]. However, human tumor cells autologous transplantation, currently considered an unethical practise, displayed that tumor-initiating cell frequencies are consistently rare in well-differentiated tumors and require injection of more than 1 million cells for tumor formation[179]. These experiments of autologous tumor cell transplantation robustly supported the notion that the CSC concept is a real biological event and not only an artifact connected to the xenotransplantation procedure. Further studies highlighted a strong connection between tumor differentiation status and patient prognosis, proposing a clinical relevant relationship able to link the tumor differentiation with the paucity of CSCs in less aggressive tumors, whereas CSCs seem to be more represented in aggressive tumors or later stages of the disease[180].

4.5.1 High TIC frequency - Melanoma PDXs

Until the last years, patients with advanced melanoma had limited therapeutic options. Recently, the approval of the BRAF inhibitor vemurafenib and the CTLA-4 inhibitor ipilimumab have opened new unexplored therapeutic directions. In this scenario, the PDX models are capturing the attention due to their ability in defining resistance pathways and rational combination strategies for the disease. Even though the establishment of PDX melanoma models was proposed many years ago, there are no comprehensive studies taking into account large numbers of models[181,182]. Preliminar experiments conducted with a melanoma PDX model from a primary tumor and a matched metastatic lesion were used to compare responses to anticancer drugs previously analyzed using cell lines derived from the tumors and PDX models. Although the studies demonstrated consistency for the majority of the responses, several different sensitivities were also detected[181]. Notably, a gene-array study executed using a panel of 22 melanoma PDX tumors was able to propose a predictive gene signature to 11 standard cytotoxic agents, even though no further clinical validation has been reported[182].

Melanoma PDX models demonstrated also to be a useful tool for the identification of melanoma tumor-initiating cells. For instance, isolation and re-implantation in limiting dilution of ABCB5+ cells from PDX models were able to regenerate the complete tumor heterogeneity and depauperation of this specific subpopulation resulted in tumor growth inhibition[183]. Another study presented the engraftment of a human uveal melanoma PDX model in NOD/SCID mice with an successful rate close to 28% and complete phenocopy of the the primary tumor in terms of histology and genetic profiles[184]. More than that, upon treatment with temozolomide, a standard chemotherapeutic agent for uveal melanoma, the response of engrafted mice was consistent with the clinical outcome of the patients[185].

4.5.2 Low TIC frequency - Pancreatic cancer PDXs

New molecular and genetic methodological tools have highlighted a significant accelleration in the scientific understanding of the complex genetics of PDAC. Whole-exome sequence analysis of primary PDAC tumors illustrated a set of 11 molecular pathways constantly affected in this disease together with an average of 63 genetic alterations interesting an individual tumor[152]. However, this accumulation of new genetic informations has not yet been able to improve the clinical outcomes for PDAC patients.

Pancreatic cancer cell lines have been routinely applied for preclinical studies with therapeutical candidates both *in vitro* and *in vivo*. However, xenografts obtained transplanting *in vitro* stabilized cancer cells determine the development of tumors characterized by masses of cancer cells with minimal stromal infiltration. So, these models could not be considered able to recapitulate the human PDAC architecture and the connections between stromal components and PDAC cells. In the same direction, it has been demonstrated that the intratumoral perfusion is negatively affected by the desmoplastic reaction, in a way that also the delivery of chemotherapeutic agents could be impacted and the antitumor effects of a given therapeutic strategy could be overestimated in cell line-based models.[186] PDX pancreatic models are based on the engraftment of primary human PDAC specimens in heterotopic or orthotopic anatomical locations.[187,188] The main feature of these PDX models is the ability of recapitulating the original tumor architecture, even though the human stroma is replaced during passaging by murine stroma. Orthotopic PDX models display a more consistent amount of stromal elements and are more prone to develop locoregional and distant metastases[187,188] (**Fig.6**). Pancreatic cancer PDX models were initially developed with the focus of optimizing the strategies for the identification of predictive and pharmacodynamic readouts for molecularly targeted therapies[144].

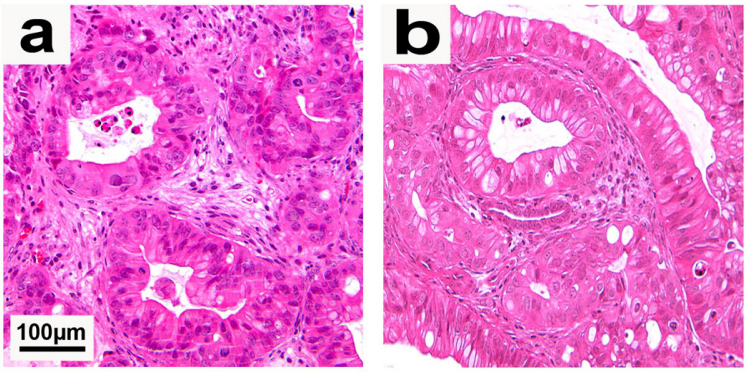


Figure 6. Pancreatic cancer xenografts. (a) Direct xenograft tumor established in a heterotopic site (subcutaneous space) of a NOD/SCID mouse. **(b)** Direct xenograft tumor established from injection of digested tumor (single cells) into the pancreas of a NOD/SCID mouse. Both heterotopic and orthotopic tumors show tumor gland formation and the presence of peri-tumoral stroma. Permission from the M.D. Anderson Cancer Center (MDACC) Institutional Regulatory Board was obtained for this study.[187]

Howevern, these achievements should have faced a massive intratumoral heterogeneity, especially taking into account gene and protein expression profiles. For example, the mTOR inhibitors were evaluated in PDX models and they were able to inform on the higher sensitivity of patients with high baseline expression of phosphorylated p70 S6 kinase, but this finding failed the transplation step towards PDAC patients[189]. This failure could be due to the parameters used to define a positive response in the PDX model or the numerouse potential feedback loops that exist for the mTOR pathway. Another proposed strategy was based on the coupling of tumor biopsy with *ex vivo* therapeutic treatment and pharmacodynamic readout[190] . This approach elucidated that a polo-like kinase (PLK) inhibitor is able to inhibit tumor growth, especially in PDX models resistant to gemcitabine treatment, and that cyclin B1 can be considered a biomarker of efficacy[191]. Another strategy that has been tested in PDAC is the treatment of a patient-specific PDX with a panel of approved drugs simultaneously with the first-line therapy in the matched patient[192]. As soon as the progression disease, the therapy demonstrating the most activity in their PDX is proposed as therapeutical option. This approach, in combination with sequencing tools, was able to prove the efficacy of a combined regimen of mitomycin C and cisplatin in a patient displaying *PALB2* mutation[193]. Limitations to this intriguing system are connected with the extreme variability in the engraftment success of the PDX and the requirement of considerable resources. A further possibility could be represented by the complete genomic profiling of a patient's tumor soon after the surgical collection in order to facilitate the

targeted therapy selection of the PDX, and then coming back with the results to the patient to direct the clinical decisions at the time of disease progression.

The stromal elements of the desmoplastic reaction tipically displayed in PDAC could represent novel targets for improving the treatment of this disease. The prominent role executed by the stroma in PDAC is highlighted by the findings that tumors cells engraftment in PDX models requires expression of stromal genes and is associated with decreased patient survival[150]. In PDAC, both genetically engineered and PDX models have illustrated that the stromal components may play a crucial role in regulating gemcitabine uptake by the tumor and can represent points of vulnerability to improve the therapy efficacy[186,193]. In the same way, inhibitions of the hedgehog pathway could have an effect in modulating the stroma, potentially due to the induction of apoptosis in pancreas stellate cells, with the result of ameliorating the tumor vascularity[186]. Notably, treatment of pancreatic cancer PDX models with a combination of gemcitabine and nabpaclitaxel affects the intratumoral desmoplastic reaction with a consequent increase in intratumoral gemcitabine concentrations and growth inhibition[193]. These experimental findings were translated into a phase I - II clinical trial of patients with advanced-stage PDAC and suggested that PDX models can be also used to propose therapeutical strategies directed to target the stromal elements in PDAC[193].

Aim of the project

The goal of this project is the identification of new potentially actionable molecular vulnerabilities in the context of pancreatic ductal adenocarcinoma (PDAC). The lack of effective therapies and the dramatic prognostic outcome classified new therapeutical options for this disease as a high priority medical need. Adapting the classical *in vivo* shRNA screening strategies to work with more predictive models, as the patient-derived xenografts, we aim to highlight new critical molecular dependencies that can be translated in more robust drug discovery programs. Specifically, we decided to focus our attention on epigentic mechanisms, normally preserved from mutational alterations in PDAC, in order to inform on specific addictions associated with pancreatic cancer. In parallel, interrogating genetically-definied mouse models in the same experimental settings we could increase the chance of associating epigenetic dependencies with the most frequent genetic landscapes in PDAC. Furthermore, conjugating novel validation models with the classical ones, we'd like to devolop a fast track pipeline for the rapid prioritization on the most promising chromatin-remodeling enzymes.

Results

1. Pancreatic cancer human xenografts and GEM models for *in vivo* screening

Transplantation of patient-derived samples in a host recipient, known as xenotransplantation, is a powerful tool to generate models able that faithfully recapitulate the tumor of origin[143,149,150]. Compared to the transplantation of in vitro established cell lines, patient-derived xenograft models provide a new layer of complexity reflecting the cellular heterogeneity of the original tumors and having avoided long term-adaptation to culturing in vitro. Numerous publications have shown that prolonged in vitro culturing drives an artificial selection of sub-clones and introduce a substantial genetic drift[134,135]. Some of these issues were recently overcome through the establishment of adhesion-independent culturing methods (organoids), but despite these efforts the ability to phenocopy the original tumor seems to be a unique feature of low-passage xenotransplantation approaches[194]. Analysis of expression profiles of patient-derived xenograft models confirmed that they maintain stability at least until the 4^{th}-5^{th} in vivo passage[151]. Patient derived-xenograft models to inform on patient vulnerabilities have been extensively exploited to test the activity of experimental and approved oncology drugs, and it appears that at least in certain instances they can predict clinical responses better than conventional xenografts[141,142]. In order to enhance our ability of identifying genetic dependencies in specific tumor contexts, we developed an in vivo screening platform in pancreatic adenocarcinoma patient-derived xenografts.

We first generated primary tumor derived xenografts (PDX) by transplanting small (approx. 3 mm^3) PDAC tumor pieces in each of five recipient mice. After this first enrichment step, tumors were excised from the animals, pooled together and single-cell preparations were obtained by applying a combination of mechanical and enzymatic dissociation protocols (**Fig.7a**). Purified single cells were seeded at high-confluence on collagen I-coated plates and grown using an optimized culturing medium supplemented with cofactors. We limited the

number of plate splitting to a maximum of 3 and introduced a “fibroblast-off” approach to eliminate mouse stromal cells. The purity of the PDX derived-cultures was determined using flow cytometry by estimating the percentage of human histocompatibility complex (HLA) positive cells (**Fig.7b**).

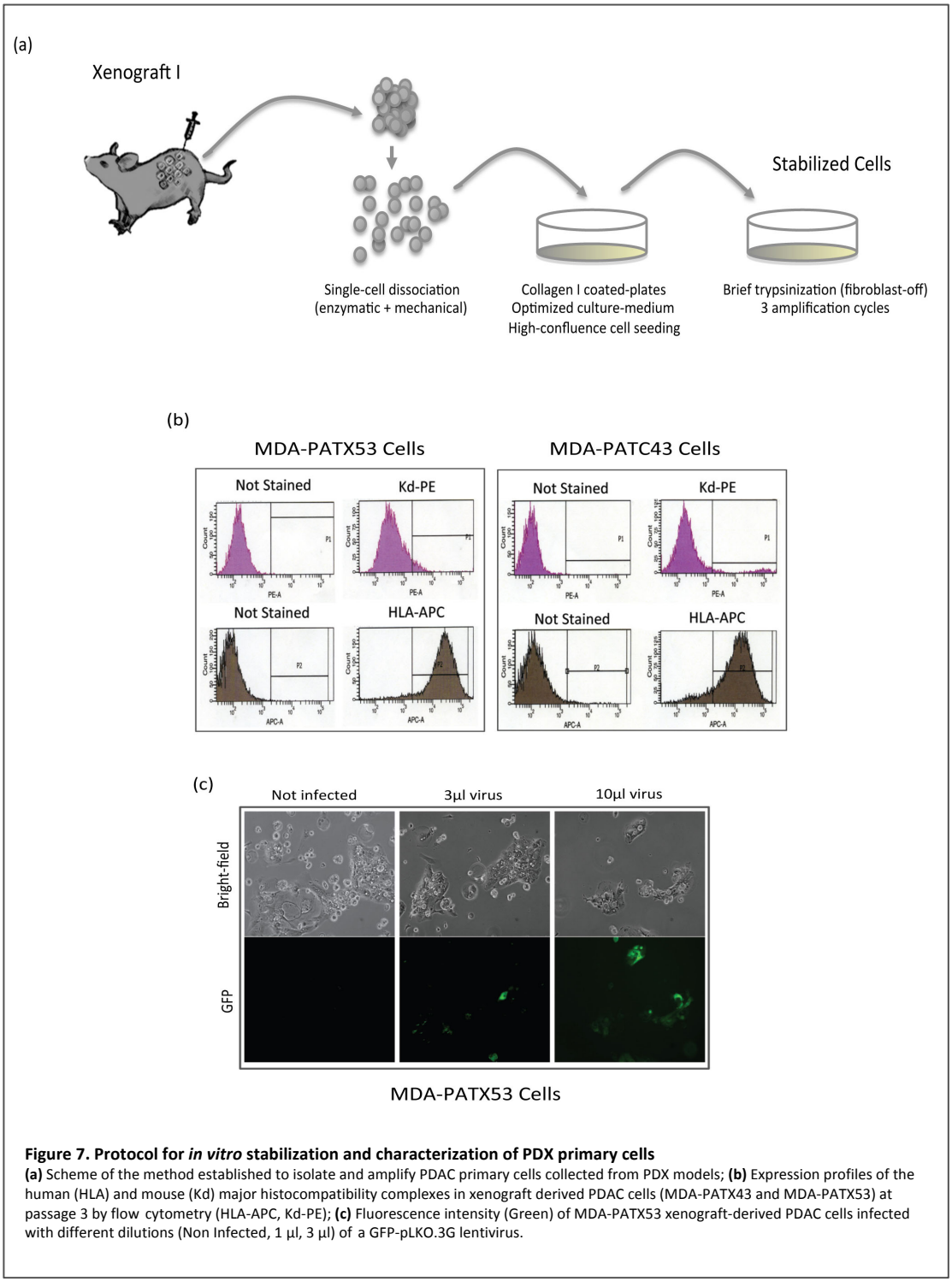


Figure 7. Protocol for *in vitro* stabilization and characterization of PDX primary cells
(a) Scheme of the method established to isolate and amplify PDAC primary cells collected from PDX models; **(b)** Expression profiles of the human (HLA) and mouse (Kd) major histocompatibility complexes in xenograft derived PDAC cells (MDA-PATX43 and MDA-PATX53) at passage 3 by flow cytometry (HLA-APC, Kd-PE); **(c)** Fluorescence intensity (Green) of MDA-PATX53 xenograft-derived PDAC cells infected with different dilutions (Non Infected, 1 µl, 3 µl) of a GFP-pLKO.3G lentivirus.

Before transplanting the isolated human cells in a secondary host, we verified not only ability of the culture to be infected with lentivirus, but also the possibility to modulate the infection rate (0, 3 and 10 µl of pLKO3.G-GFP virus) (**Fig.7c**). Tumors harvested from secondary xenografts were profiled for PDAC histological and histochemical features (H&E, Cytokeratin 19, Vimentin, 10x) and compared with primary xenograft and patient-derived specimens to verify that the histological characteristics of the original tumors were as much as possible maintained during this procedure (**Fig.8**).

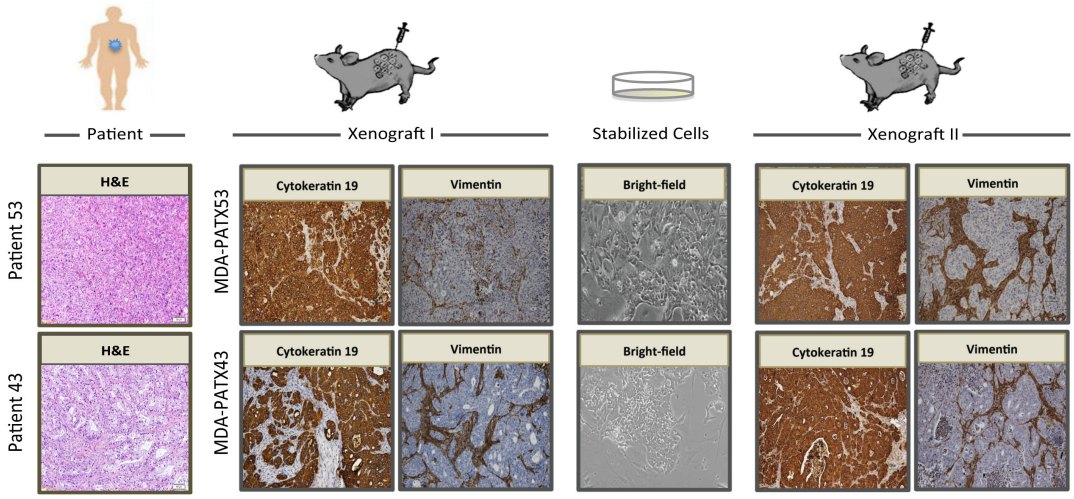


Figure 8. PDX models able to phenocopy the pancreatic cancer of patients
Histological sections of PDAC diagnosed patients (Patient 53, Patient43, 10x) were stained with Hematoxylin and Eosin (H&E) and compared with the sections collected from the matched PDX models (MDA-PATX53, MDA-PATX43). Primary (xenograft I) and secondary (xenograft II) xenografts were stained for Cytokeratin 19 and Vimentin (10x). Bright-fild images of the stabilized cells were taken during *in vitro* amplification steps (20x).

The architectural structure of the original PDAC tumors seemed to be conserved during passaging in mouse recipients and the histological properties of the tissue were preserved even after an *in vitro* step.

Using this method, 4 different PDAC xenograft models (MDA-PATX43, MDA-PATX50, MDA-PATX53 and MDA-PATX66) were generated and characterized for genetic alterations (mutations or deletions) most frequently associated with pancreatic cancer: KRAS, TP53, CDKN2A and SMAD4 (**Fig.9a**). Sequencing analysis showed that KRAS mutations were detectable in all the 4 PDX-derived models, while TP53 mutations were detected in only 2 of them (MDA-PATX50 and MDA-PATX66). In addition, 3 out of 4 PDX samples (MDA-PATX43, MDA-PATX50 and MDA-PATX66) had deletions in CDKN2A and SMAD4 genes. These

deletions were also confirmed at the level of protein expression by western blot. Interestingly, PDX-derived MDA-PATX53 cells, which reported no mutations or deletions except for KRAS, showed absence of expression of the tumor-suppressors p53 and p16 (**Fig.9b**). These peculiar features could be related to the fact that the MDA-PATX53 model was derived from a liver metastasis of a diagnosed PDAC tumor and not from a primary pancreatic location as all the others.

(a)

Patient Sample	Sex	Age	Pathology	Stage	Metastasis Site	Kras	TP53	CDKN2A	DPC4/ Smad4
MDA-PATX43	F	71	PDAC	3.0.0	No	Cod 12: GGT;CGT	WT	Del	Del
MDA-PATX50	F	60	PDAC	3.1.0	Liver & Lymph nodes	Cod 12: GGT;CGT	C(945,1061)>T	Del	Del
MDA-PATX53[2]	M	65	PDAC		Liver	Cod 12: GGT;GAT	WT	WT	WT
MDA-PATX66	F	63	PDAC	3.1.0	Lymph nodes	Cod 12: GGT; GAT	T(909,591)>A	Del	Del

[1]Patients received no pre-operation treatment [2]Tumor from liver metastasis site

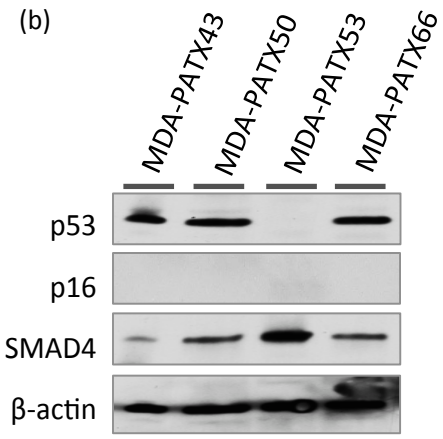


Figure 9. Clinico-pathological and mutational features of PDAC patient-derived samples
(a) Clinico-pathological (Sex, Age, Pathology, Stage, Metastasis Site) and mutational (Kras, TP53, CDKN2A, DPC4/Smad4) features of PDAC xenograft-derived samples (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66); **(b)** Protein expression of p53, p16, SMAD4 and β-actin in PDX-derived cells (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66).

Upon transplantation in a secondary xenograft, PDX-derived models displayed either epithelial (MDA-PATX43 and MDA-PATX53) or mesenchymal (MDA-PATX50 and MDA-PATX66) features. The contribution of mutant KRAS in sustaining the activation of its downstream pathways in PDX-derived cells was assessed by immunohistochemical staining for the phosphorylation level of its downstream targets ERK1/2 (**Fig.10**).

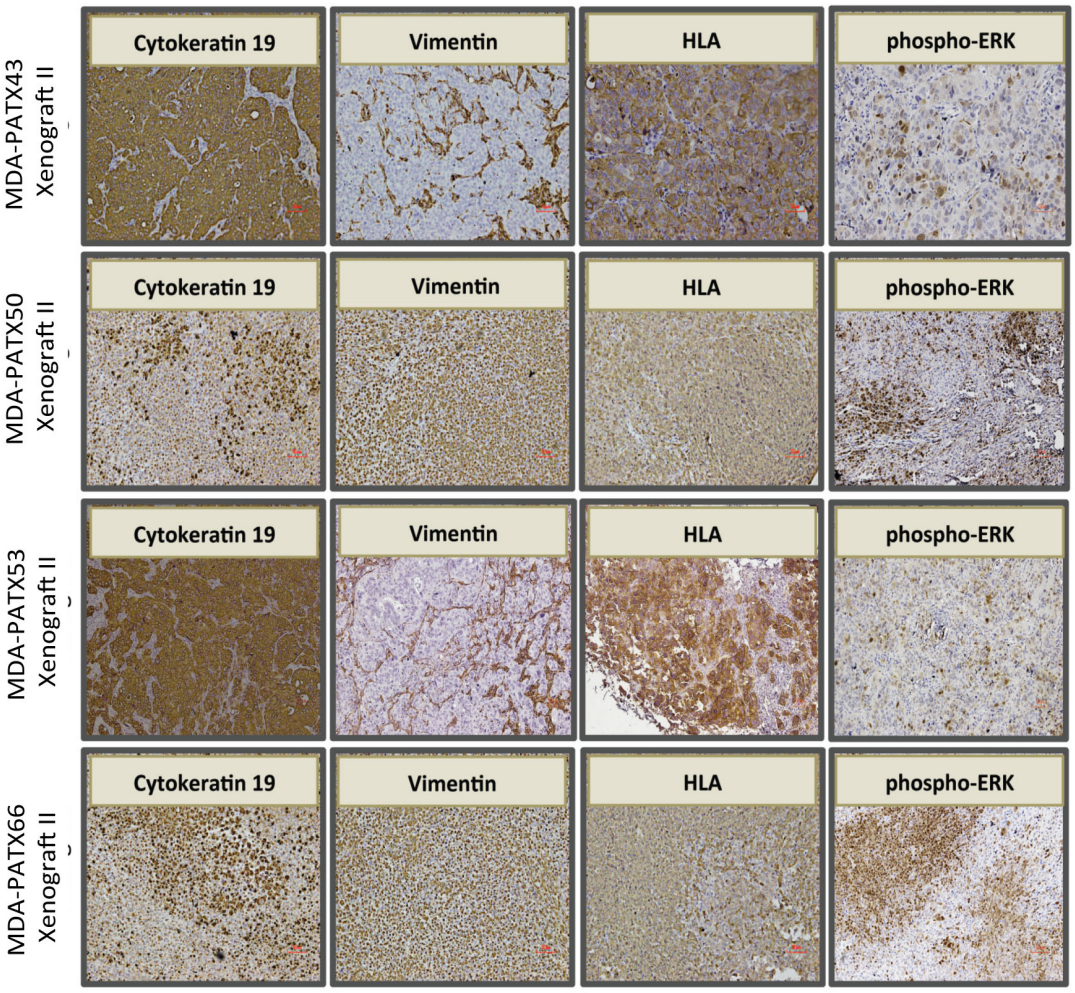


Figure 10. *In vivo* characterization of PDX secondary xenografts generated transplanting PDAC stabilized primary cells
Immunohistochemistry profiles (Cytokeratin 19, Vimentin, human major histocompatibility complex (HLA), phospho ERK, 10x) of the secondary xenografts (xenograft II) generated transplanting the PDAC xenograft-derived cells (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66) upon the stabilization step *in vitro*.

With a very similar approach, primary cultures were also established starting from PDAC genetic engineered mouse models (GEMMs):

- p48-Cre, $Kras^{G12D\_LSL/+}$;

- p48-Cre, $Kras^{G12D\_LSL/+}$, $Trp53^{L/L}$;
- p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$.

p48 is a critical transcription factor in pancreas development and its expression is essential to commit cells to a pancreatic fate[195]. To target the expression of mutant Kras in pancreatic progenitor cells, a conditional allele was generated through genetic elements inhibiting transcription and translation flanked by LoxP sites. Specifically, a Lox-Stop-Lox (LSL) cassette was inserted into the mouse genomic Kras locus to contain the G-A transition in locus 12 (G12D) in order to drive the expression of mutant Kras in pancreatic tissue expressing the Cre recombinase[196]. The p48-Cre, $Kras^{G12D\_LSL/+}$ model showed development of all the three stages of pre-neoplastic lesions (PanINs) associated with pancreatic cancer, but low-frequency progression to invasive and metastatic adenocarcinoma[196,197]. Single cell preparations were isolated from pancreata of 8 weeks old mice and in vitro cultured. The presence of mutant Kras in pancreatic cells drove over only a few passages oncogene-induced senescence, as assessed by β -galactosidase assay (data not shown). Spontaneously, some cells showed the ability to bypass senescence and proliferate. The characterization of these cells highlighted the *in vitro* expression of mesenchymal markers and the capacity of forming tumors with mesenchymal profiles when transplanted in the pancreas of a recipient animal. This tumorigenic cell population was called p48-Cre, $Kras^{G12D\_LSL/+}$ senescence-escaper (**Fig.11**). As Kras mutations are not sufficient to induce progression to the invasive stage of pancreatic adenocarcinoma, inactivation of either p53 or Ink4a has been used to generate combined models that progress to invasive PDAC[198,199]. Pancreatic tumors were harvested from p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ and p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$ mouse models, processed for single cell isolation and established in vitro as cell cultures. Orthotopic and subcutaneous transplantation of these cell populations generated tumors with epithelial profiles for the p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ background and mesenchymal ones for the p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$ (**Fig.11**).

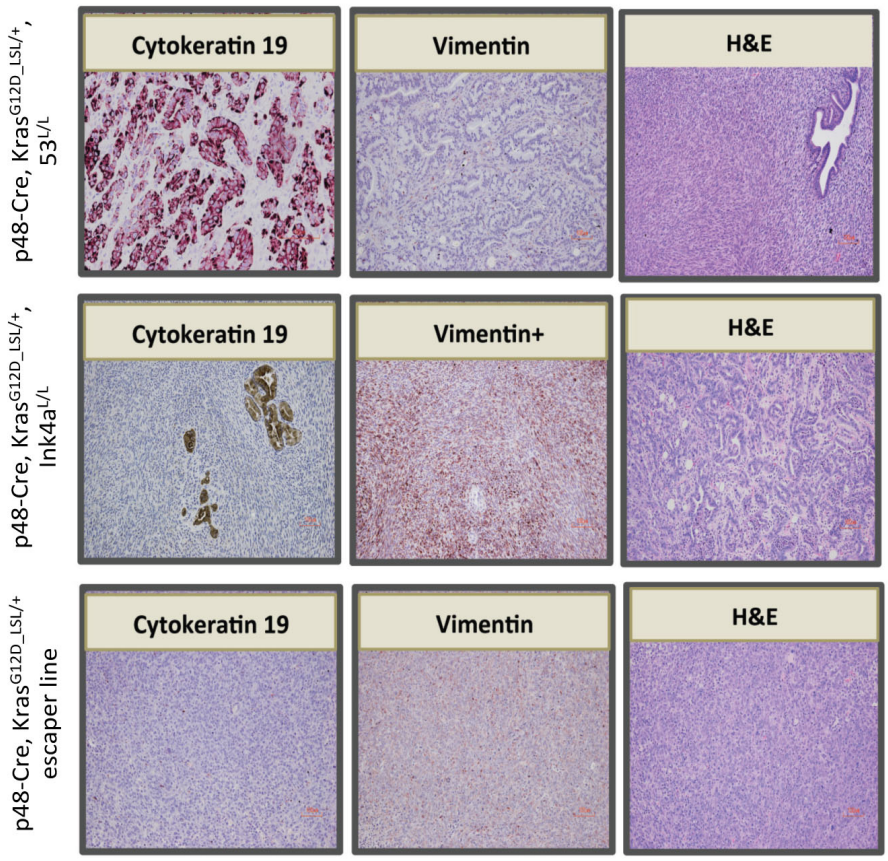


Figure 11. *In vivo* characterization of mouse allogarfts generated transplanting PDAC GEMM-derived cells
Immunohistochemistry profiles (Hematoxylin and Eosin, Cytokeratin 19, Vimentin, 10x) of the allografts generated translanting the PDAC mouse model-derived cells (p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper) upon *in vitro* stabilizaion.

2. Developing methods for rapid in vivo assessment of tumor-initiating cell frequency

A number of parameters have to be carefully controlled when attempting to identify gene products that are essential for in vivo tumor growth in a pooled screen approach[131]. In vivo shRNA screens rely on the silencing of a library of targeted genes in a cell population endowed with tumor engraftment capacity when implanted into recipient mice. Tumor initiating cell (TIC) frequency, i.e. the ability of a tumor cell population to engraft and propagate itself when implanted in vivo, varies dramatically between tumor types and must be accurately determined to ensure faithful representation of complex, pooled shRNA libraries[168,173]. The strategy most commonly applied to estimate TIC frequency in a cancer cell population is based on in vivo transplantation upon extreme limiting dilution assay (ELDA)[168-172]. To more rapidly and accurately determine TIC frequency in PDAC xenograft models, we first assessed the distribution and representation of a non-targeting library expressing 12,500 unique molecular barcodes in early passage tumor samples implanted in recipient mice. The barcoded-library was designed in a way that a molecular barcode of 18 unique nucleotides is cloned in the pRSI17 lentiviral vector (Cellecta) carrying a polycistronic site with the puromycin resistance ($Puro^R$) and RFP reporter gene (TagRFP) separated by the 2A peptide (**Fig.12a**). By "tagging" individual PDAC tumor cells, we could essentially track cell fate by comparing clone representation in a reference population prior to implantation into recipient mice, with that emerging after in vivo tumor establishment. If coverage (cells per barcode) is sufficient to sustain library complexity in vivo, we would expect a normal distribution when comparing barcode representation in xenografts with reference cells. If library complexity is not covered within the experimental system, the distribution of individual barcodes would be shifted compared to reference (**Fig.12b**).

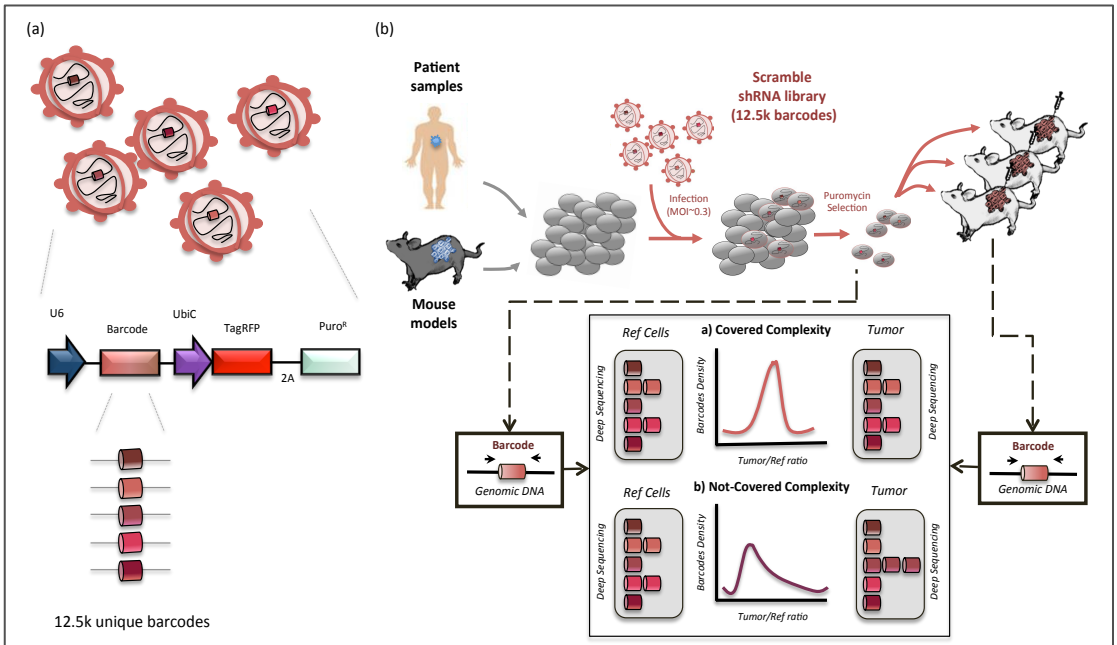
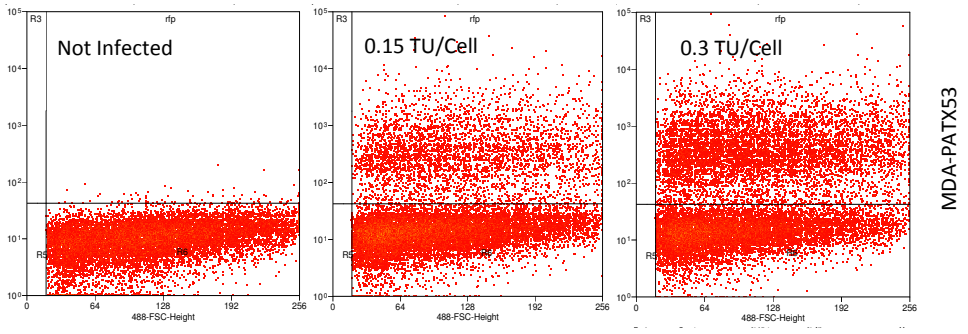


Figure 12. New method for rapid *in vivo* assessment of tumor-initiating cell (TIC) frequency
(a) Highlight of pRSI17 lentiviral vector cloning site carrying the 12.5k molecular barcodes and the downstream polycistronic site carrying the puromycin resistance ($Puro^R$) and RFP reporter gene (TagRFP) separated by the 2A peptide; **(b)** Outline of experimental design for in vivo TIC coverage study in patient-derived and mouse models of pancreatic ductal adenocarcinoma (MOI=multiplicity of infection, Ref Cells=Infected and selected cells before injection, Barcode=18 unique nucleotides), representative analysis (Tumor/Ref Log_2 ratio) of cells infected with a pool of unique molecular barcodes (18 nucleotides) in the same transplantation setting of the effective shRNA screen (i=model for covered complexity, ii=model for not-covered complexity);

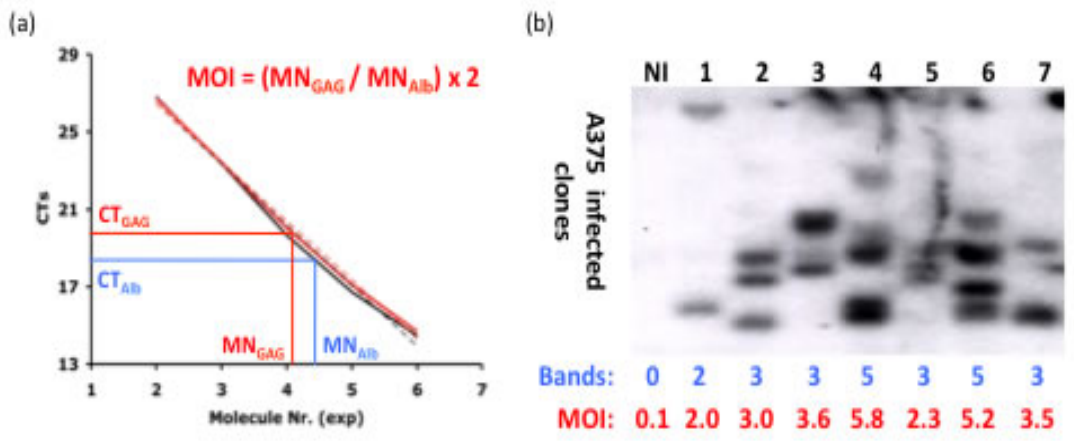
Through this effort, we could accurately determine the number of cells that are participating in tumor engraftment (TICs). To highlight the power of the approach, we transduced cells isolated from the early-passage human PDAC xenografts (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66) and PDAC GEM models (p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$; p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$; p48-Cre, $Kras^{G12D\_LSL/+}$ senescence-escaper). We infected cells with a lentiviral library expressing 12,500 unique molecular barcodes at a low multiplicity of infection to ensure that each cell received a single viral integrant. The Poisson's distribution of lentiviral integrants suggested the ideal infection to maximize the number of cells with only one integration as the one able to generate 25-30% infected cells (data not shown)[200]. This infection rate was established sample-by-sample and confirmed by FACS analysis through the fluorescent marker inserted in the lentiviral vector. The optimization of the multiplicity of infection (MOI) was performed in patient-derived cells applying three theoretical viral titers (0.15, 0.3 and 0.6 transducing units per cells (TU/cells)) of the barcoded library.



TU/cell	PATX53 tRFP (%)	PATX43 tRFP (%)	PATX50 tRFP (%)	PATX66 tRFP (%)	p48-Cre, Kras$^{G12D_LSL/+}$ escaper line tRFP (%)	p48-Cre, Kras$^{G12D_LSL/+}$, 53$^{L/L}$ tRFP (%)	p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a$^{L/+}$ tRFP (%)
0.15	16	14	12.5	20			
0.3	30	24	22	31	4.5	9	5.8
0.6					9	14	9.1
1.2					15	18	16.3

Figure 13. MOI optimization study for infection with barcoded libraries
(Flow-cytometry analysis for the RFP expression in MDA-PATX53 cells upon infection with different transducing units (Not Infected, 0.15 TU/Cells, 0.3 TU/Cells) of the 12.5k barcoded library, Day 2 post-infection (Upper Panel), Summary of the RFP-positive cells (percentage) from PDAC xenograft-derived (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66) and GEM (p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper) models upon infection with different titers of the 12.5k barcoded library: 0.15 TU/Cells, 0.3 TU/Cells, 0.6 TU/Cells and 1.2 TU/Cells, Day 2 post-infection (Lower Panel);

Due to the previously shown higher reticence of mouse cells of being infected with lentiviral vectors, we decided to increase the theoretical titers (0.3, 0.6 and 1.2 TU/cells) applied to the PDAC GEM models for the MOI optimization study. Cytofluorimetric analysis of the RFP positive cells, performed 2 days after infection, showed the optimal conditions for infecting human or mouse cells respectively at 0.3 and 1.2 TU/cell (**Fig.13**). Further confirmations that these established infections are able to maximize the number of cells with only one integrant were performed introducing a qPCR system for MOI assesment[201]. This method is based on the calculation of the number of molecules (NM) of a lentiviral component (Gag) integrated in the genomic DNA of an infected cell population and the successive comparison with the Albumin (Alb) gene, known to be present in the genome in 2 copies. The number of molecules was determined applying a titration curve with known amounts of a lentiviral vector carriyng a portion of the Albumin gene. A ratio close to 0.5 between Gag and Alb is the prove the lentivirus integrated only one time in the majority of the cells (**Fig.14a**).



(c)

TU/Cell	MOI
PATX43 0.3	1.2
PATX43 0.6	1.54
PATX50 0.3	1.16
PATX50 0.6	1.62
PATX53 0.3	0.92
PATX53 0.6	1.18
PATX66 0.3	1.25
PATX66 0.6	1.89

Figure 14. qPCR method for validation of the number of lentiviral integrants
(a) Titration curve generated amplifying known numbers of molecule of Albumine and GAG (lentiviral component) of a lentiviral vector carrying a fragment of the Albumin gene. Threshold cycles were calculated for Albumine (CT_{Alb}) and GAG (CT_{GAG}) using the genomic DNA extracted from infected cells and the same set of primers. Molecule numbers of Albumine (MN_{Alb}) and GAG (MN_{GAG}) were calculated from the genomic DNA and titration curve comparison and used to determine the MOI applying the formula: MOI= (MN_{GAG}/MN_{Alb}) x 2; **(b)** Southern blot of viral integrants performed on genomic DNA collected from A375 clones infected with a lentivirus at different MOIs (NI=Not infected, 1-7=Infected clones). Comparison between number of integrants determined by Southern Blot (Bands) and qPCR (MOI) methods; **(c)** Number of integrants calculated by qPCR (MOI) in PDAC xenograft-derived cells infected with 2 different dilutions of virus (0.3 and 0.6 TU/Cell).

Prior to apply to our studies, we confirmed the reliability of this tool comparing the MOI calculated through the qPCR method with the effective number of integrants that we were able to detect by Southern Blot in isolated clones of infected A375 cells (**Fig.14b**). The qPCR system demonstrated that the above defined infection conditions (0.3 and 1.2 TU/cells) are able to maximize the number of cells with a single integration in human and mouse PDAC models (**Fig.14c**). Upon controlled infection, cells were then selected with puromycin and implanted subcutaneously into NSG mice. Tumors were isolated from mice and the barcodes amplified from the genomic DNA by a 2-step PCR. We modified the secondary PCR in order to allow the employment of a set of primers carrying Illumina adapters (P5 and P7) to

facilitate the quantification of individual barcodes by next-generation sequencing (NGS) reducing the number of total cycles (**Fig15a**). To model optimal distribution of the complex barcode library, we implanted different cell numbers with each barcode expressed in 80, 240, or 400 individual cells (**Fig.15b,c**).

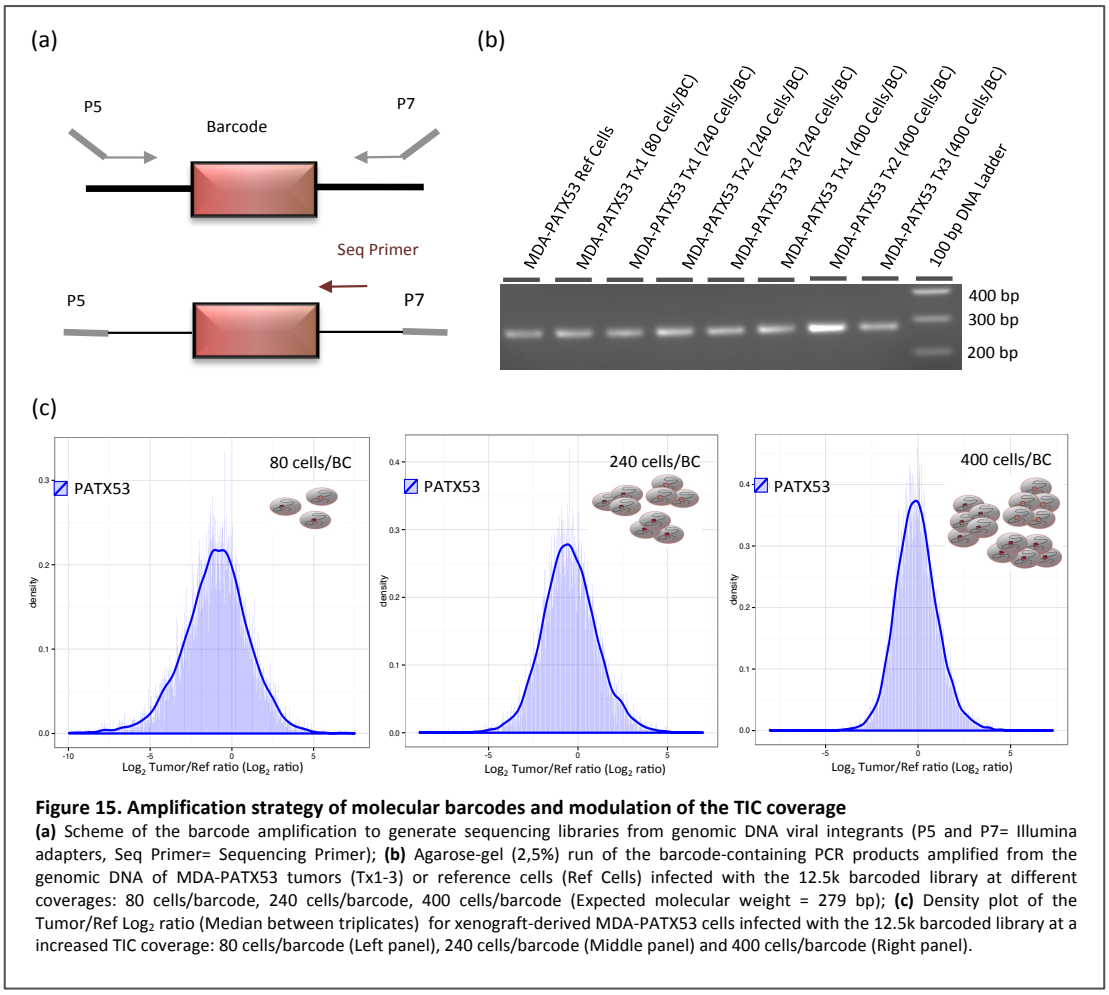


Figure 15. Amplification strategy of molecular barcodes and modulation of the TIC coverage
(a) Scheme of the barcode amplification to generate sequencing libraries from genomic DNA viral integrants (P5 and P7= Illumina adapters, Seq Primer= Sequencing Primer); **(b)** Agarose-gel (2,5%) run of the barcode-containing PCR products amplified from the genomic DNA of MDA-PATX53 tumors (Tx1-3) or reference cells (Ref Cells) infected with the 12.5k barcoded library at different coverages: 80 cells/barcode, 240 cells/barcode, 400 cells/barcode (Expected molecular weight = 279 bp); **(c)** Density plot of the Tumor/Ref Log_2 ratio (Median between triplicates) for xenograft-derived MDA-PATX53 cells infected with the 12.5k barcoded library at a increased TIC coverage: 80 cells/barcode (Left panel), 240 cells/barcode (Middle panel) and 400 cells/barcode (Right panel).

The reads generated by the HiSeq2000 were primarily filtered for a common portion of the lentiviral integrant to remove the non-specific errors produced during sequencing. The positions of the 18 nucleotide sequences corresponding to the 12,500 molecular barcodes were identified in the read string, trimmed and quantified. Upon normalization, the relative number of counts (Log_2) for each individual barcode in the tumor was compared with the one in the reference cells before transplantation. The Log_2 ratio between tumors and reference cells (Tumor/Ref Log_2 ratio) informed on the ability of the barcoded cell populations to be represented in the tumors as they grew. Applying a density plot to the

Tumor/Ref ratio we were able to appreciate the distribution of the 12,500 barcodes as a function of the change in representation (counts).

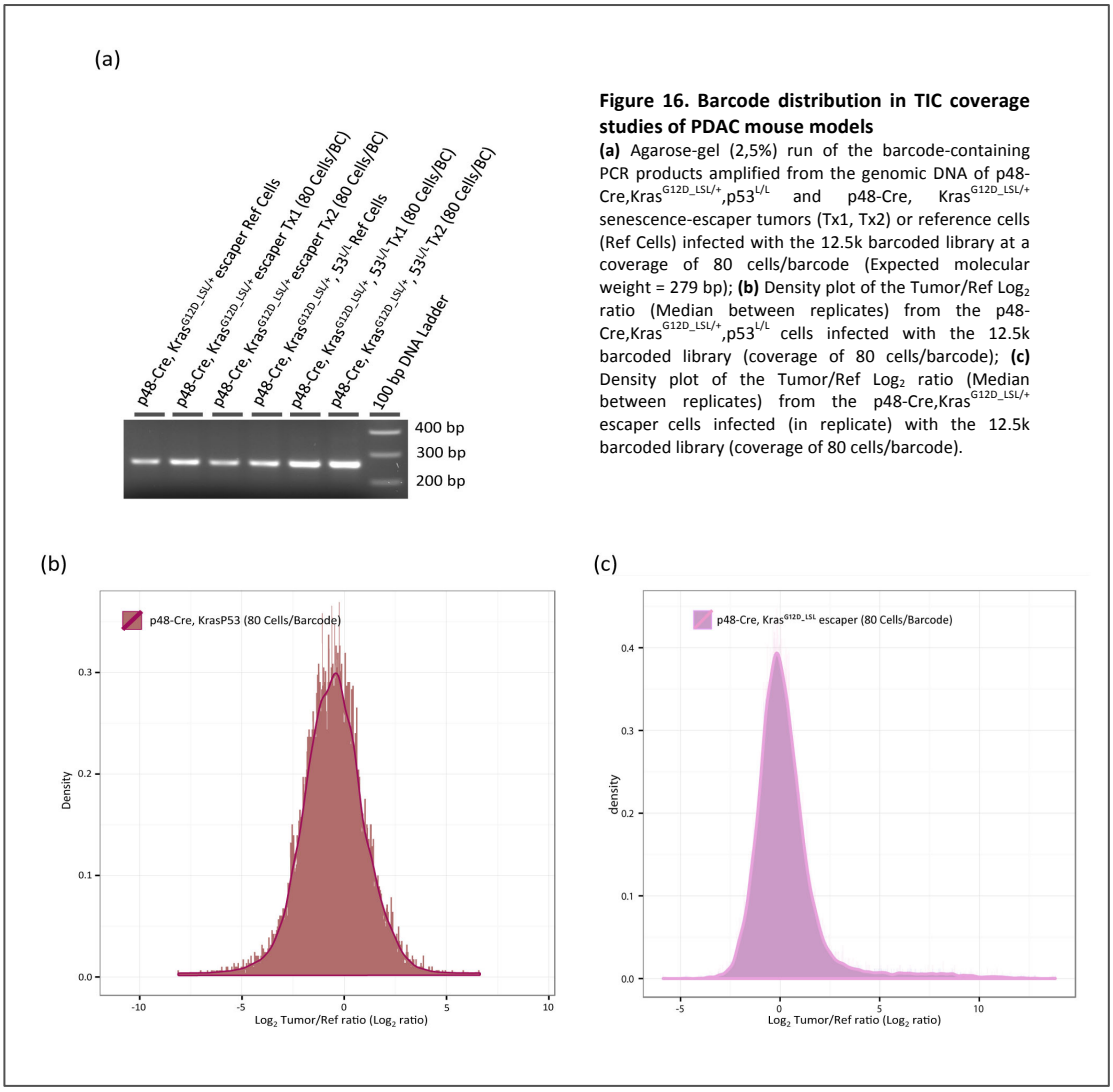


Figure 16. Barcode distribution in TIC coverage studies of PDAC mouse models
(a) Agarose-gel (2,5%) run of the barcode-containing PCR products amplified from the genomic DNA of p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$ and p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper tumors (Tx1, Tx2) or reference cells (Ref Cells) infected with the 12.5k barcoded library at a coverage of 80 cells/barcode (Expected molecular weight = 279 bp); **(b)** Density plot of the Tumor/Ref Log$_2$ ratio (Median between replicates) from the p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$ cells infected with the 12.5k barcoded library (coverage of 80 cells/barcode); **(c)** Density plot of the Tumor/Ref Log$_2$ ratio (Median between replicates) from the p48-Cre,Kras$^{G12D_LSL/+}$ escaper cells infected (in replicate) with the 12.5k barcoded library (coverage of 80 cells/barcode).

A centered normal distribution is indicative of a complex library coverage (cells per barcode) sufficient to represent each one of the barcodes in the established tumor. Alterations of the Gaussian curve or *in vivo* disappearance of barcodes could be considered as a failure in covering the number of the tumor initiating cells (TICs) that are participating in tumor engraftment (**Fig.15c**). A ratio of 80 cells per barcode was sufficient to retain representation of 12,500 different barcoded vectors in mouse PDAC implants derived from either the p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ or the p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper models (**Fig.16a-c**).

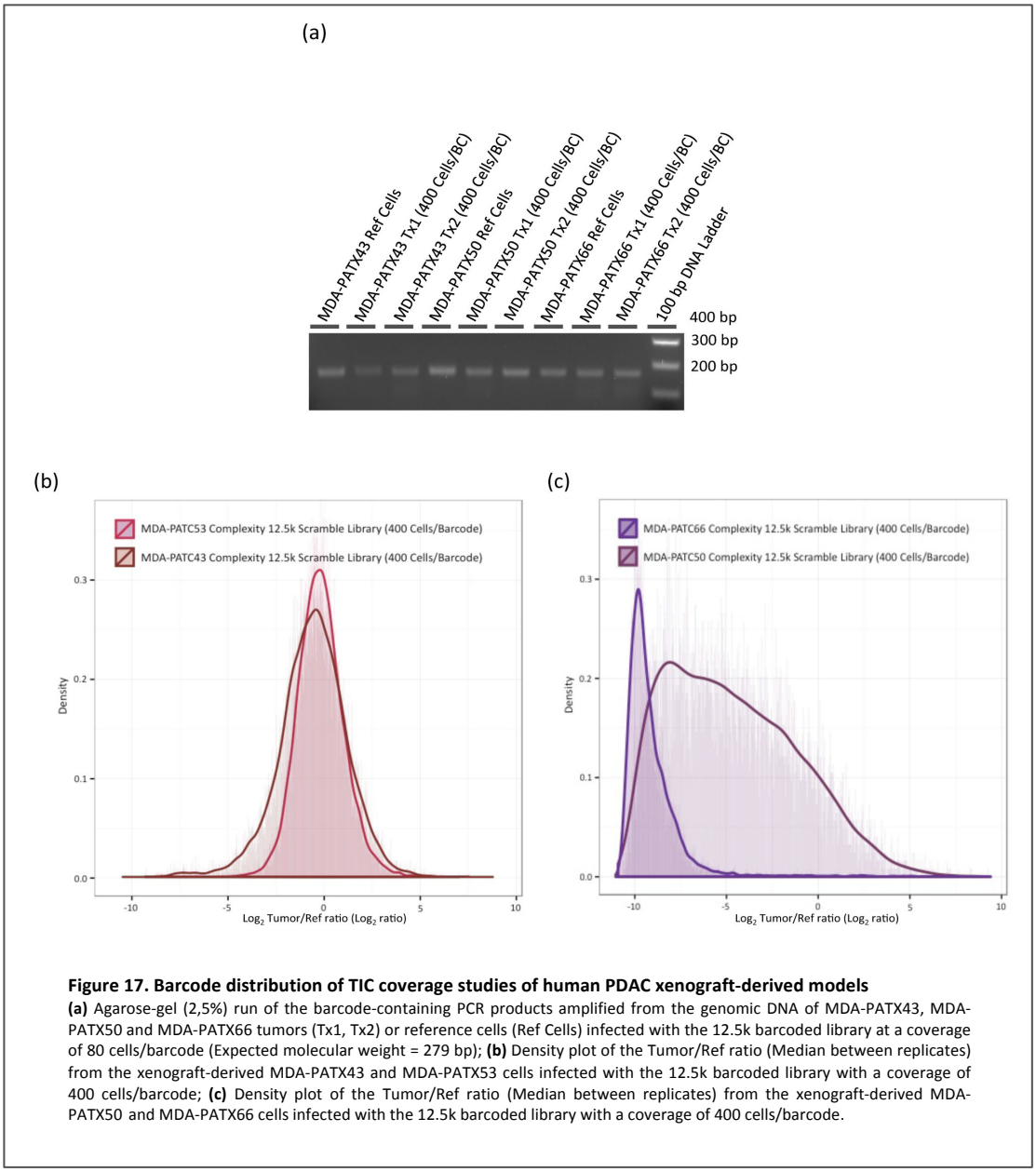


Figure 17. Barcode distribution of TIC coverage studies of human PDAC xenograft-derived models
(a) Agarose-gel (2,5%) run of the barcode-containing PCR products amplified from the genomic DNA of MDA-PATX43, MDA-PATX50 and MDA-PATX66 tumors (Tx1, Tx2) or reference cells (Ref Cells) infected with the 12.5k barcoded library at a coverage of 80 cells/barcode (Expected molecular weight = 279 bp); **(b)** Density plot of the Tumor/Ref ratio (Median between replicates) from the xenograft-derived MDA-PATX43 and MDA-PATX53 cells infected with the 12.5k barcoded library with a coverage of 400 cells/barcode; **(c)** Density plot of the Tumor/Ref ratio (Median between replicates) from the xenograft-derived MDA-PATX50 and MDA-PATX66 cells infected with the 12.5k barcoded library with a coverage of 400 cells/barcode.

Conversely, for either PDX-derived MDA-PATX53 or MDA-PATX43 samples, the number of cells required to retain complexity was at least 5-fold higher (400 cells/barcode) owning to their lower TIC frequency (**Fig.17a,b**). Two additional human PDX-derived models (MDA-PATX50 and MDA-PATX66) failed to provide a number of tumor-initiating cells sufficient to cover the complexity of a 12,500-barcode library, with the majority of barcodes being randomly depleted during mouse tumor engraftment (**Fig.17a,c**).

To confirm the relative TIC frequencies as estimated by barcode representation, we performed extreme limiting dilution assays (ELDA) on the same mouse and human PDAC tumor models. Through ELDA transplantation of primary mouse PDAC tumors we estimated

TIC frequencies in the p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ and the p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper models of 1:27 and 1:19 respectively.[202,203] Consistent with the results of our barcode experiments, ELDA estimated a 10-20-fold shift in TIC frequency between mouse and human PDAC, and further confirmed the variability between human PDAC samples with respect to TIC frequency (**Fig.18**). Taken together, these data suggest that the rapid determination of barcode representation in these tumor explant models can accurately predict TIC frequency compared to the time-intensive ELDA transplantation assays. Using this approach we can rapidly assess TIC frequency across patient-derived PDAC tumors and adjust barcode representation per cell to ensure adequate coverage, opening the possibility of performing targeted library screen on any patient obtainable tumor sample.

p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$

Cell Nr	Mice Nr
100,000	2/2
10,000	4/4
1,000	4/4
100	10/10
10	2/9
TIC Frequency: 1/27 *(95% C.I.: 1/12-1/60)*	

p48-Cre, Kras$^{G12D_LSL/+}$ escaper line

Cell Nr	Mice Nr
100,000	2/2
10,000	4/4
1,000	4/4
100	10/10
10	4/10
TIC Frequency: 1/19 *(95% C.I.: 1/8-1/43)*	

MDA-PATX53

Cell Nr	Mice Nr
100,000	2/2
10,000	4/4
1,000	4/4
100	2/10
10	1/10
TIC Frequency: 1/288 *(95% C.I.: 1/118-1/704)*	

MDA-PATX43

Cell Nr	Mice Nr
100,000	2/2
10,000	4/4
1,000	3/4
100	3/10
10	0/10
TIC Frequency: 1/495 *(95% C.I.: 1/197-1/1245)*	

MDA-PATX50

Cell Nr	Mice Nr
100,000	2/2
10,000	3/4
1,000	2/4
100	0/10
10	0/10
TIC Frequency: 1/4848 *(95% C.I.: 1/687-1/13936)*	

MDA-PATX66

Cell Nr	Mice Nr
100,000	2/2
10,000	3/4
1,000	0/4
100	0/10
10	0/10
TIC Frequency: 1/9138 *(95% C.I.: 1/2921-1/28583)*	

Figure 18. Pancreatic cancer TIC frequency estimation by limiting dilution assay
Extreme limiting dilution analysis (ELDA) in PDAC mouse models (p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper) and PDX-derived cells (MDA-PATX43, MDA-PATX50, MDA-PATX53, MDA-PATX66). Tumor-initiating cell (TIC) frequencies and 95% confidence intervals were determined by the Extreme Limiting Dilution Analysis (ELDA) software.[202]

3. In vivo shRNA screens of new epigenetic vulnerabilities in human and mouse PDAC models

To identify candidate epigenetic mechanisms required for PDAC growth and survival, we performed *in vivo* shRNA screens in two PDAC xenograft-derived samples (MDA-PATX53 and MDA-PATX43) and three PDAC GEM tumor models (p48-Cre, Kras$^{G12D_LSL/+,}$ Trp53$^{L/L}$; p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a/Arf$^{L/L}$ and p48-Cre, Kras$^{G12D_LSL/+}$senescence-escaper) (**Fig.19a**).

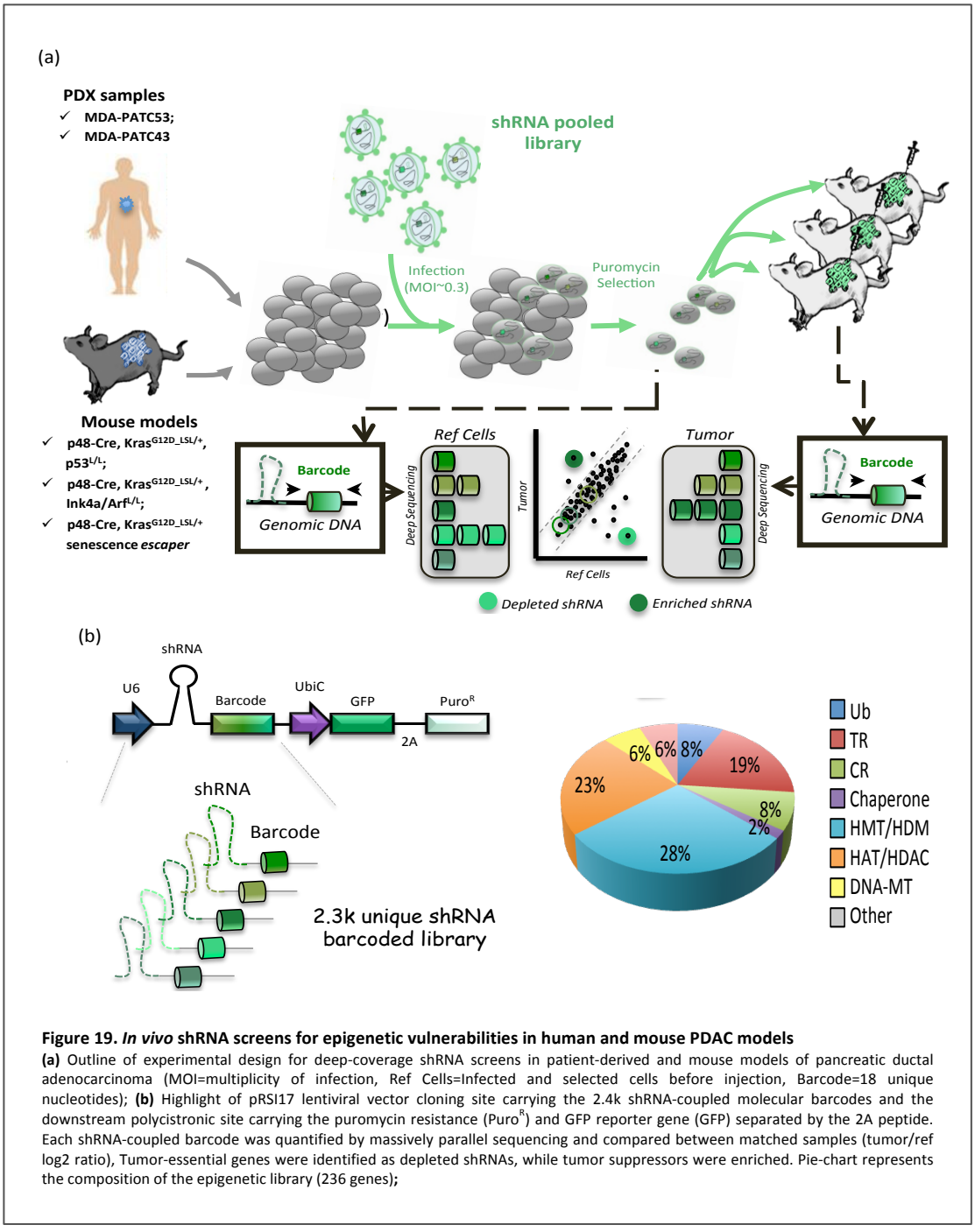


Figure 19. *In vivo* shRNA screens for epigenetic vulnerabilities in human and mouse PDAC models
(a) Outline of experimental design for deep-coverage shRNA screens in patient-derived and mouse models of pancreatic ductal adenocarcinoma (MOI=multiplicity of infection, Ref Cells=Infected and selected cells before injection, Barcode=18 unique nucleotides); **(b)** Highlight of pRSI17 lentiviral vector cloning site carrying the 2.4k shRNA-coupled molecular barcodes and the downstream polycistronic site carrying the puromycin resistance (Puro^{R}) and GFP reporter gene (GFP) separated by the 2A peptide. Each shRNA-coupled barcode was quantified by massively parallel sequencing and compared between matched samples (tumor/ref log2 ratio), Tumor-essential genes were identified as depleted shRNAs, while tumor suppressors were enriched. Pie-chart represents the composition of the epigenetic library (236 genes);

The workflow design was adjusted with the purpose of performing the shRNA *in vivo* screens in exactly the same experimental setting previously applied for the TIC coverage studies. We developed an shRNA library targeting 236 human or mouse epigenetic regulators. To enhance the robustness of the screen and limit the potential for non-specific off-target activity we employed 10 unique shRNAs per each gene in the library. Each shRNA, constituted by 2 G/U mismatches in the passenger strand, a 7 nucleotides loop and a 21 nucleotides targeting sequence, was cloned into the pRSI16 lentiviral vector (Cellecta) carrying a downstream polycistronic site with the puromycin resistance ($Puro^R$) and GFP reporter gene (GFP) separated by the 2A peptide. The oligo corresponding to each shRNA was synthesized with a unique molecular barcode (18 nucleotides) for measuring representation by NGS (**Fig.19b**).

The infection steps were performed mimicking the MOI optimized conditions (0.3 TU/cell in human, 1.2 TU/cell in mouse) we determined in each sample with the 12,500-barcode library during the TIC coverage studies. The PCR amplification of the barcodes and the sequencing data deconvolution and normalization were performed in accordance with the methods explained above. We screened the PDX-derived MDA-PATX53 and MDA-PATX43 cells in replicate with ≈2,000 cells/shRNA to ensure that library complexity was covered in each transplanted mouse (**Fig.20a,b**). As suggested by the TIC coverage study with 12,500 barcodes, the chosen shRNA library would not be maintained in the MDA-PATX66 tumors with a coverage of ≈2,000 cells/shRNA (**Fig.20c**). Replicate screens in the PDAC GEM models were performed with ≈400 cells/shRNA (**Fig.21a-c**).

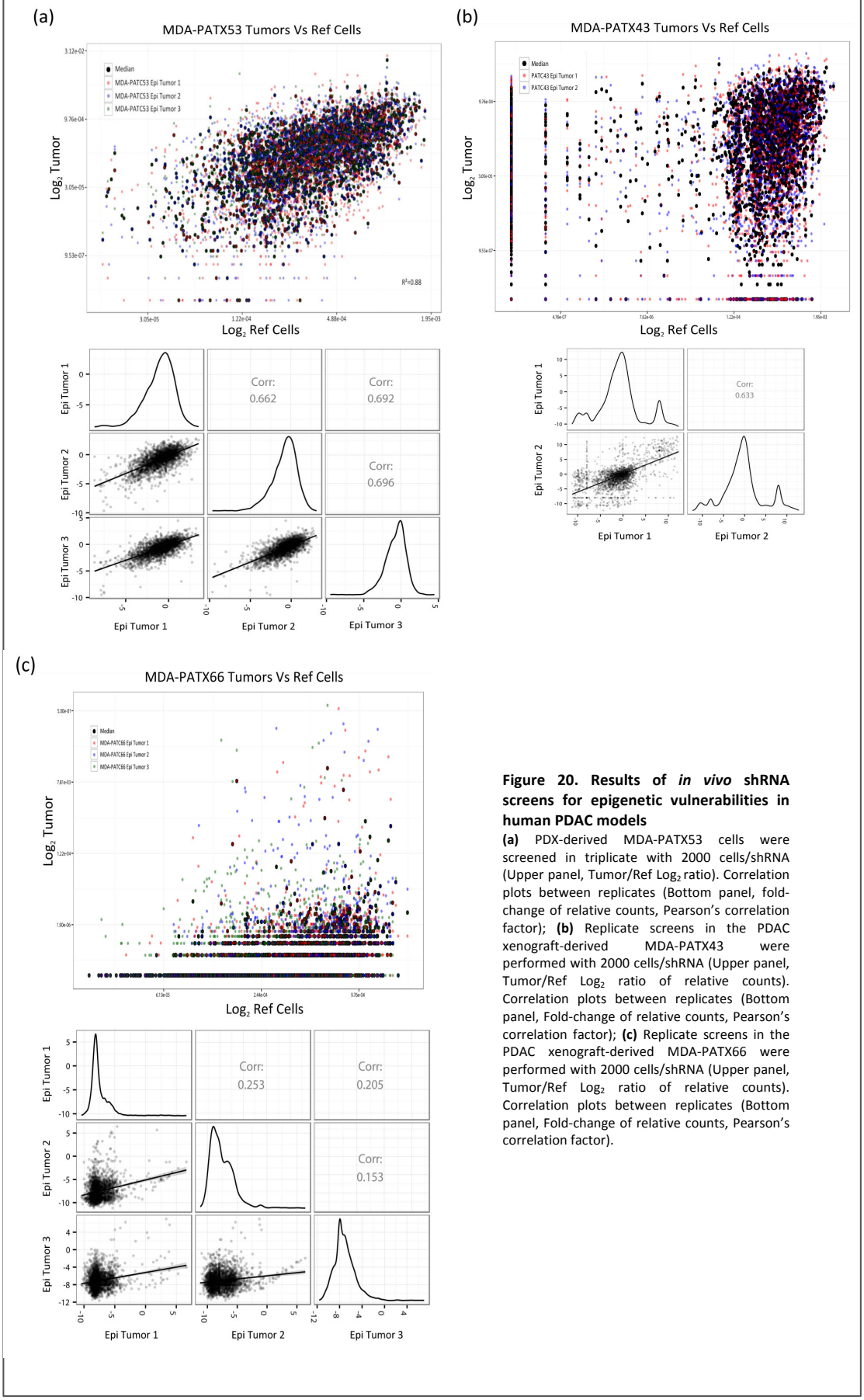


Figure 20. Results of *in vivo* shRNA screens for epigenetic vulnerabilities in human PDAC models

(a) PDX-derived MDA-PATX53 cells were screened in triplicate with 2000 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio). Correlation plots between replicates (Bottom panel, fold-change of relative counts, Pearson's correlation factor); **(b)** Replicate screens in the PDAC xenograft-derived MDA-PATX43 were performed with 2000 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio of relative counts). Correlation plots between replicates (Bottom panel, Fold-change of relative counts, Pearson's correlation factor); **(c)** Replicate screens in the PDAC xenograft-derived MDA-PATX66 were performed with 2000 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio of relative counts). Correlation plots between replicates (Bottom panel, Fold-change of relative counts, Pearson's correlation factor).

The robustness of this screening approach was demonstrated comparing each replicate and calculating the associated Pearson's correlation factor. Among all the performed screens, only the MDA-PATX66 replicates showed a very low correlation (R^2=0.15-0.25) (**Fig.20-21**).

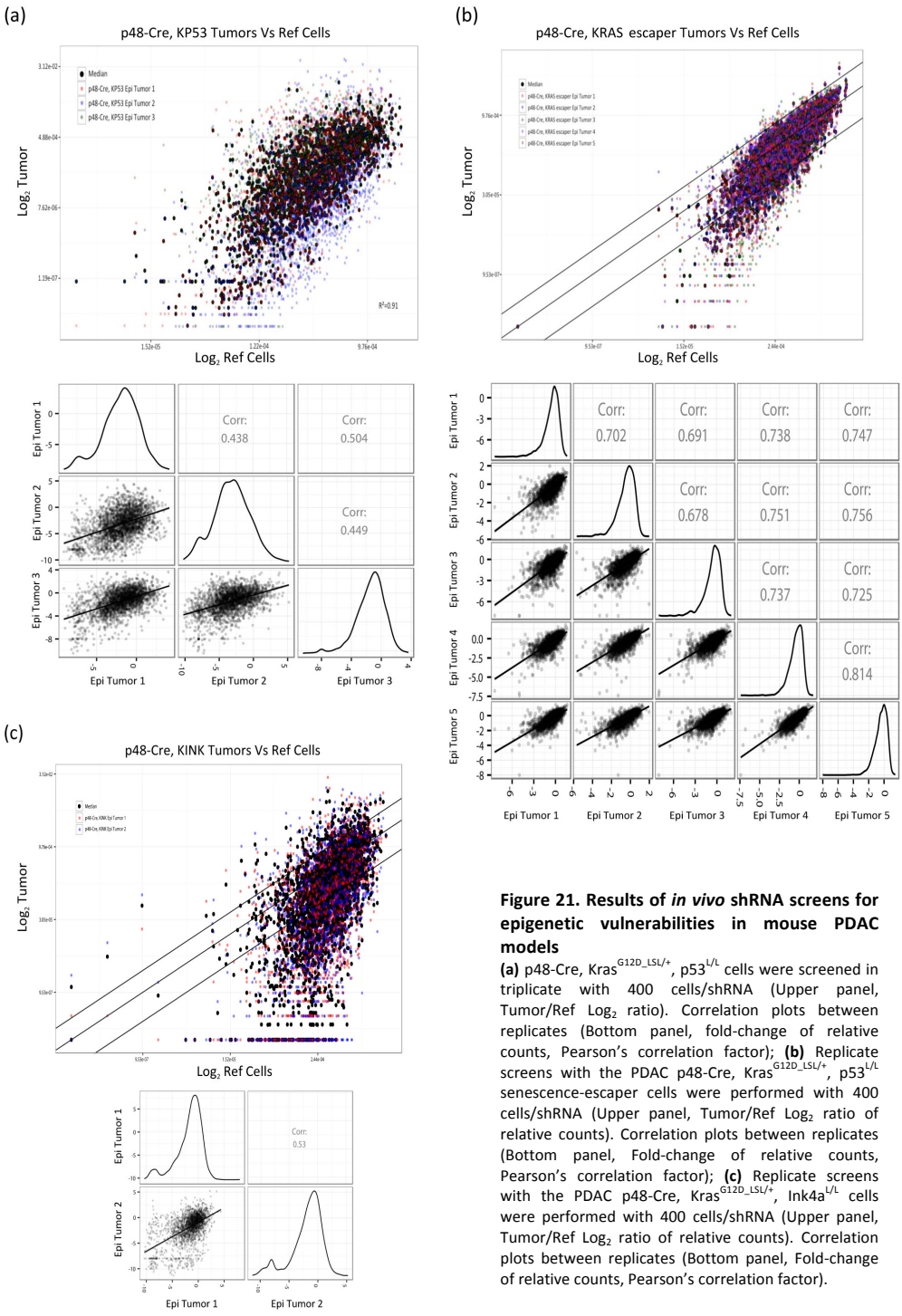


Figure 21. Results of *in vivo* shRNA screens for epigenetic vulnerabilities in mouse PDAC models

(a) p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ cells were screened in triplicate with 400 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio). Correlation plots between replicates (Bottom panel, fold-change of relative counts, Pearson's correlation factor); **(b)** Replicate screens with the PDAC p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ senescence-escaper cells were performed with 400 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio of relative counts). Correlation plots between replicates (Bottom panel, Fold-change of relative counts, Pearson's correlation factor); **(c)** Replicate screens with the PDAC p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$ cells were performed with 400 cells/shRNA (Upper panel, Tumor/Ref Log_2 ratio of relative counts). Correlation plots between replicates (Bottom panel, Fold-change of relative counts, Pearson's correlation factor).

To confirm an accurate coverage of each shRNA in the performed screens, we applied a cumulative distribution function (CDF) to the Tumor/Ref ratio as a first-line filter to detect the top 15-30% depleted shRNAs (**Fig.22a**).

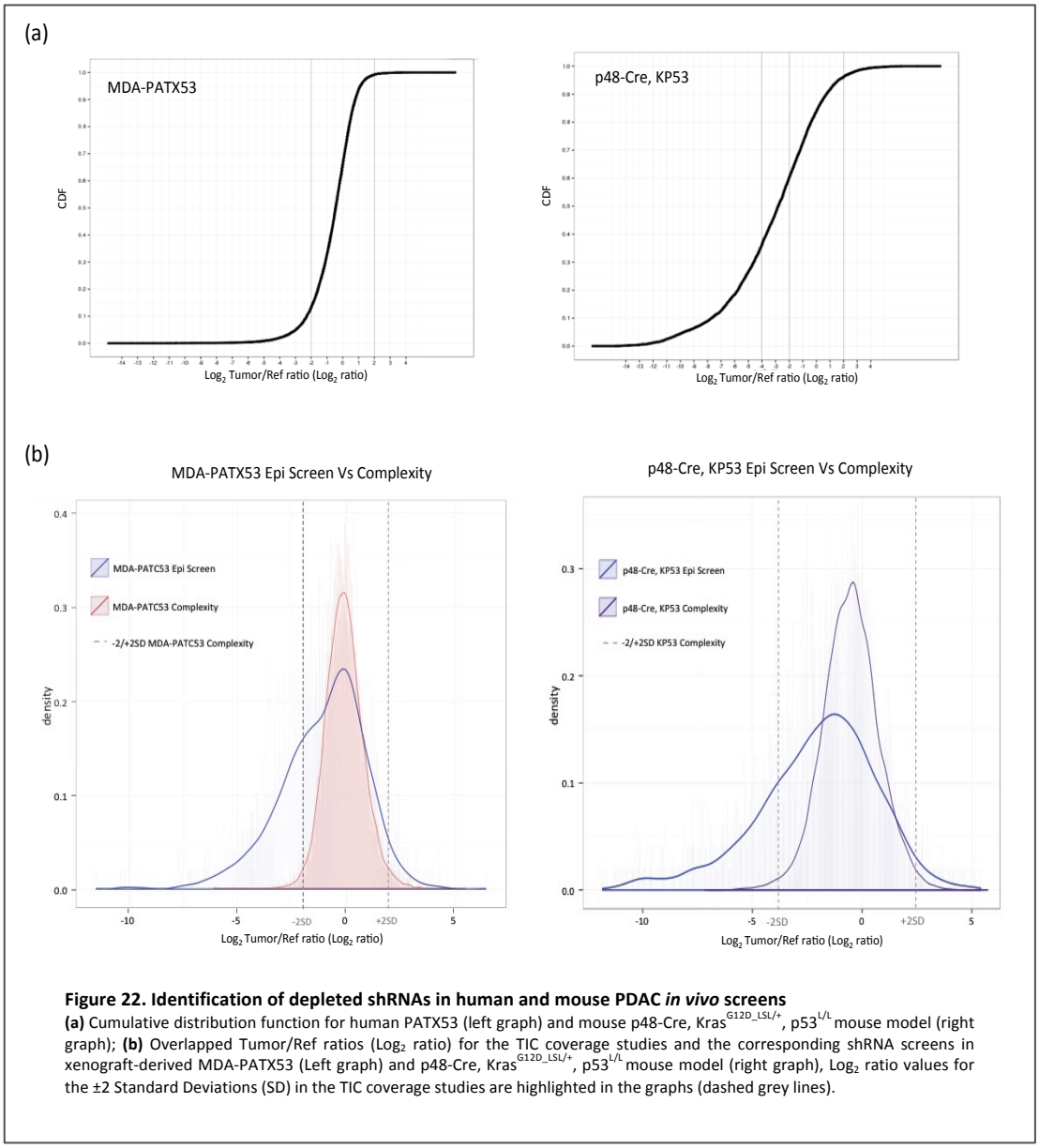


Figure 22. Identification of depleted shRNAs in human and mouse PDAC *in vivo* screens
(a) Cumulative distribution function for human PATX53 (left graph) and mouse p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ mouse model (right graph); **(b)** Overlapped Tumor/Ref ratios (Log_2 ratio) for the TIC coverage studies and the corresponding shRNA screens in xenograft-derived MDA-PATX53 (Left graph) and p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ mouse model (right graph), Log_2 ratio values for the ±2 Standard Deviations (SD) in the TIC coverage studies are highlighted in the graphs (dashed grey lines).

For each sample, the cutoff (-2/-4 Log_2) determined by the CDF analysis demonstrated to match, with good approximation, the Log_2 value of the -2 Standard Deviation (-2SD) in the corresponding TIC coverage study (**Fig.22b**).

To identify the top-scoring "hits" in each screened PDAC model we associated robust Z-scores to each gene, taking into account the fold-change of top 3 shRNAs, and ranked them on the base of the corresponding p-value (**Fig.23**).

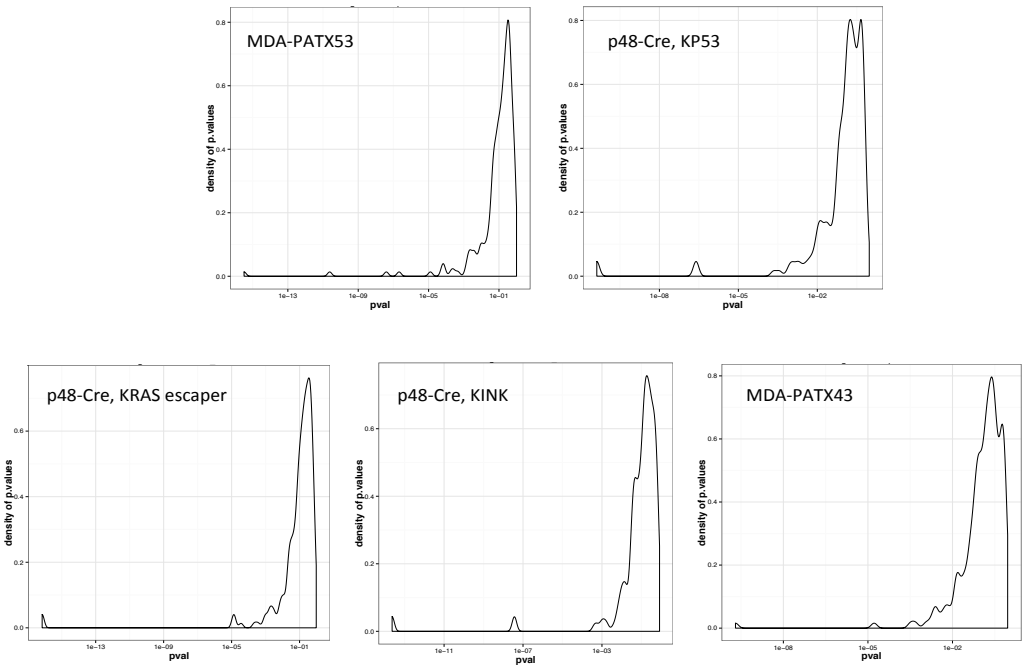


Figure 23. Identification of depleted genes in human and mouse PDAC *in vivo* screens
Density plot of the p-value associated to the mean Z-score (fold-change, average of the top 3 shRNAs) for each gene in MDA-PATX53 and MDA.PATX43 human PDAC screens and p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$, p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ senescence-escaper and p48-Cre, $Kras^{G12D\_LSL/+}$, $Ink4a^{L/L}$ mouse PDAC screens;

A p-value associated to the mean Z-score among all the screening models has been applied to rank the most potent genes across all of these PDAC relevant contexts (**Fig.24**).

2 well-known essential genes (PSMA1, RLP30) were inserted into the library as positive controls and they scored among the top depleted hits. Interestingly, one of the top scoring genes, PHF5A, has been proposed as a relevant key player in selectively sustaining RNA splicing and survival of glioblastoma stem cells (GSCs) through a genome-wide RNAi screening approach[204] .

Among the cohesins (SMC1a, SMC3) and condensins (SMC2, SMC4), SMC2 confirmed a prominent role in sustaining PDAC cell proliferation, in line with previous findings[205]. In the

same direction, we identified BRD4 in our top-scoring gene list as expected by the very-well demonstrated PDAC cells sensitivity to BET inhibitors and BRD4 interference[206] (**Fig.24**).

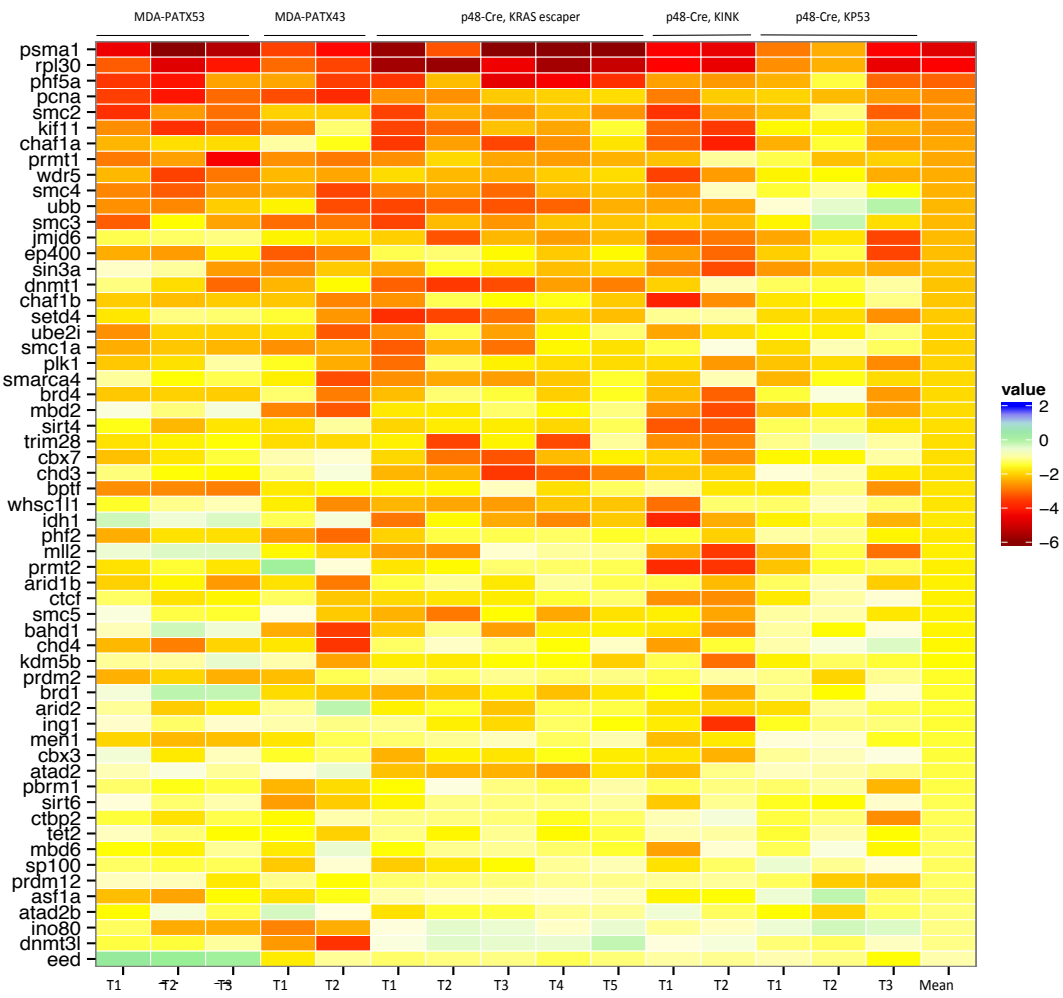


Figure 24. Top-scoring epigenes in human and mouse PDAC in vivo screens
Heatmap of the top-scoring hits identified applying a p-value-based cut-off ($p<0.05$) associated to the z-score (fold-change, average of the top 3 shRNAs) for each gene. Genes were ranked calculating the mean z-score among all the screened samples (T = tumor).

In addition, an unsupervised clustering method, applied to all the genes in the library, demonstrated not only the robustness of replicates from each PDAC model, but also the possibility to inform on functional epigenetic networks in PDAC and identifying patient-specific dependent vulnerabilities (**Fig.25**). As expected, positive controls clustered together and also leaded a very well defined sub-group constituted by the most potent hits among all the PDAC screened models. Negative controls clustered together as well and were at the top of a bigger cluster containing genes that were inactive and genes that showed significant depletion only in specific PDAC sub-sets.

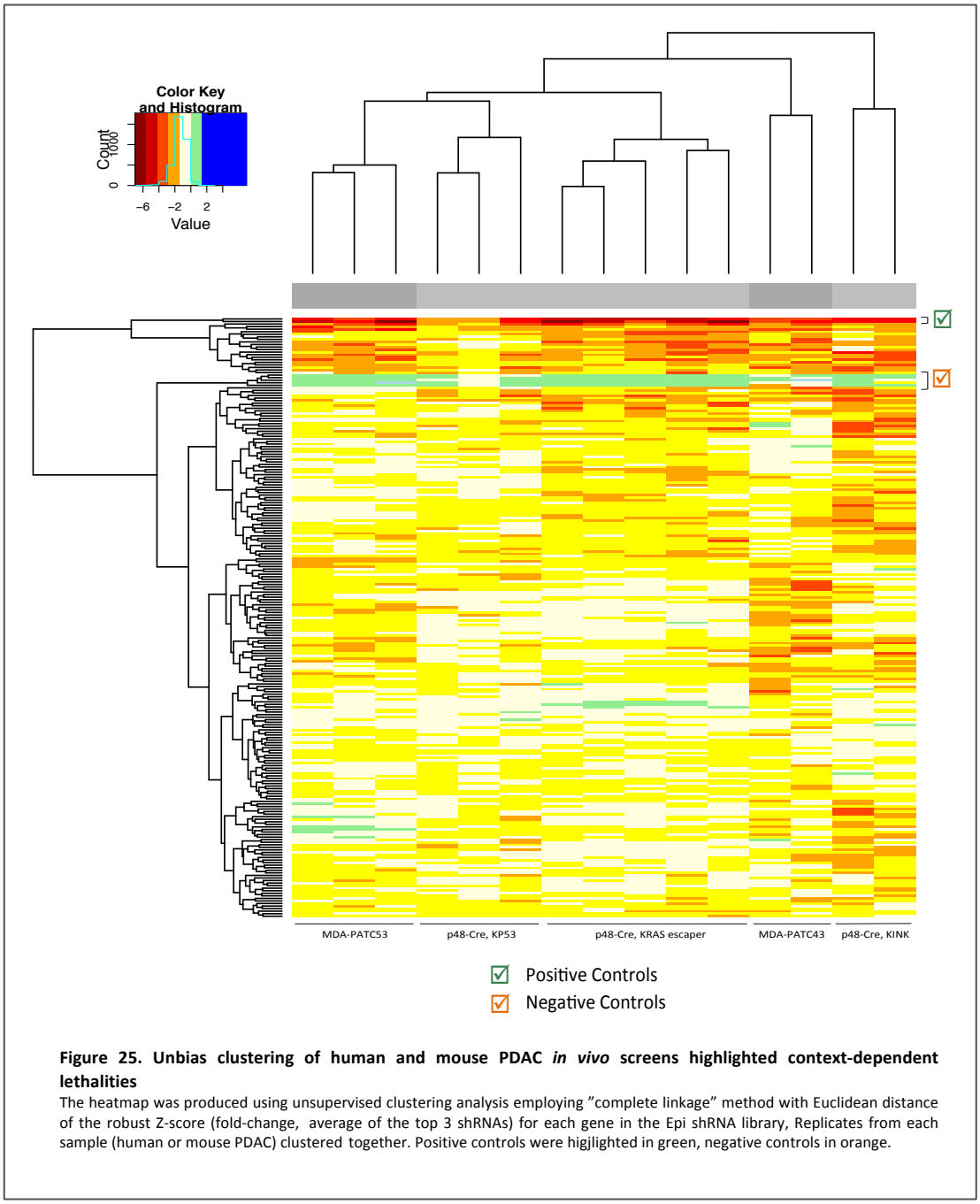


Figure 25. Unbias clustering of human and mouse PDAC *in vivo* screens highlighted context-dependent lethalities

The heatmap was produced using unsupervised clustering analysis employing "complete linkage" method with Euclidean distance of the robust Z-score (fold-change, average of the top 3 shRNAs) for each gene in the Epi shRNA library, Replicates from each sample (human or mouse PDAC) clustered together. Positive controls were higjlighted in green, negative controls in orange.

Differently high-ranked common hits (PHF5A, SMC2, WDR5) were individually validated with 2 independent shRNA, aiming at linking target knock-down level (72 hours after infection), in vivo xenotransplantation and in vitro colony formation assay (CFA). Down-regulation of these genes, confirmed by western blot, significantly impaired new colony formation and tumor growth in both human (**Fig.26a**) and mouse PDAC models (**Fig.26b**).

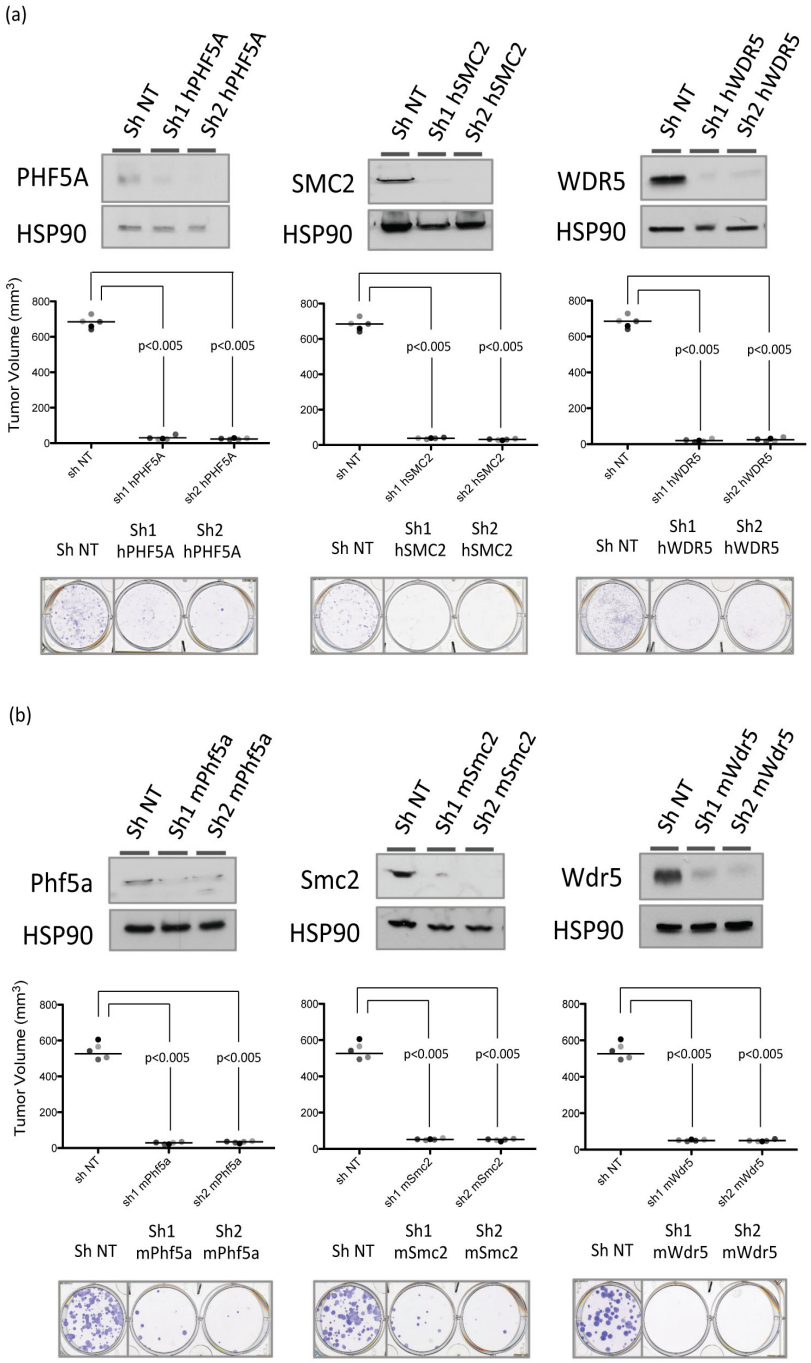


Figure 26. Validation of the top-scoring hits identified by the *in vivo* shRNA screens for epigenetic vulnerabilities in human and mouse PDAC models

(a) Western blot, xenotransplantation tumor size (mm^3) and colony formation assay for xenograft-derived PATX53 cells infected with 2 independent shRNAs against the top-scoring hits: WDR5, PHF5A and SMC2 (h=human); **(b)** Western blot, xenotransplantation tumor size (mm^3) and colony formation assay for the mouse p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ model infected with 2 independent shRNAs against the top-scoring hits: Wdr5, Phf5a and Smc2 (m=mouse).

4. WDR5 is essential for PDAC establishment and maintenance

Unbiased in vivo shRNA screens provide the opportunity to inform on novel genetic associations and to query dependencies across common pathways, nodes, networks or even multiprotein complexes in a disease-relevant context. One of the strongest "hits" to emerge across the multiple PDAC screens was the WD40 protein WDR5, a core member of the COMPASS histone H3 Lys4 (H3K4) methyltransferase complex[44,207,208].

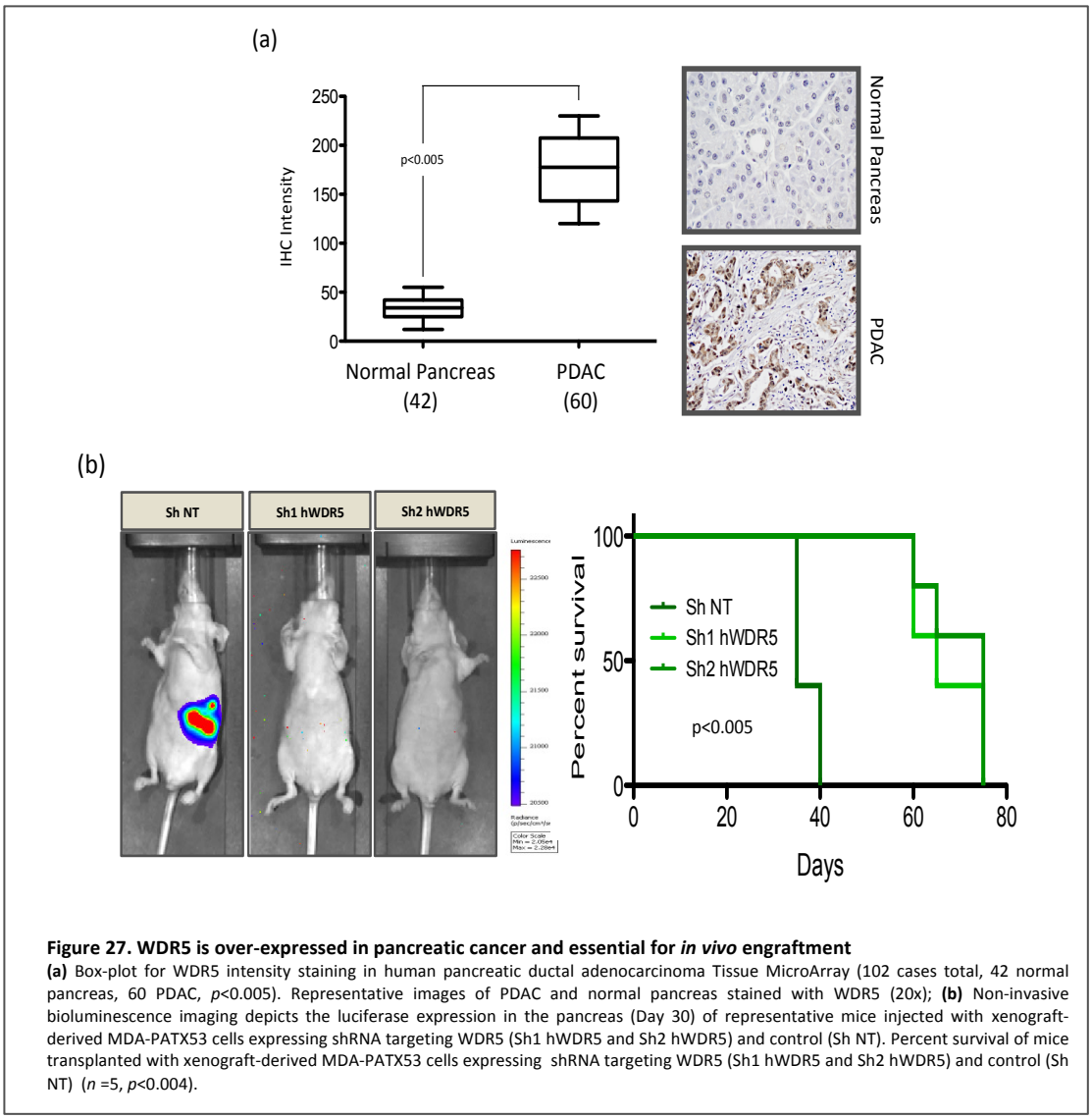


Figure 27. WDR5 is over-expressed in pancreatic cancer and essential for *in vivo* engraftment
(a) Box-plot for WDR5 intensity staining in human pancreatic ductal adenocarcinoma Tissue MicroArray (102 cases total, 42 normal pancreas, 60 PDAC, $p<0.005$). Representative images of PDAC and normal pancreas stained with WDR5 (20x); **(b)** Non-invasive bioluminescence imaging depicts the luciferase expression in the pancreas (Day 30) of representative mice injected with xenograft-derived MDA-PATX53 cells expressing shRNA targeting WDR5 (Sh1 hWDR5 and Sh2 hWDR5) and control (Sh NT). Percent survival of mice transplanted with xenograft-derived MDA-PATX53 cells expressing shRNA targeting WDR5 (Sh1 hWDR5 and Sh2 hWDR5) and control (Sh NT) (n =5, $p<0.004$).

Recently, WDR5 up-regulation was detected in prostate and bladder cancers, where it also demonstrated to be critical for cancer cells proliferation[209,210]. We first confirmed deregulated expression of WDR5 in human PDAC compared to normal control pancreas. A

Tissue MicroArray (TMA) representing 102 specimens (46 normal pancreas, 60 PDAC) was stained for WDR5 expression and an intensity score (0-300, percentage of positive cells by the intensity) was assigned by a pathologist to each sample (**Fig.27a**). PDAC specimens showed a significant higher level of WDR5 expression in comparison with the normal pancreas (Student's T-test, $p<0.005$). To functionally validate the screen results, PDX-derived PDAC cells transduced with individual shRNAs targeting WDR5 were orthotopically implanted in the pancreas of NSG mice. In agreement with the above results, WDR5 knock-down substantially delayed tumor growth and extended survival compared to non-targeting (NT) shRNA controls (**Fig.27b**).

These effects seemed specific as expression of a cDNA encoding the WDR5 coding sequence (ORF) lacking the 3' UTR nucleotides rescued the observed effects upon targeting by the shRNA (**Fig.28a,b**).

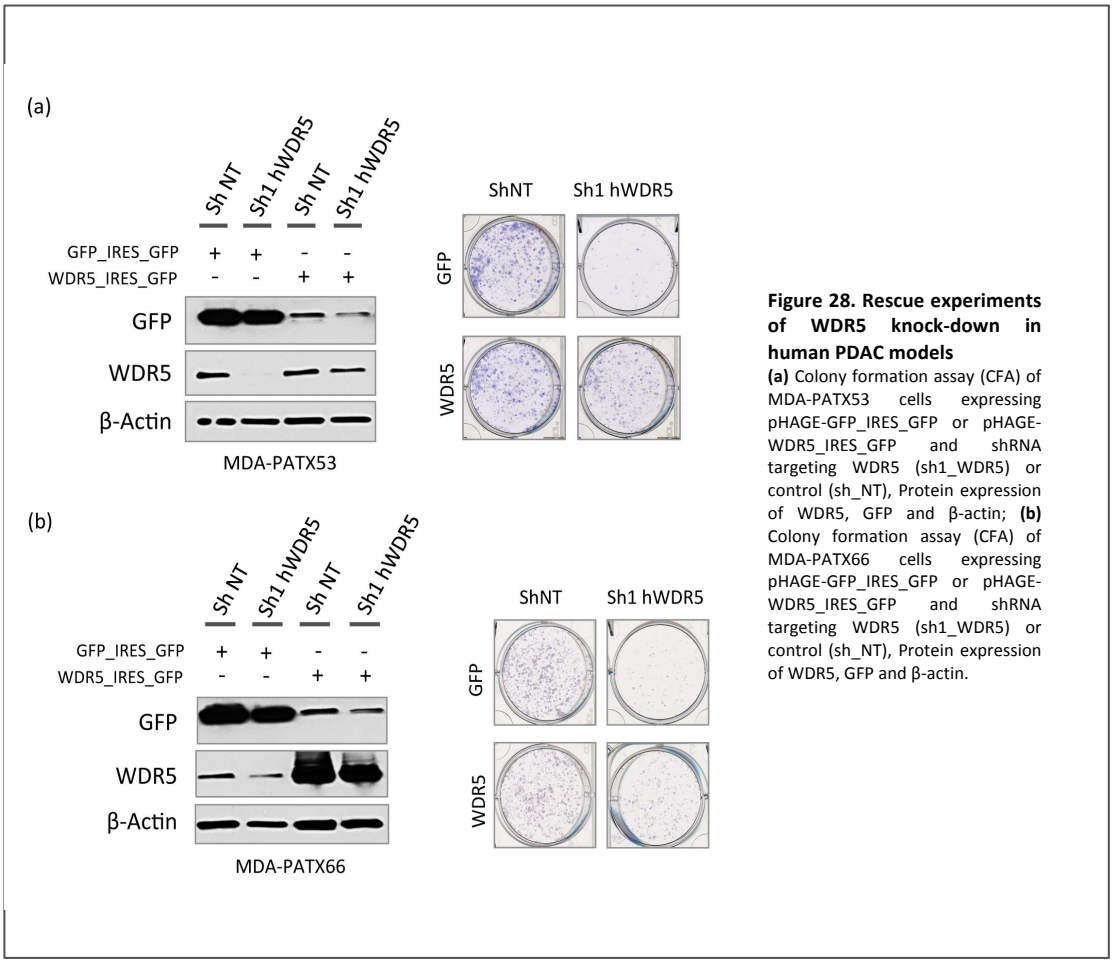


Figure 28. Rescue experiments of WDR5 knock-down in human PDAC models
(a) Colony formation assay (CFA) of MDA-PATX53 cells expressing pHAGE-GFP_IRES_GFP or pHAGE-WDR5_IRES_GFP and shRNA targeting WDR5 (sh1_WDR5) or control (sh_NT), Protein expression of WDR5, GFP and β-actin; **(b)** Colony formation assay (CFA) of MDA-PATX66 cells expressing pHAGE-GFP_IRES_GFP or pHAGE-WDR5_IRES_GFP and shRNA targeting WDR5 (sh1_WDR5) or control (sh_NT), Protein expression of WDR5, GFP and β-actin.

Using a serum-free 3D culture, we were able to generate pancreatic cancer spheres from either human or mouse PDAC samples[211]. WDR5 knock-down significantly impaired the spherogenic potential of these cultures as demonstrated by calcein staining and spheroid counts (**Fig.29a,d**).

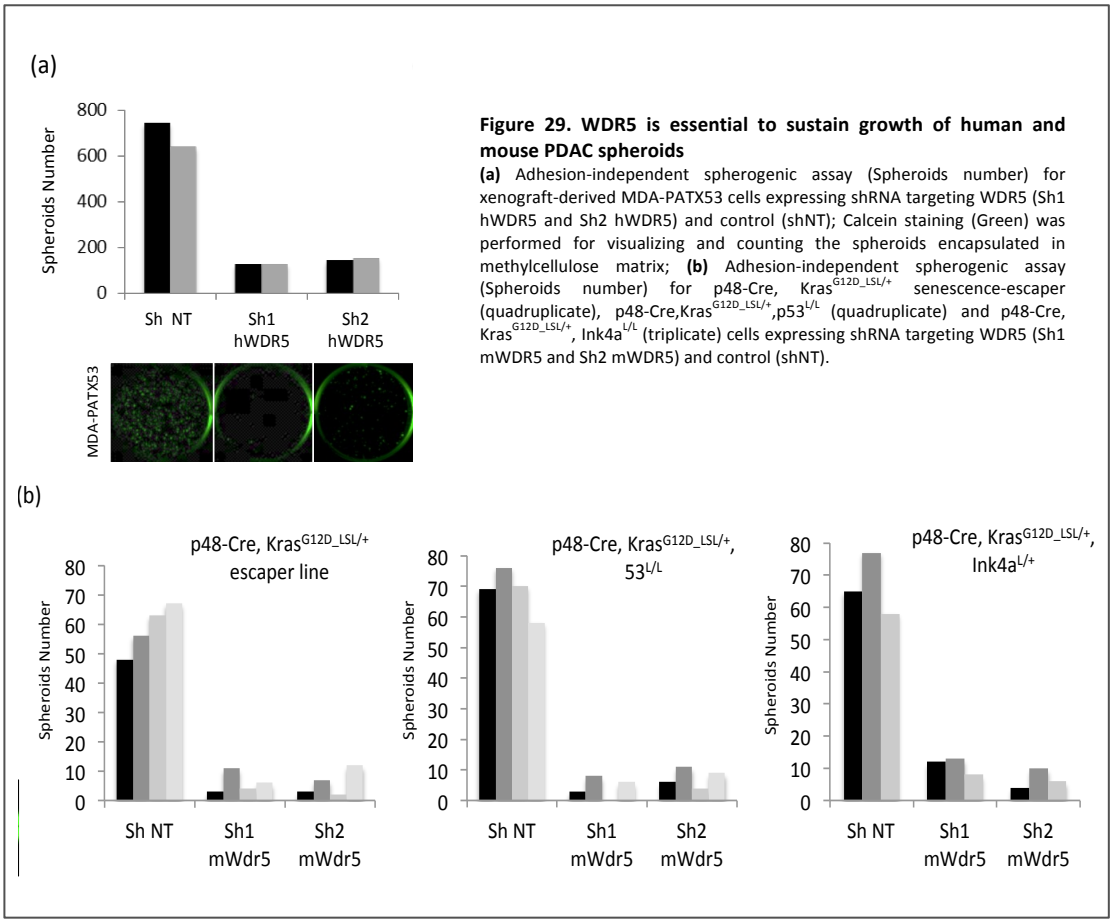


Figure 29. WDR5 is essential to sustain growth of human and mouse PDAC spheroids

(a) Adhesion-independent spherogenic assay (Spheroids number) for xenograft-derived MDA-PATX53 cells expressing shRNA targeting WDR5 (Sh1 hWDR5 and Sh2 hWDR5) and control (shNT); Calcein staining (Green) was performed for visualizing and counting the spheroids encapsulated in methylcellulose matrix; **(b)** Adhesion-independent spherogenic assay (Spheroids number) for p48-Cre, Kras$^{G12D_LSL/+}$ senescence-escaper (quadruplicate), p48-Cre,Kras$^{G12D_LSL/+}$,p53$^{L/L}$ (quadruplicate) and p48-Cre, Kras$^{G12D_LSL/+}$, Ink4a$^{L/L}$ (triplicate) cells expressing shRNA targeting WDR5 (Sh1 mWDR5 and Sh2 mWDR5) and control (shNT).

Next, we aimed to determine if WDR5 was required for the in vivo growth of additional patient-derived PDAC models. Using our optimized ex vivo rapid culture conditions we infected stabilized PDC cells with 2 independent shRNA for WDR5 or control. Target knock-down was confirmed 48 hours after puromycin selection and NSG mice were transplanted for each condition (n=5). Measuring the tumor size we were able to confirm that WDR5 knock-down blocked tumor growth of transplanted PDAC samples (**Fig.30**).

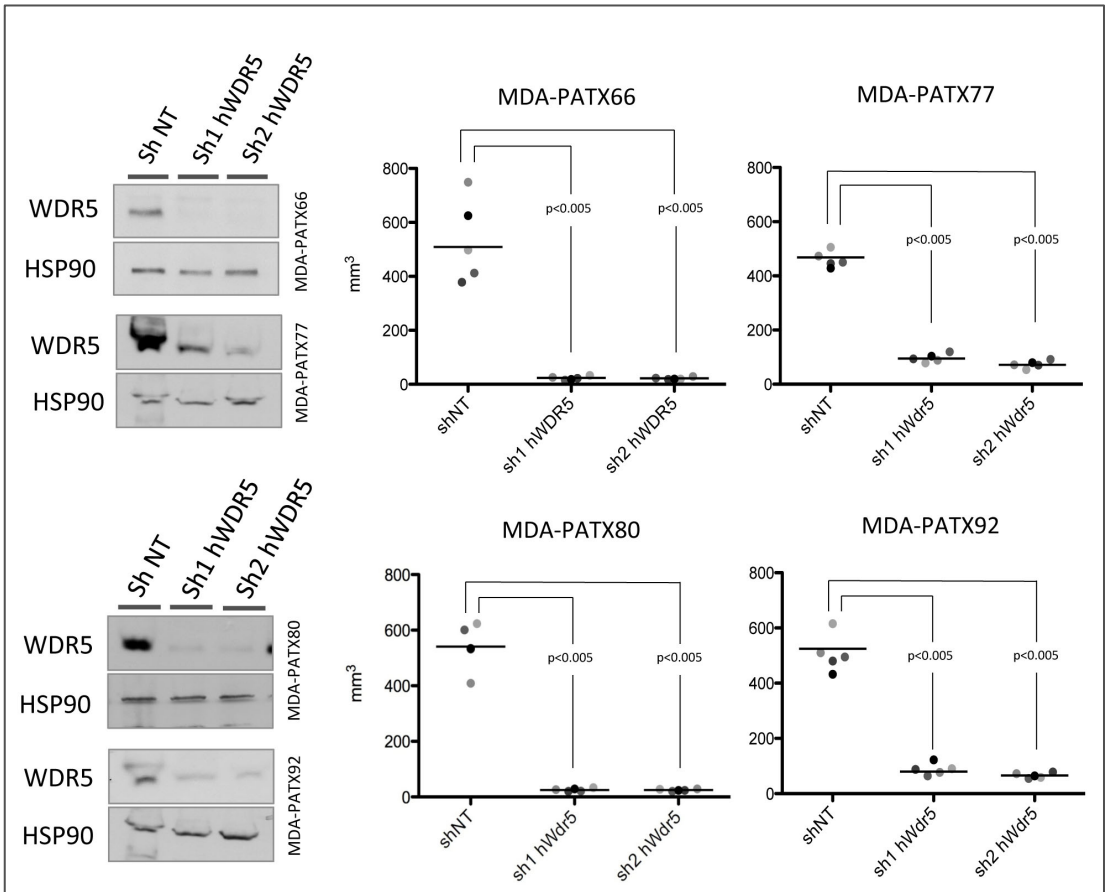


Figure 30. WDR5 knock-down impairs the engraftment potential of human xenograft-derived PDAC models
Western blot for xenograft-derived PDAC cells (MDA-PATX66, MDA-PATX77, MDA-PATX80, MDA-PATX92) infected with 2 independent shRNAs against WDR5 (Sh1 hWDR5 and Sh2 hWDR5) or control (Sh NT), Protein expression of WDR5 ans HSP90; Tumor measurement (tumor size, mm^3) of 3 primary PDAC xenograft-derived cells (MDA-PATX77, MDA-PATX80, MDA-PATX92) expressing shRNAs targeting WDR5 (sh1_WDR5, sh2_WDR5), and control (sh NT), Median (n = 5, p< 0.005).

To evaluate whether WDR5 would provide an essential function in tumor maintenance, we infected PDX-derived samples (MDA-PATX53 and MDA-PATX66) with Tet-inducible WDR5 shRNAs (Sh1 WDR5i, Sh2 WDR5i) or control shRNA (Sh NT) and sorted them for the top 20% RFP-positive cells. WDR5-inducible cells were grown using Tet-free culturing media and the protein knock-down was confirmed after doxycycline addition (+ DOX) by western blot (**Fig.31a-c**). Applying colony formation assay and supplementing with doxycycline upon cell seeding (+ DOX), we observed a significant arrest in colony growth, not detectable when colonies were kept in absence of doxycycline (- DOX) (**Fig.32a**). The same WDR5-inducible PDAC cells were transplanted into host mice and preliminarily monitored for their growing ability in the absence of a doxycycline-containing drinking water regimen (n=5). (**Fig.32b**).

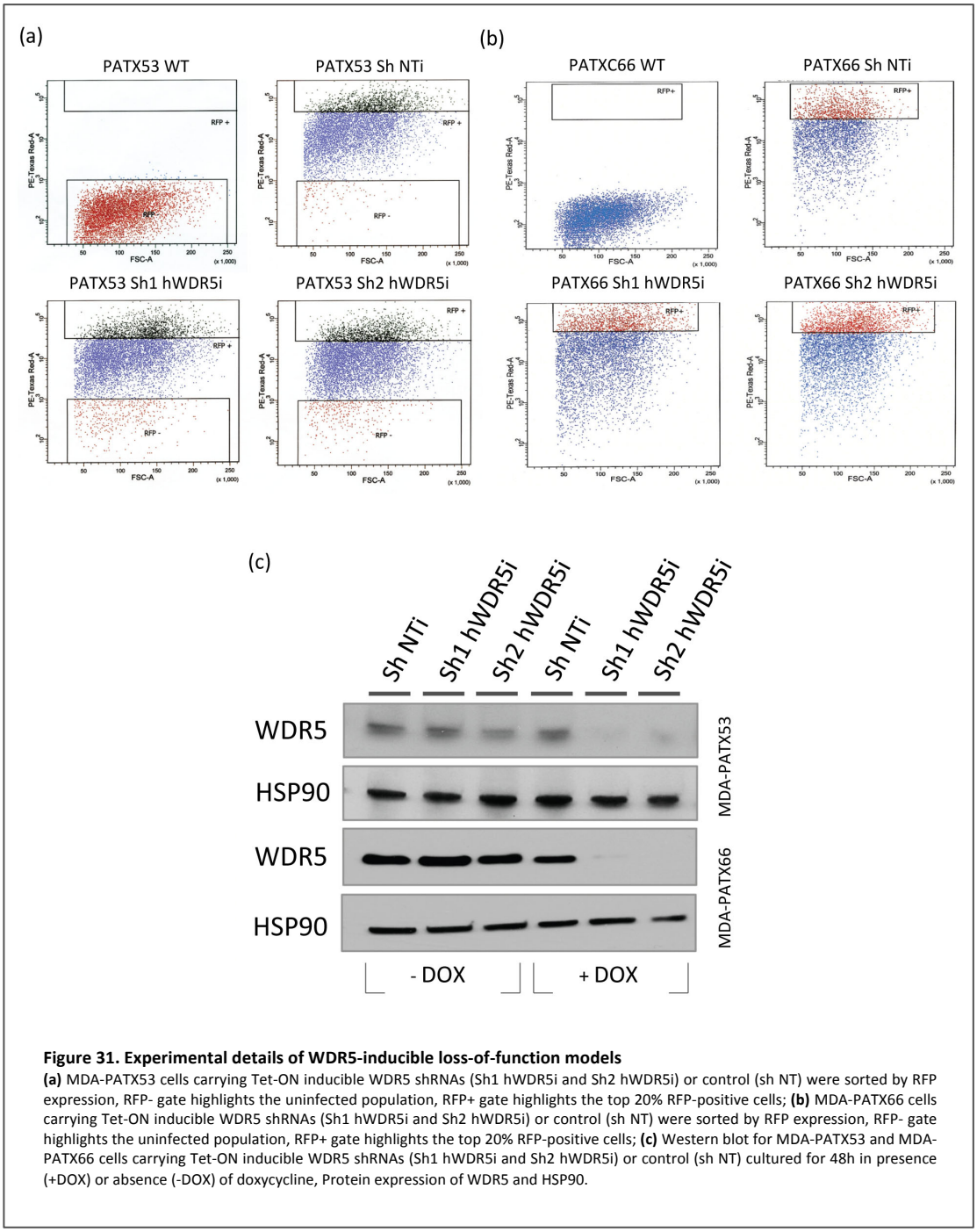


Figure 31. Experimental details of WDR5-inducible loss-of-function models
(a) MDA-PATX53 cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (sh NT) were sorted by RFP expression, RFP- gate highlights the uninfected population, RFP+ gate highlights the top 20% RFP-positive cells; **(b)** MDA-PATX66 cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (sh NT) were sorted by RFP expression, RFP- gate highlights the uninfected population, RFP+ gate highlights the top 20% RFP-positive cells; **(c)** Western blot for MDA-PATX53 and MDA-PATX66 cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (sh NT) cultured for 48h in presence (+DOX) or absence (-DOX) of doxycycline, Protein expression of WDR5 and HSP90.

We observed a significant arrest in the growth ability of the WDR5 down-regulated tumors under doxycycline treatment (+ DOX) that was maintained until the end of the study (Student's T-test, $p<0.05$) (**Fig.32c**). Immunohistochemistry staining confirmed *in vivo* WDR5 knock-down associated with a reduction of the proliferation marker Ki67 (**Fig.32d**).

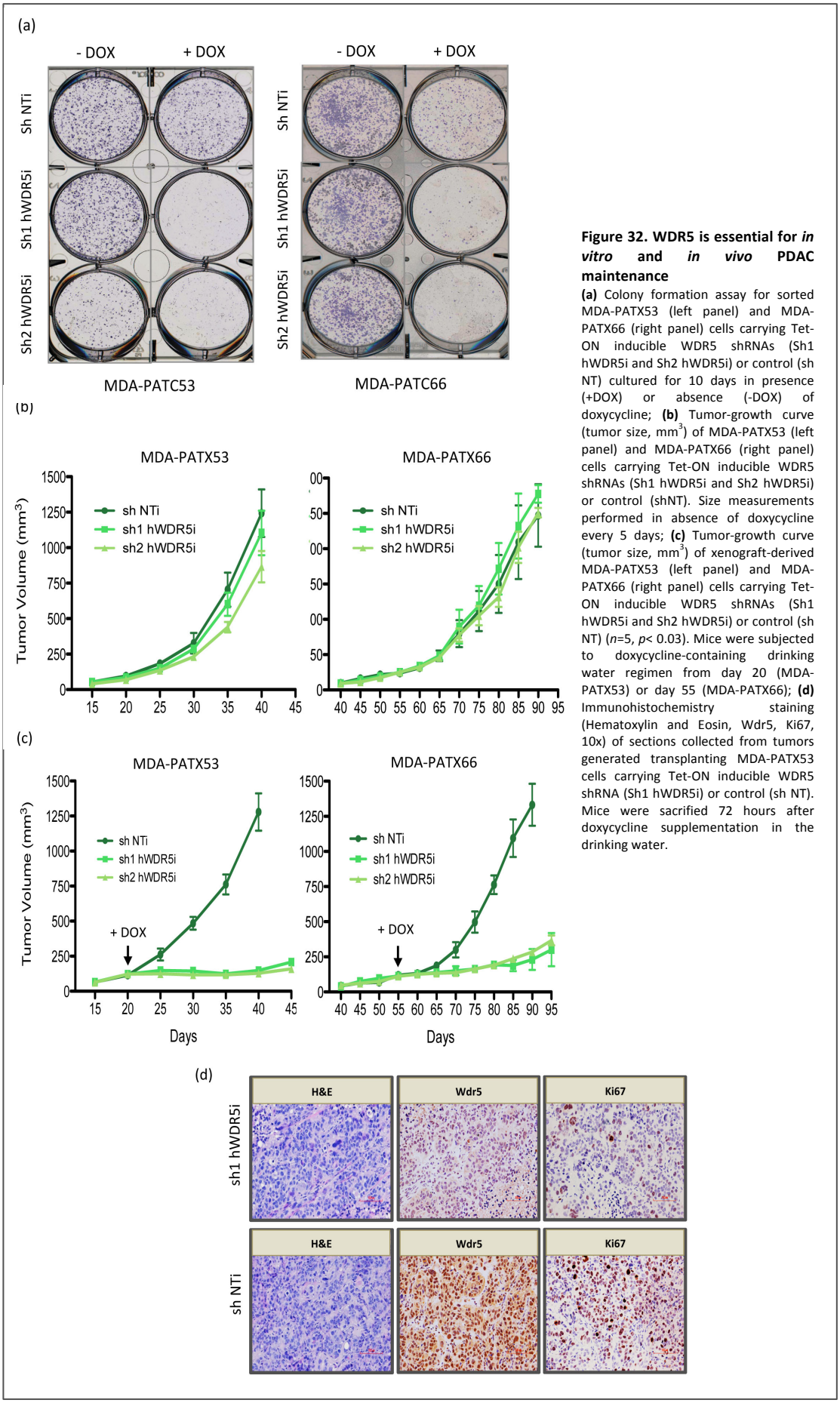


Figure 32. WDR5 is essential for *in vitro* and *in vivo* PDAC maintenance

(a) Colony formation assay for sorted MDA-PATX53 (left panel) and MDA-PATX66 (right panel) cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (sh NT) cultured for 10 days in presence (+DOX) or absence (-DOX) of doxycycline; **(b)** Tumor-growth curve (tumor size, mm^3) of MDA-PATX53 (left panel) and MDA-PATX66 (right panel) cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (shNT). Size measurements performed in absence of doxycycline every 5 days; **(c)** Tumor-growth curve (tumor size, mm^3) of xenograft-derived MDA-PATX53 (left panel) and MDA-PATX66 (right panel) cells carrying Tet-ON inducible WDR5 shRNAs (Sh1 hWDR5i and Sh2 hWDR5i) or control (sh NT) (n=5, p< 0.03). Mice were subjected to doxycycline-containing drinking water regimen from day 20 (MDA-PATX53) or day 55 (MDA-PATX66); **(d)** Immunohistochemistry staining (Hematoxylin and Eosin, Wdr5, Ki67, 10x) of sections collected from tumors generated transplanting MDA-PATX53 cells carrying Tet-ON inducible WDR5 shRNA (Sh1 hWDR5i) or control (sh NT). Mice were sacrified 72 hours after doxycycline supplementation in the drinking water.

To demonstrate the role exerted by Wdr5 in autochthonous pancreatic cancer, we developed a Lentiviral based-Somatic-Mosaic system (pLSM5) allowing for tissue specific, time restricted activation of a latent shRNA in a transplantation syngeneic GEM model of PDAC. The system was designed to be modular, where, by combining in one vector the Cre-LoxP and Flpo-Frt technologies, we were able to generate PDAC carrying a latent shRNA allowing for the time restricted acute inactivation of a gene of interest in full blown tumors established from cells transplanted in Rag2$^{-/-}$ immune-compromised mice (**Fig.33a**).

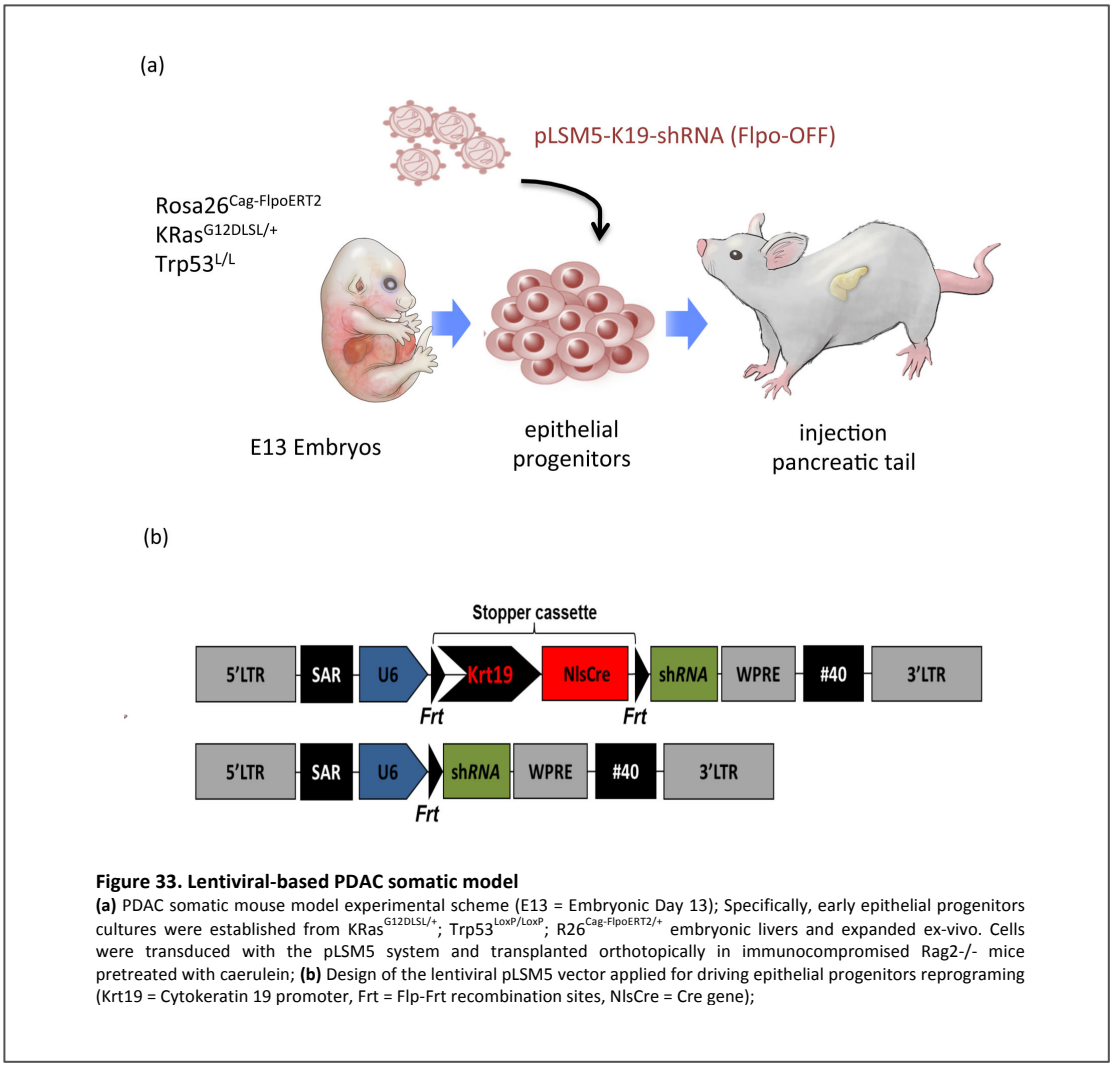


Figure 33. Lentiviral-based PDAC somatic model
(a) PDAC somatic mouse model experimental scheme (E13 = Embryonic Day 13); Specifically, early epithelial progenitors cultures were established from KRas$^{G12DLSL/+}$; Trp53$^{LoxP/LoxP}$; R26$^{Cag\text{-}FlpoERT2/+}$ embryonic livers and expanded ex-vivo. Cells were transduced with the pLSM5 system and transplanted orthotopically in immunocompromised Rag2-/- mice pretreated with caerulein; **(b)** Design of the lentiviral pLSM5 vector applied for driving epithelial progenitors reprograming (Krt19 = Cytokeratin 19 promoter, Frt = Flp-Frt recombination sites, NlsCre = Cre gene);

Specifically, early epithelial progenitors cultures were established from KRas$^{G12DLSL/+}$; Trp53$^{LoxP/LoxP}$; R26$^{Cag\text{-}FlpoERT2/+}$ embryonic livers and expanded ex-vivo. Cells were transduced with the pLSM5 system (where a Frt-stop-Frt cassette containing the Cre recombinase under

the Krt19 promoter was cloned between the U6 promoter and the shRNA) and transplanted orthotopically in immune-compromised $Rag2^{-/-}$ mice pretreated with caerulein (**Fig.33b**). Transplants-derived tumors expressed the epithelial markers 19 and Sox9 and the pancreatic specific marker Pdx1 suggesting that embryonic endodermal progenitors display a remarkable plasticity and are able to originate bona fide pancreatic tumors (**Fig.34a**).

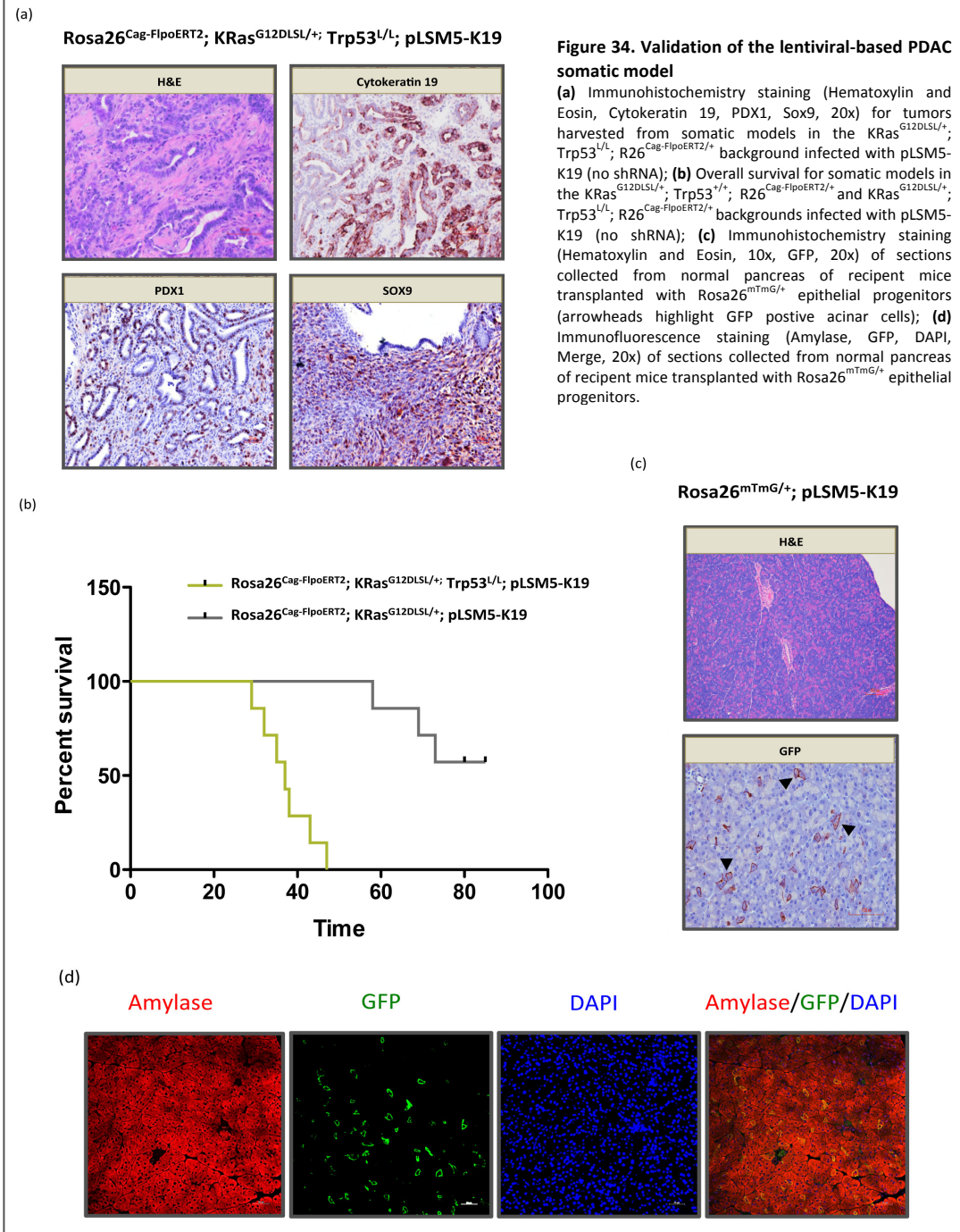


Figure 34. Validation of the lentiviral-based PDAC somatic model
(a) Immunohistochemistry staining (Hematoxylin and Eosin, Cytokeratin 19, PDX1, Sox9, 20x) for tumors harvested from somatic models in the $KRas^{G12DLSL/+}$; $Trp53^{L/L}$; $R26^{Cag\text{-}FlpoERT2/+}$ background infected with pLSM5-K19 (no shRNA); **(b)** Overall survival for somatic models in the $KRas^{G12DLSL/+}$; $Trp53^{+/+}$; $R26^{Cag\text{-}FlpoERT2/+}$ and $KRas^{G12DLSL/+}$; $Trp53^{L/L}$; $R26^{Cag\text{-}FlpoERT2/+}$ backgrounds infected with pLSM5-K19 (no shRNA); **(c)** Immunohistochemistry staining (Hematoxylin and Eosin, 10x, GFP, 20x) of sections collected from normal pancreas of recipent mice transplanted with $Rosa26^{mTmG/+}$ epithelial progenitors (arrowheads highlight GFP postive acinar cells); **(d)** Immunofluorescence staining (Amylase, GFP, DAPI, Merge, 20x) of sections collected from normal pancreas of recipent mice transplanted with $Rosa26^{mTmG/+}$ epithelial progenitors.

Survival curves showed 100% penetrance and a latency around 40 days for KRas$^{G12DLSL/+}$; Trp53$^{LoxP/LoxP}$; R26$^{Cag\text{-}FlpoERT2/+}$ mice, incomplete penetrance and prolonged survival was instead detected for KRas$^{G12DLSL/+}$; R26$^{Cag\text{-}FlpoERT2/+}$ mice (**Fig.34b**). The mT/mG is a double-fluorescent Cre reporter mouse that expresses Tomato reporter (mT) prior to Cre-mediated excision and GFP reporter (mG) after excision.

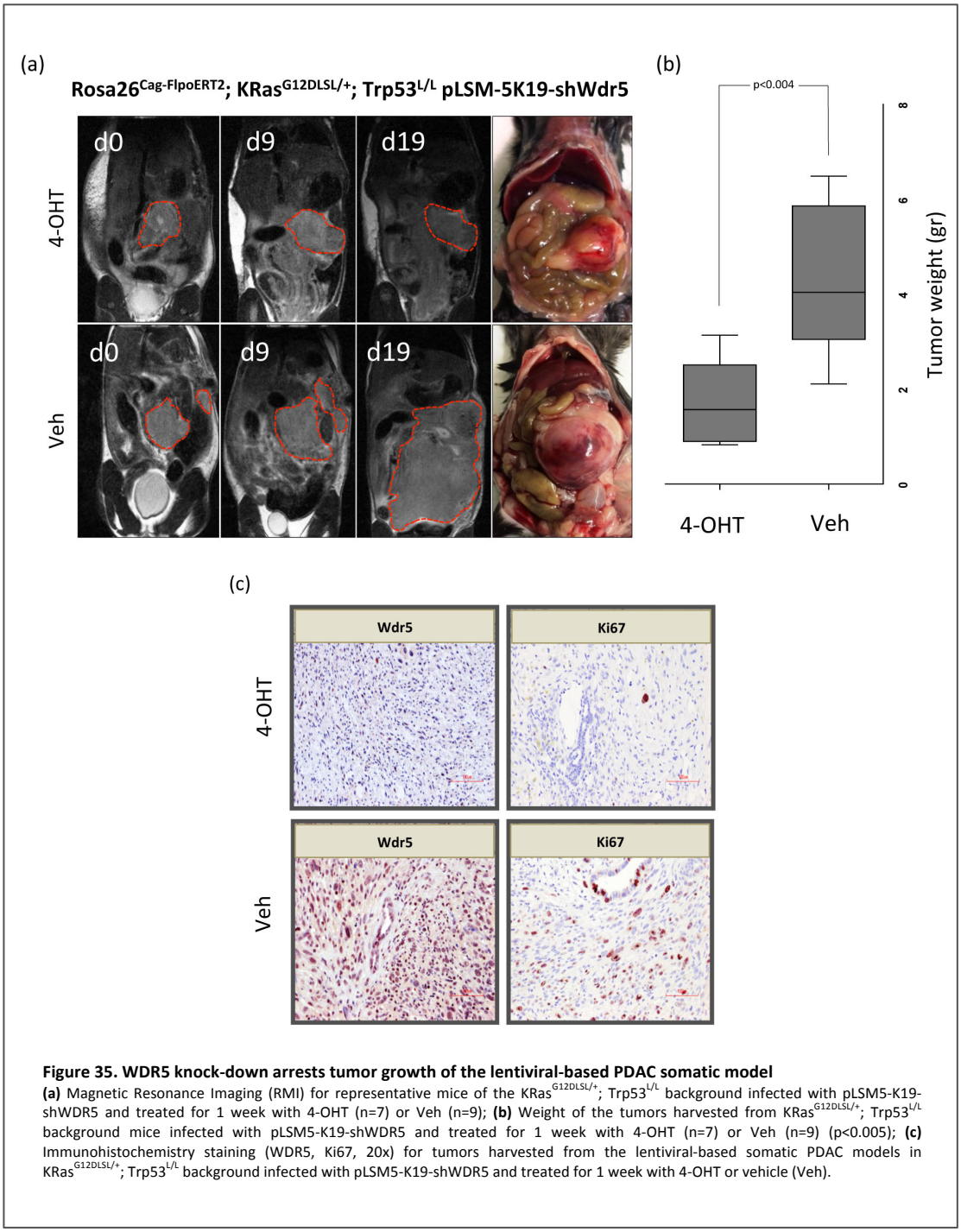


Figure 35. WDR5 knock-down arrests tumor growth of the lentiviral-based PDAC somatic model
(a) Magnetic Resonance Imaging (RMI) for representative mice of the KRas$^{G12DLSL/+}$; Trp53$^{L/L}$ background infected with pLSM5-K19-shWDR5 and treated for 1 week with 4-OHT (n=7) or Veh (n=9); **(b)** Weight of the tumors harvested from KRas$^{G12DLSL/+}$; Trp53$^{L/L}$ background mice infected with pLSM5-K19-shWDR5 and treated for 1 week with 4-OHT (n=7) or Veh (n=9) (p<0.005); **(c)** Immunohistochemistry staining (WDR5, Ki67, 20x) for tumors harvested from the lentiviral-based somatic PDAC models in KRas$^{G12DLSL/+}$; Trp53$^{L/L}$ background infected with pLSM5-K19-shWDR5 and treated for 1 week with 4-OHT or vehicle (Veh).

Transplanting epithelial progenitors from Rosa26$^{mTmG/+}$ mice, upon infection with the pLSM5-K19 lentiviral vector, we observed no tumor formation and differentiation in pancreatic acini, as demonstrated by double positivity for GFP and Amylase of pancreatic sections (**Fig.34c,d**). Mice were monitored for tumor formation by MRI imaging weekly. The FlpoERT2 was activated by repeated tamoxifen treatments to remove the stopper cassette and activate the shRNA in advanced pancreatic tumors. In line with our previous findings, the acute inactivation of Wdr5 in vivo resulted in a delayed growth and in smaller lesions (**Fig.35a,b**) characterized by a decrease in the numbers of Ki67 positive cells (**Fig.35c**). These data confirm a role for WDR5 as a critical tumor dependency in human PDAC.

5. WDR5 complex protects PDAC cells from DNA damage and aneuploidy stabilizing the DNA replication forks

The power of our approach was further demonstrated by the fact that we identified multiple members of the COMPASS complex (ASH2L, MEN1, MLL1, MLL2) as "hits" in both the human and mouse PDAC screens (**Fig.36a**).

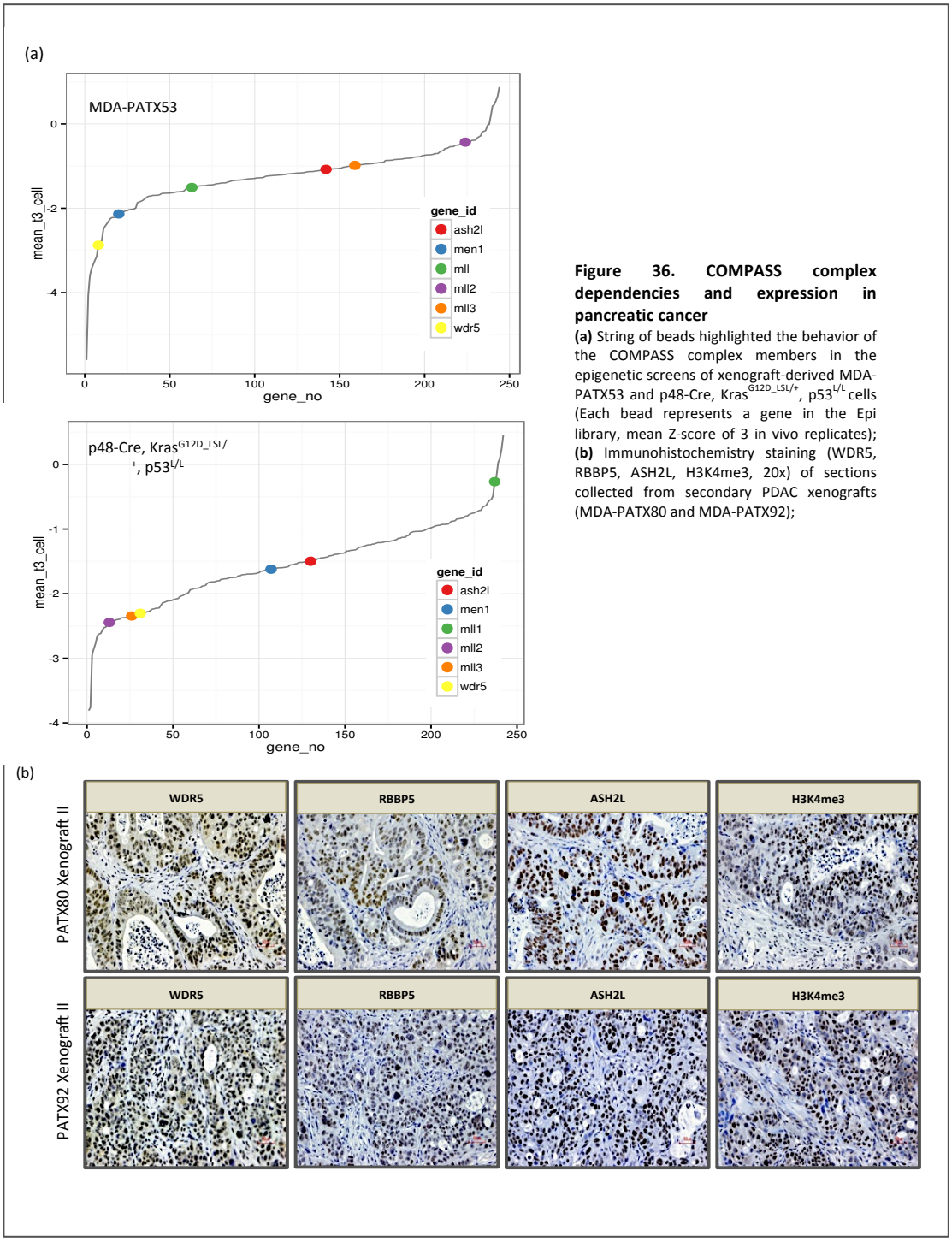


Figure 36. COMPASS complex dependencies and expression in pancreatic cancer
(a) String of beads highlighted the behavior of the COMPASS complex members in the epigenetic screens of xenograft-derived MDA-PATX53 and p48-Cre, $Kras^{G12D\_LSL/+}$, $p53^{L/L}$ cells (Each bead represents a gene in the Epi library, mean Z-score of 3 in vivo replicates); **(b)** Immunohistochemistry staining (WDR5, RBBP5, ASH2L, H3K4me3, 20x) of sections collected from secondary PDAC xenografts (MDA-PATX80 and MDA-PATX92);

COMPASS and COMPASS-like complexes are characterized by their unique subunit composition, and individual subunits appear to dictate the biological functions of each complex[44,47,48]. For example, even though both MLL1 and MLL2 are recruited to the Hox loci through MEN1-specific interactions, they also have non-redundant functions as exemplified by the MLL1 and MLL2 knock-out mouse models.[212] Similarly, MLL3 and MLL4 may share redundant functions in regulating the Hox genes, but they also interact with subunits important for targeting nuclear receptors, such as PTIP[48]. In line with these data, MLL1 and MLL2 seem to drive context-specific essential functions also in PDAC, as shown by the non-overlapping results of our screens (**Fig.36a**).

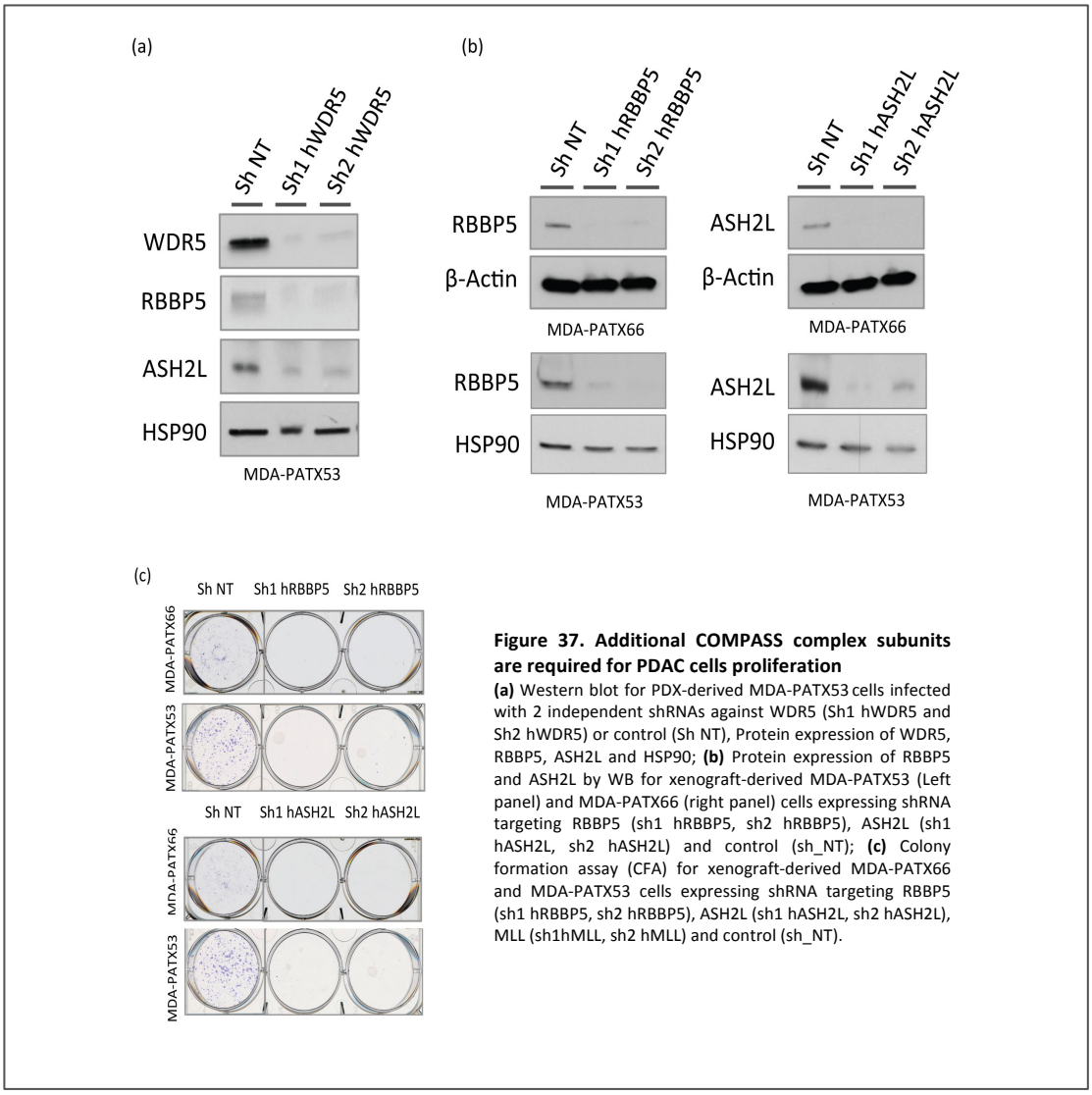
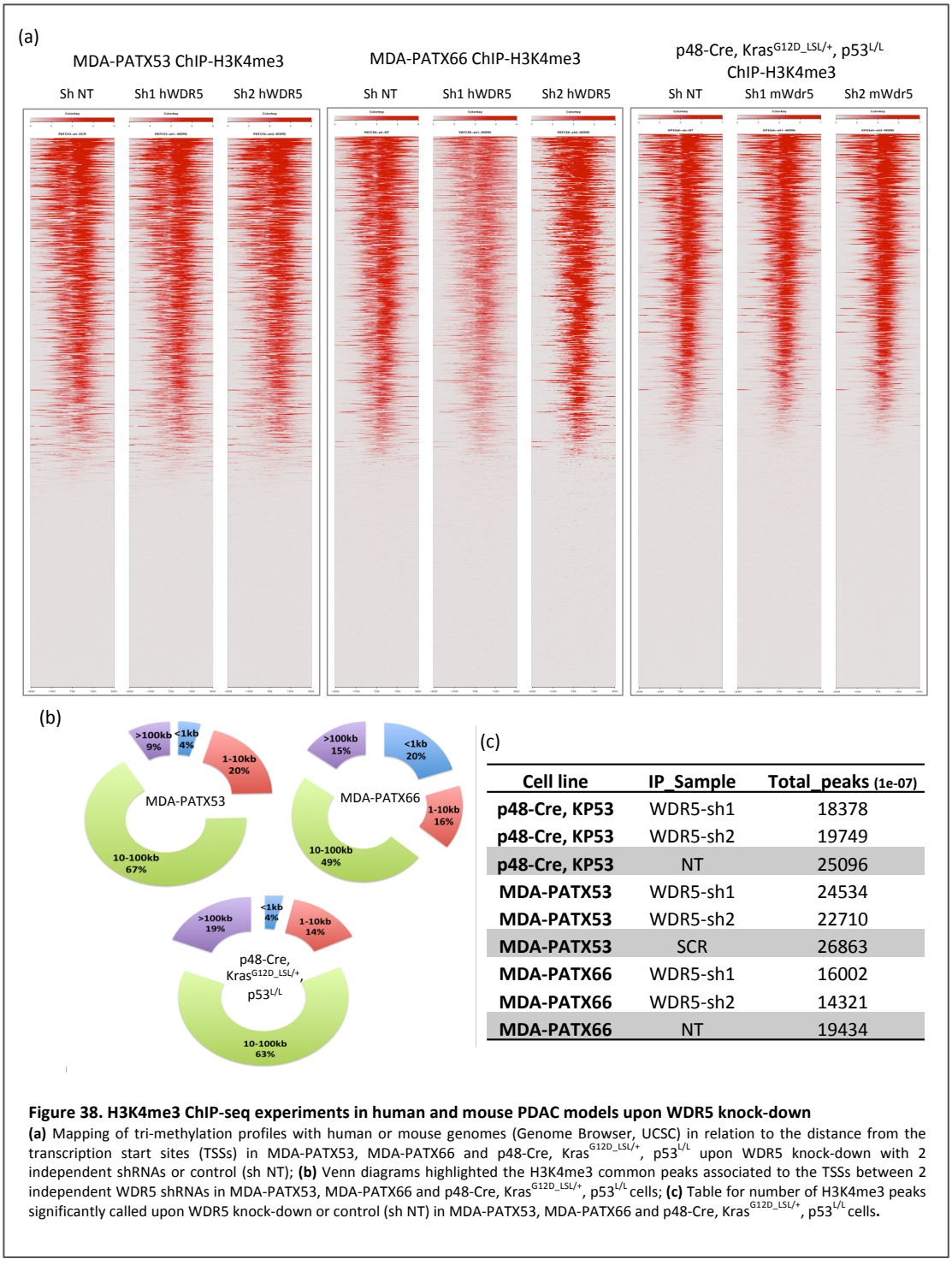


Figure 37. Additional COMPASS complex subunits are required for PDAC cells proliferation
(a) Western blot for PDX-derived MDA-PATX53 cells infected with 2 independent shRNAs against WDR5 (Sh1 hWDR5 and Sh2 hWDR5) or control (Sh NT), Protein expression of WDR5, RBBP5, ASH2L and HSP90; **(b)** Protein expression of RBBP5 and ASH2L by WB for xenograft-derived MDA-PATX53 (Left panel) and MDA-PATX66 (right panel) cells expressing shRNA targeting RBBP5 (sh1 hRBBP5, sh2 hRBBP5), ASH2L (sh1 hASH2L, sh2 hASH2L) and control (sh_NT); **(c)** Colony formation assay (CFA) for xenograft-derived MDA-PATX66 and MDA-PATX53 cells expressing shRNA targeting RBBP5 (sh1 hRBBP5, sh2 hRBBP5), ASH2L (sh1 hASH2L, sh2 hASH2L), MLL (sh1hMLL, sh2 hMLL) and control (sh_NT).

The WDR5-RBBP5-ASH2L (WAR) core showed high protein expression level in human PDAC xenografts (**Fig.36b**). The functional non-redundant role of WAR complex in PDAC was proven by the significant impairment of colony formation ability we observed when ASH2L and RBBP5 were down regulated in PDX-derived samples using 2 independent shRNAs (**Fig.37a-c**).



Cell line	IP_Sample	Total_peaks (1e-07)
p48-Cre, KP53	WDR5-sh1	18378
p48-Cre, KP53	WDR5-sh2	19749
p48-Cre, KP53	NT	25096
MDA-PATX53	WDR5-sh1	24534
MDA-PATX53	WDR5-sh2	22710
MDA-PATX53	SCR	26863
MDA-PATX66	WDR5-sh1	16002
MDA-PATX66	WDR5-sh2	14321
MDA-PATX66	NT	19434

Figure 38. H3K4me3 ChIP-seq experiments in human and mouse PDAC models upon WDR5 knock-down
(a) Mapping of tri-methylation profiles with human or mouse genomes (Genome Browser, UCSC) in relation to the distance from the transcription start sites (TSSs) in MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ upon WDR5 knock-down with 2 independent shRNAs or control (sh NT); **(b)** Venn diagrams highlighted the H3K4me3 common peaks associated to the TSSs between 2 independent WDR5 shRNAs in MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ cells; **(c)** Table for number of H3K4me3 peaks significantly called upon WDR5 knock-down or control (sh NT) in MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ cells.

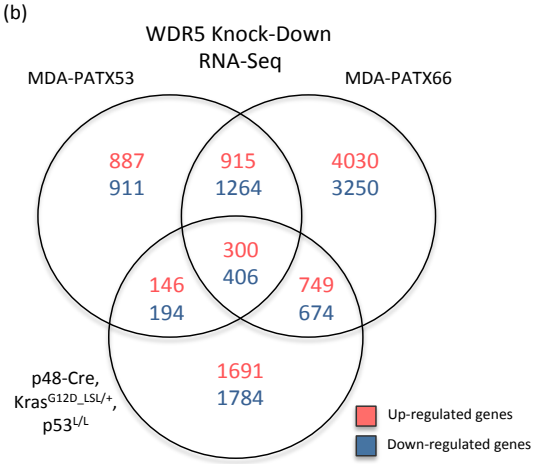
The methyltransferase subunit of the COMPASS complex is catalyzing the methylation of lysine 4 on histone H3, a validated marker of open-chromatin conformation and active transcription[44,213]. Knock-down of WDR5 in both human and mouse PDAC cell lines induced a modest overall reduction in global methylation levels as confirmed by anti-H3K4me3 ChIP-Seq experiments (**Fig.38a**). Mapping of tri-methylation profiles with functional elements of the genome confirmed that only a small fraction (4-20%) of the significantly altered methylation regions were in proximity (‹1kb) of transcriptional start sites (TSSs) and these peaks were statistically significant (**Fig.38b,c**).

In addition, the WAR tripartite complex has been reported to execute additional H3K4 methylation-independent functions, as the enhancement of the transcriptional activation of nuclear hormone receptor-responsive genes[214]. We next performed RNA-Seq to inform on transcriptional changes conferred upon WDR5 knock-down with 2 independent shRNAs. A consistent number of genes displayed modification in mRNA expression level with both the shRNAs for WDR5 (1% FDR, 1.5 FC) and crossing results from human and mouse PDAC models we robustly identified a common set of genes affected by WDR5 silencing in pancreatic cancer (**Fig.39a,b**). Remarkably, genes for which a reduction in the level of H3K4me3 was detected in the promoter regions showed a significant association with down-regulation at the transcription level upon WDR5 knock-down, in accordance with the role of the COMPASS complex as activator of transcription (data not shown). However, gene set enrichment analysis (GSEA) identified up-regulated genes involved in the control of DNA replication and progression through the different phases of the cell cycle in all the PDAC models (**Fig.39c,d**). The increase in transcriptional levels of genes essential for supporting replicative mechanisms and cell proliferation could be explained as compensatory events carried out by the PDAC cells to counteract the WDR5 loss of function.

(a)

Cell line	Comparison	1% FDR		5% FDR		10% FDR		1%FDR, 1.5 FC	1% FDR, 1.5FC
		Up	Down	Up	Down	Up	Down	Up	Down
MDA-PATX53	sh1-control	1125	1295	1926	2182	2532	2928	988	1207
	sh2-control	1620	1994	2646	2965	3383	3588	1490	1807
MDA-PATX66	sh1-control	3632	4055	4889	5351	5603	6144	3632	4055
	sh2-control	2633	2826	3703	4001	4475	4833	2633	2826
P48-Cre, KP53	sh1-control	3124	3324	4097	4246	4690	4784	2809	3152
	sh2-control	1247	1237	1996	2013	2535	2547	1153	1162

(b)



(d)

PATX53	NES	p-value	FDR
REACTOME_MITOTIC_G1_G1_S_PHASES	2.15	0	0
REACTOME_G1_S_TRANSITION	2.21	0	0
REACTOME_LAGGING_STRAND_SYNTHESIS	1.76	0.006	0.025

PATX66	NES	p-value	FDR
REACTOME_MITOTIC_G1_G1_S_PHASES	1.98	0	0.001
REACTOME_G1_S_TRANSITION	2.01	0	0.001
REACTOME_LAGGING_STRAND_SYNTHESIS	1.41	0.091	0.183

P48-Cre, KP53	NES	p-value	FDR
REACTOME_MITOTIC_G1_G1_S_PHASES	1.72	0.001	0.093
REACTOME_G1_S_TRANSITION	1.66	0.001	0.124
REACTOME_LAGGING_STRAND_SYNTHESIS	1.65	0.009	0.104

(c)

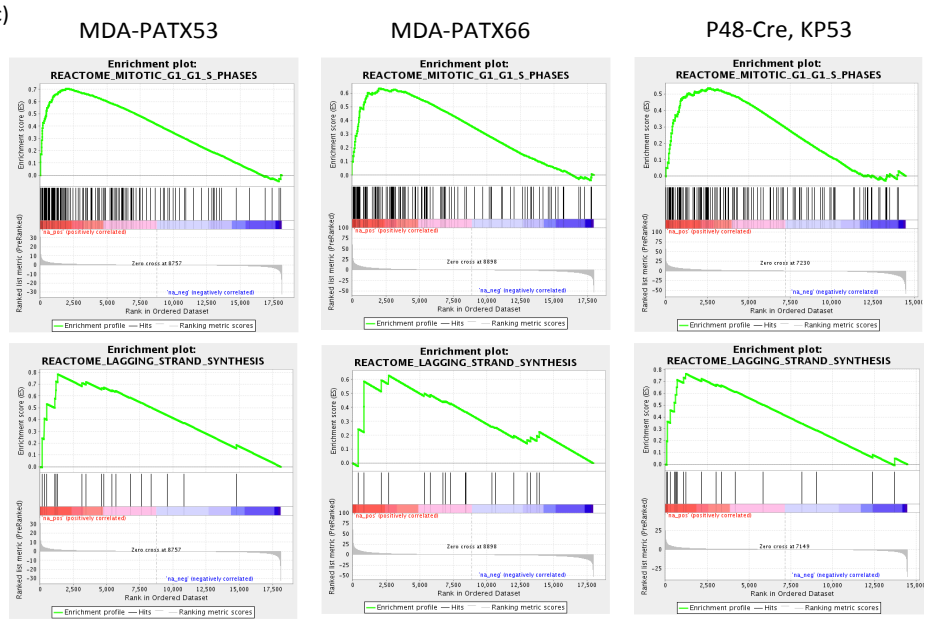


Figure 39. RNA-Seq experiments in human and mouse PDAC models upon WDR5 knock-down

(a) Number of genes differentially expressed (Up- or Down-regulated) in MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ (KP53) cells comparing 2 independent WDR5 shRNAs (sh1 and sh2) with control (Sh NT) (FDR = False Discovery Rate, FC = Fold-Change); **(b)** Venn diagram of the genes differentially expressed in human (MDA-PATX53, MDA-PATX66) and mouse (p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$) PDAC models upon WDR5 knockdown (RNA-Seq replicates, FC>1.5, FDR 1%); **(c)** Gene Set Enrichment Analysis (GSEA) for pathways significantly enriched and in common between MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ (KP53) cells comparing 2 independent WDR5 shRNAs (sh1 and sh2) with control (Sh NT), **(d)** Table for NES, p-value and false discovery rate (FDR) associated with significantly enriched pathways in MDA-PATX53, MDA-PATX66 and p48-Cre, Kras$^{G12D_LSL/+}$, p53$^{L/L}$ (KP53) cells comparing 2 independent WDR5 shRNAs (sh1 and sh2) with control (Sh NT).

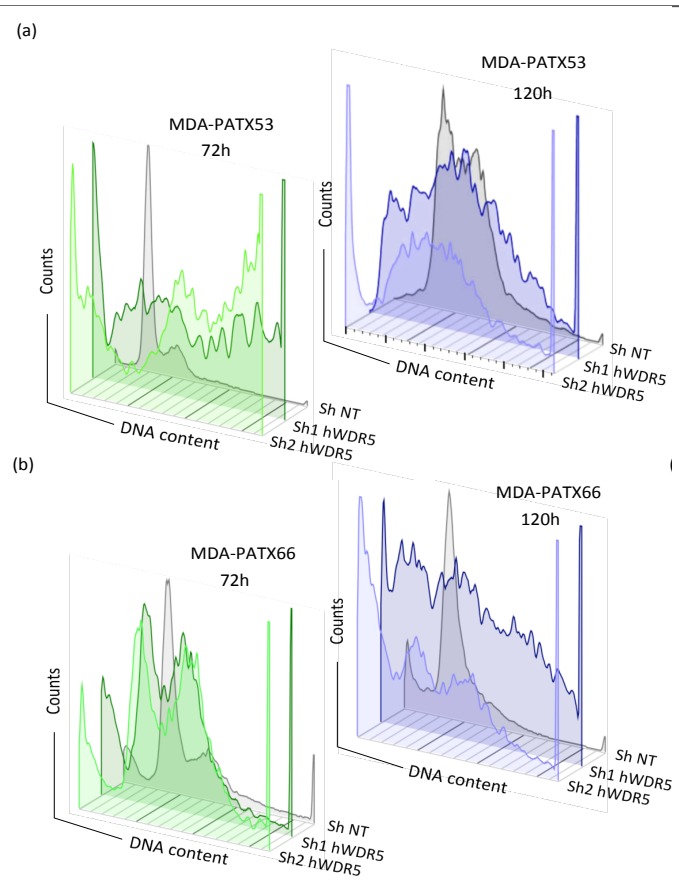
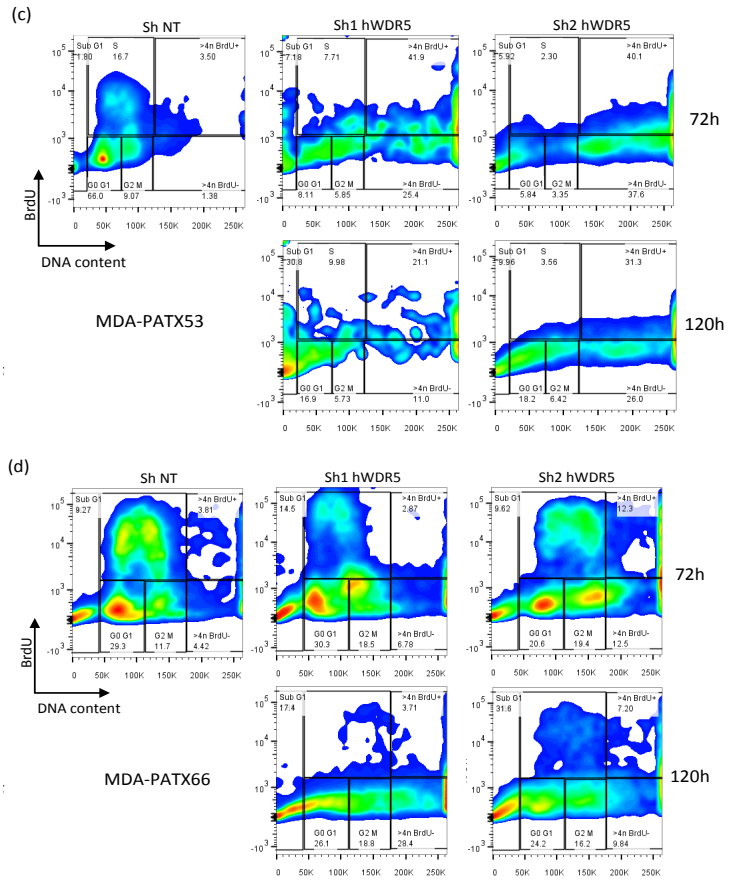


Figure 40. WDR5 protects PDAC cells from replicative defects and aneuploidy

(a) DNA content analysis (DAPI) by flow cytometry for PDX-derived MDA-PATX53 cells with 2 independent shRNA targeting WDR5 or shNT (72 and 120 hrs after selection); **(b)** DNA content analysis (DAPI) by flow cytometry for PDX-derived MDA-PATX66 cells with 2 independent shRNA targeting WDR5 or shNT (72 and 120 hrs after selection); **(c)** Flow cytometry analysis of BrdU uptake (2h pulse) and DNA content in MDA-PATX53 cells infected with 2 independent shRNA targeting WDR5 (Sh1 hWDR5 and Sh2 hWDR5) or sh NT (72 or 120 hrs after infection), Gates highlight the different phases of the cell-cycle (Sub-G1, G0/G1, S, G2/M) and cells with extra amount of DNA content (>4n Brd+, >4n Brd-); **(d)** Flow cytometry analysis of BrdU uptake (2h pulse) and DNA content in MDA-PATX66 cells infected with 2 independent shRNA targeting WDR5 (Sh1 hWDR5 and Sh2 hWDR5) or sh NT (72 or 120 hrs after infection), Gates highlight the different phases of the cell-cycle (Sub-G1, G0/G1, S, G2/M) and cells with extra amount of DNA content (>4n Brd+, >4n Brd-).



To further inform on a functional role of the WAR complex in regulating DNA replication we performed BrdU-labeling studies in PDAC cells. Consistent with the observed transcriptional changes, knock-down of WDR5 resulted in reduction of BrdU incorporation, with a paradoxical increase in overall DNA content suggesting failure to sustain DNA replication checkpoint or mitotic defects (**Fig.40a-d**). First of all, we designed a single guide RNA (sgRNA) guide specific for targeting WDR5 with the CRISPR/CAS9 lentiviral system to interrogate the origin of the aneuploidy as a consequence of WDR5 silencing (**Fig.41a**).

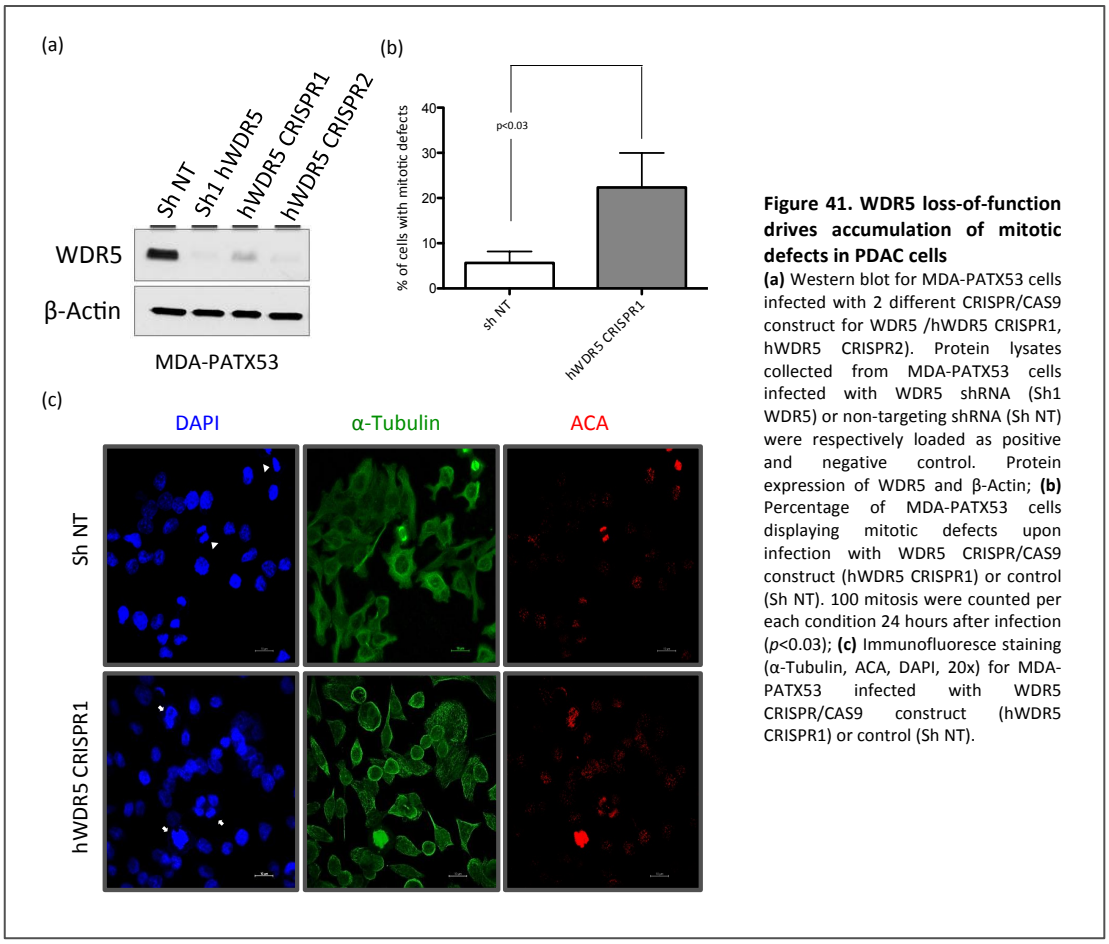


Figure 41. WDR5 loss-of-function drives accumulation of mitotic defects in PDAC cells
(a) Western blot for MDA-PATX53 cells infected with 2 different CRISPR/CAS9 construct for WDR5 /hWDR5 CRISPR1, hWDR5 CRISPR2). Protein lysates collected from MDA-PATX53 cells infected with WDR5 shRNA (Sh1 WDR5) or non-targeting shRNA (Sh NT) were respectively loaded as positive and negative control. Protein expression of WDR5 and β-Actin; **(b)** Percentage of MDA-PATX53 cells displaying mitotic defects upon infection with WDR5 CRISPR/CAS9 construct (hWDR5 CRISPR1) or control (Sh NT). 100 mitosis were counted per each condition 24 hours after infection (p<0.03); **(c)** Immunofluoresce staining (α-Tubulin, ACA, DAPI, 20x) for MDA-PATX53 infected with WDR5 CRISPR/CAS9 construct (hWDR5 CRISPR1) or control (Sh NT).

This approach gave us the possibility to investigate early events and the first mitosis upon WDR5 withdrawal. PDAC cells infected with the CRISPR/CAS9 construct for WDR5 displayed a significant increase in defective mitosis (multipolar mitotic spindles) and cytokinesis, as we were able to appreciate staining for anti-centromere antibodies (ACA), DAPI and α-Tubulin (**Fig.41b,c**).

High sensitivity of tumor cells to S-phase defects was associated with an increase in the number of active replication forks and concomitant frequent alterations in arresting G1 checkpoints[215]. In line with these data, PDAC cells with defects in G1-checkpoint showed accumulation of replicative damage when treated with a specific ATR inhibitor (VE-821) (**Fig.42a,b**). The accumulation of deregulated genetic materials observed in PDAC cells was associated with the induction of DNA damage (phospho-H2AX staining) upon WDR5 silencing that accompanied the deregulated DNA replication events (**Fig.36d**). The excessive DNA content and the accumulation of phospho-H2AX were not detected in the wild-type MEF upon WDR5 knock-down or ATR inhibition (**Fig.42c,d** and data not shown).

These findings indicated a critical role for WDR5 in sustaining the proper execution of DNA replication in PDAC cells The assembly of a proficient and active replication machinery is a multistep process strictly regulated during G1- and S-phases of the cell cycle through the involvement of CDKs and CDK inhibitors (CDKIs)[213]. The licensing of the replication origins happens in G1 with the assembly of the pre-replication complex (pre-RC), characterized by ORC1-6, MCM2-7, CDC6 and CDT1. Signals promoting the entering into the S-phase start the firing of the replication origins and the formation of the pre-initiation complex (pre-IC). Features of this critical step are the disassociation from the pre-replication complex of CDC6 and CDT1, upon phosphorylation, and the recruitment of CDC45 and GNIS at the replication forks[216,217].

In this direction, the subcellular protein fractionation of PDAC cells upon WDR5 knock-down demonstrated a significant reduction of CDC45 recruitment on the chromatin, in front of a CDT1 accumulation (**Fig.43a**). Displacement of CDC45 from the pre-initiation complex suggested that the origin of the DNA damage we observed in PDAC cells could be related to a prolonged stalling and further collapse of the replications forks.

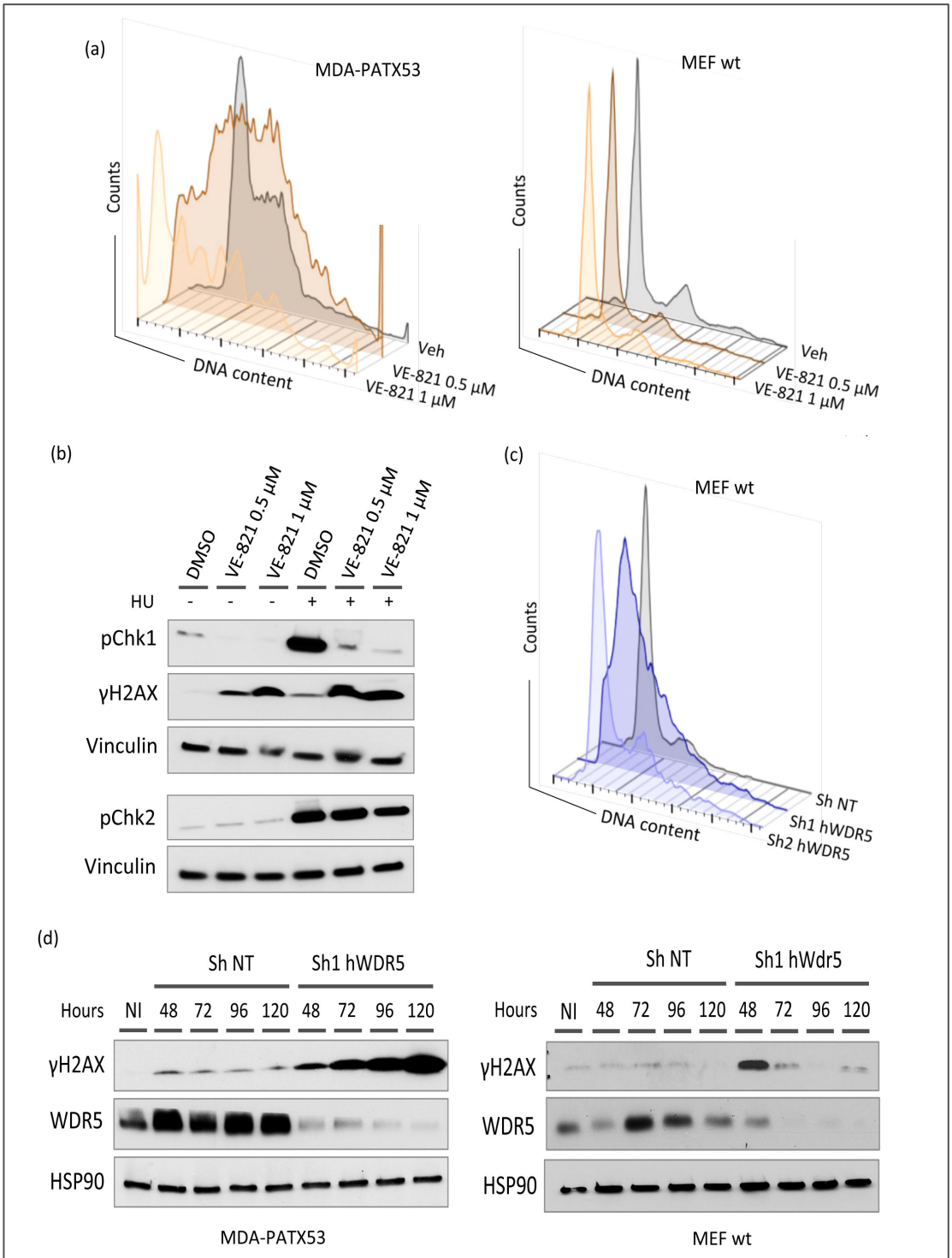


Figure 42. PDAC cells are sensitive to replicative defects and accumulate DNA damage upon WDR5 knock-down
(a) DNA content analysis (DAPI) by flow cytometry for PDX-derived MDA-PATX53 cells treated with ATR inhibitor VE-821 (0.5 µM and 1 µM) or vehicle (96 hrs, left panel), DNA content analysis (DAPI) by flow cytometry for wild-type mouse embryo fibroblast (MEF) cells treated with ATR inhibitor VE-821 (0.5 µM and 1 µM) or vehicle (96 hrs, right panel); **(b)** Western blot for MDA-PATX53 cells treated for 72 hours with 2 different concentrations (0.5 µM and 1 µM) of the ATR inhibitor VE-821 or vehicle, protein expression of phospho-Chk1 (Ser345), phospho-Chk2 (Thr68), phospho H2AX (γH2AX) and Vinculin; **(c)** DNA content analysis (DAPI) by flow cytometry for wild-type mouse embryo fibroblast (MEF) cells infected with 2 independent shRNA targeting WDR5 or shNT (120 hrs after selection); **(d)** Western blot for MDA-PATX53 cells (left panel) or wild-type MEF (right panel) expressing shRNA targeting WDR5 (sh1 hWdr1) and control (shNT) at different time-points after infection (48, 72, 96, 120h, NI=not infected). Protein expression of phospho-H2AX (γH2AX), WDR5 and HSP90;

The recent finding of a direct interaction between WDR5 and c-Myc and the reduction of chromatin-bound c-Myc upon WDR5 silencing open the possibility that WDR5 may be recruited to replication origins to sustain replication stress and tumorigenicity in PDAC cells[218]. These results demonstrated a critical role of WDR5 in stabilizing the replication forks in the presence of replicative stress (**Fig.43b**).

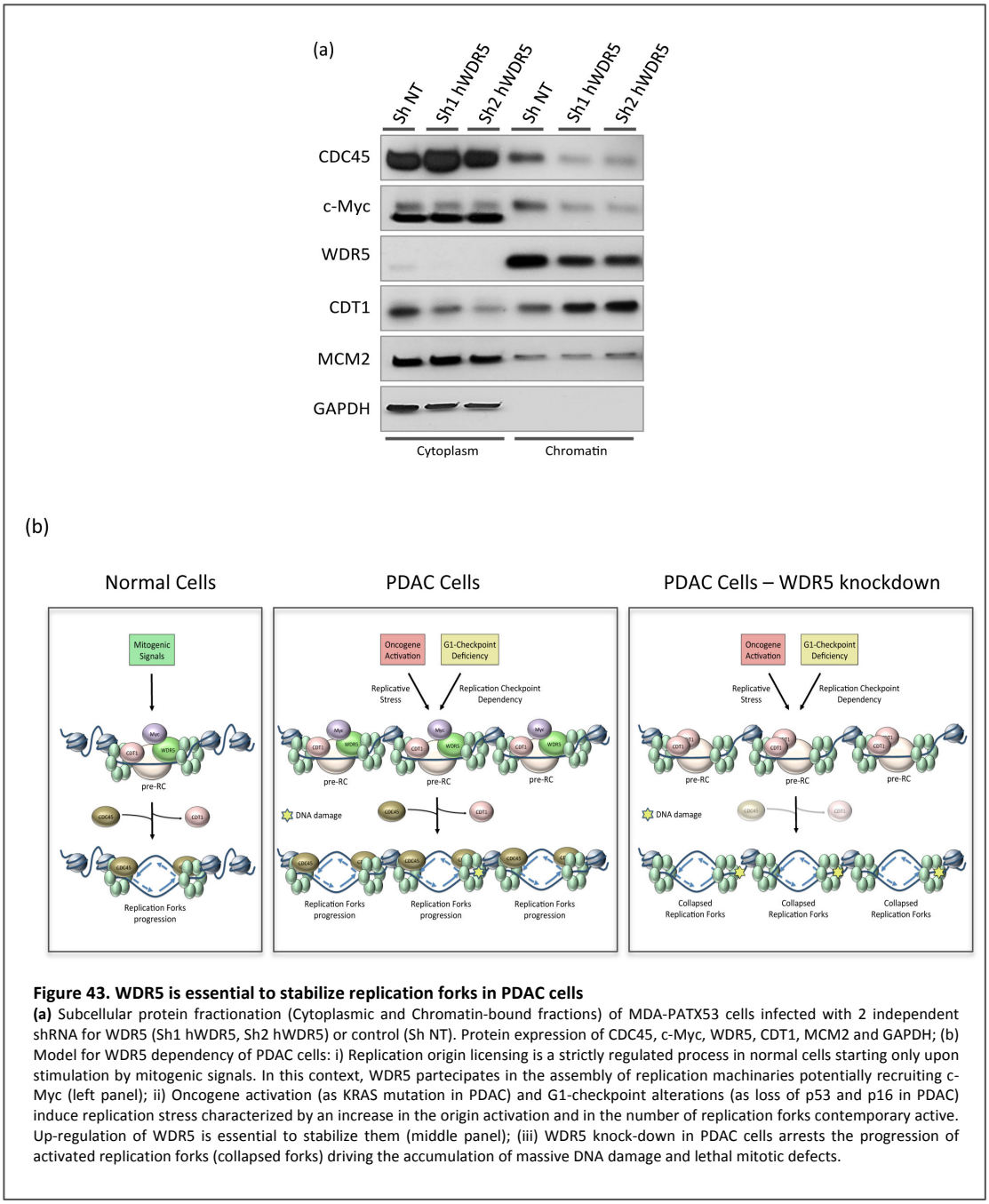


Figure 43. WDR5 is essential to stabilize replication forks in PDAC cells
(a) Subcellular protein fractionation (Cytoplasmic and Chromatin-bound fractions) of MDA-PATX53 cells infected with 2 independent shRNA for WDR5 (Sh1 hWDR5, Sh2 hWDR5) or control (Sh NT). Protein expression of CDC45, c-Myc, WDR5, CDT1, MCM2 and GAPDH; (b) Model for WDR5 dependency of PDAC cells: i) Replication origin licensing is a strictly regulated process in normal cells starting only upon stimulation by mitogenic signals. In this context, WDR5 partecipates in the assembly of replication machinaries potentially recruiting c-Myc (left panel); ii) Oncogene activation (as KRAS mutation in PDAC) and G1-checkpoint alterations (as loss of p53 and p16 in PDAC) induce replication stress characterized by an increase in the origin activation and in the number of replication forks contemporary active. Up-regulation of WDR5 is essential to stabilize them (middle panel); (iii) WDR5 knock-down in PDAC cells arrests the progression of activated replication forks (collapsed forks) driving the accumulation of massive DNA damage and lethal mitotic defects.

Summary of results

- Adaptation of the pancreatic cancer PDX models to *in vivo* functional studies

- New barcode-based TIC frequency tool for a sample-by-sample rapid estimation of the minimum required coverage of complex libraries

- *In vivo* loss-of-function screens in human PDX and mouse models to inform on new epigenetic context-dependent vulnerabilities in PDAC

- Integration of the top-scoring "hits" in a comprehensive triage pipeline for target prioritization

- Establishment of a new somatic model of PDAC able to recapitulate the progression of the disease

- WDR5 over-expression in PDAC is associated with cell-cycle functions essential to sustain tumor proliferation and maintenance

- WDR5 inhibition selectively promotes DNA damage accumulation in PDAC cells with p53 and p16 alterations (G1 checkpoint-deficiency), potentially destabilizing the DNA replication forks

Discussion and future directions

Large-scale genomics efforts have provided the opportunity to access a comprehensive catalog of genetic alterations in multiple cancers. They have also highlighted both the high frequency at which alterations are detected and the significant intra- and inter-tumor heterogeneity of the diseases. It has also become apparent that very few driver mutations are emerging and as a consequence of that, limited opportunities exist to target mutated oncogenic proteins. It is imperative therefore to develop alternative approaches to therapy that can leverage the selective vulnerabilities of tumor cells resulting from the engagement of abnormal pathway connectivity. These can be best exploited *in vivo*, in a context that is closer to the environment tumors strive in[143,149,150]. To identify new relevant actionable dependencies we have developed a system for rapid identification and validation of potential therapeutic targets in PDX tumor cells. By optimizing primary rapid tumor explant and expansion, determination of tumor initiating cell frequency through retroviral-mediated transduction, next generation sequencing and bioinformatics, we have been able to identify a first set of factors that are dramatically impacting pancreatic cancer in a most resistance disease context (activating mutation of KRAS and p53/p16 inactivations).

Biology-applied robotics and small-scale optimizations had started the so-called "high-throughput screening era". Despite its tremendous potential, the very high percentage of gene candidates failing the *in vitro-in vivo* step rapidly unmasked the intrinsic weak spot: non-physiological conditions. In the last few years, many laboratories endeavored to find a way to couple the high-throughput techniques with the *in vivo* models. So far, the majority of performed *in vivo* screens were focused on the identification of cooperating factors in cancer mouse models[125,126,127]. Opportunities for translation of these valuable tools to identify new actionable functionalities in human contexts have been hampered by the fact that tumor-initiating cells (TICs) are rare in many human cancers[168-172]. TIC frequency assessment is traditionally based on serial transplantation in limiting dilution, a time-

consuming and imprecise method[202,203]. Recently, Possemato and colleagues showed the first pooled-shRNA in vivo screen on a human breast cancer cell line (MCF10DCIS.com) and relied on the statistical power of the average between eleven replicates for hits identification[132]. In our study, we propose a loss-of-function in vivo screen on patient-derived samples from one of the most aggressive solid tumors, the pancreatic ductal adenocarcinoma, adjusted for the effective number of tumor initiating cells. A complexity study using thousands of molecular barcodes demonstrated to be a useful tool to increase the sensitivity of the method and to rapidly optimize conditions for shRNA library transduction reducing the number of required animals. The integration of this approach with barcoded on-target shRNA libraries in exactly the same experimental setting established a flexible format that can take into account tumor variability between patients. This point acquires even more relevance taking into account the adaptation of our barcode platform into a context of patient-derived xenografts (PDXs). So far, limitations of extensive use of PDX models for functional genomic studies were intrinsically related to the extreme variability observed in establishing stable *in vitro* cultures[134,135]. Our new protocol, optimized in each one of the involved steps (single cell isolation, cell seeding and culturing media), displayed 100% rate of success (7/7). The potential caveat of genetic biases introduced by a prolonged *in vitro* culture was limited in our approach restricting the number of passages before transplantation. *In vivo* expansion of the tumor cell population in primary xenograft and cell seeding at high confluency further guaranteed the possibility to obtain the required number of cells for performing drop-out screens in a short period of time. Interestingly, the extension of our in vivo screening platform to other medical indications with longer survival expectations compared to PDAC (ex. Melanoma, Triple-Negative Breast Cancer) will potentially allow for identification of selective vulnerabilities in time to impact treatment recommendations.

To increase the stringency of our studies, we chose to perform parallel screens in human and mouse PDAC cells, using shRNA sets specific for either context. This increased the chance to

find cancer vulnerabilities related to a specific context dependency and allowed the application of functional and structural criteria for hits prioritization. The robustness of our analytics pipeline was proven by the absence of false positive hits among the top-scoring genes, as confirmed by the experiments with independent shRNAs. Developing a high-throughput validation scheme to triage hits emerged from each screen we were able to enroll the most promising candidates in functional and clinico-pathological studies to determine the highest priority targets. The idea of proving the efficacy of our *in vivo* screening platform applying a set of shRNA specific for chromatin-remodeling enzymes was dictated by the very low mutation frequency of these genes in PDAC, in front of a consistent deregulation at the protein level. Focusing on different PDAC genetic backgrounds, with the common denominator of mutant KRAS, we increased the chance of identifying actionable epigenetic dependencies to overcome actual KRAS "undruggability".

Genetic lesions in chromatin regulators have been identified in a variety of cancers and new epigenetic cancer dependencies are emerging as intriguing actionable vulnerabilities[32,219]. Chromosomal translocations involving MLL gene are frequent events characterizing the Mixed Lineage Leukemia. In this disease, it has been shown that fusion events with a variety of different partners compromise the MLL methyltransferase activity[220]. However, multiple members of the MLL family could be deregulated via different oncogenic mechanisms in PDAC, as the genetic alterations in MLL2 (amplification) and MLL3 (mutation) suggested[34,221]. Therefore, a deeper understanding of genes and pathways regulated by each MLL subunit in the context of PDAC will be critical to elucidate the molecular dynamics of this disease better and to identify additional key points of vulnerability. Experiments with the MM-401 peptide, able to selectively disrupt the interaction between MLL1 and WDR5 without impacting the other MLL subunits, could be performed to dissect these specific PDAC context dependencies[222].

In this study we identified the core different subunits of the COMPASS complexes (WDR5, ASH2L, RBBP5) as broad relevant players in sustaining PDAC progression and maintenance,

while the dependency on the MLL subunits appears to be more context-dependent and potentially consequent to specific genetic alterations. These findings were further enforced by minimal overlap between transcriptionally regulated gene sets in 3 different PDAC models upon COMPASS complex depletion, as proved by the intersection between expression profiles and promoter methylation. However, our actual analysis was limited to the methylation level of the promoter regions and can not exclude a relevant contribution of enhancer regions in PDAC transcriptional regulation. So, we propose to investigate the contribution of these distal genomic regions in sustaining essential PDAC mechanisms crossing our H3K4 methylation profiles with the span of H3K27 acetylation, a histone mark associated to the active enhancers[223]. Moreover, Blobel and colleagues recently proposed an intriguing hypothesis for the COMPASS role in mitosis as a positional bookmark for transcriptional reactivation of immediate early genes (IEGs) after mitotic exit[224]. In this scenario, it could be valuable investigating COMPASS regulation of IEGs in PADC contexts where mutant KRAS imposes a higher requirement of their associated transcription factors (TFs), as previously demonstrated by Fos.

Different works clearly demonstrated a cell-cycle regulated expression of COMPASS complex members and presented as peaks at the G1–S and G2–M transitions. The interactions between MLL1/MLL2 complexes and E2F proteins are unique, or partially redundant, and may activate or repress E2F cell cycle target genes in addition to cyclins and cyclin-dependent kinase (CDK) inhibitors (CDKIs)[213]. Along the same line, a COMPASS complex methylation-independent role was recently proposed as a mechanism to regulate cell-cycle execution in U2OS osteosarcoma cell line[56]. Another possibility is a direct control of the cell-cycle mechanism, as demonstrated in yeast by a cross-talk between histone and non-histone methylations, controlled by the Set1-containing complex and able to influence the kinetochore subunit Dam1[225]. So, our results are helping to clarify the hijacked functions of the COMPASS complex in solid tumors, as demonstrated by the fact that PDAC cells under the constant stimulus of a potent oncogene are not able to repair DNA damages upon WDR5

silencing. The analysis of the early events is suggesting potential defects starting in G1-S phases of the cell-cycle that cannot be counteracted during chromosome segregation or cell division and are relentlessly driving cells to aneuploidy and mitotic catastrophe. These observations are further intriguing taking into account the prominent role of ATR and MLL in S-phase checkpoint activation and premature chromosome condensation (PCC) escaping in the presence of damaged DNA[55,215]. In this direction, the COMPASS complex could represent a new weakness point for PDAC tumors with G1-checkpoint alterations and compromised cell cycle-arrest checkpoints. Tumor cells possess an increased number of replication forks and endure elevated levels of replication stress compared to normal cells, opening the intriguing possibility that WDR5 over-expression in tumors may be required to stabilize the replication machinery. Whether this is a direct effect of WDR5 (or a WDR5 complex) binding or an indirect effect (possibly mediated through transcriptional control by a COMPASS-like complex), cannot be determined from our current studies. The recent finding of a direct interaction between WDR5 and c-Myc could imply a further possibility, that WDR5 may be recruited to replication origins to regulate c-Myc association with the replication machinary[218]. Dominguez-Sola and colleagues have recently demonstrated a direct role of c-Myc in the initiation of DNA replication that is independent from the transcriptional regulation[226]. More than that, in this study, WDR5, MCM2, MCM4 and c-Myc coimmunoprecipitated as part of the same complex enforcing the hypothesis of WDR5-mediated c-Myc recruitment at the replicative forks. Direct confirmation of this physical interaction at the replication origins of PDAC contexts will open the possibility to think new inhibitors selectively oriented towards the transcription-independent function of c-Myc and not at the interaction with Max. So, our findings are proposing a new layer of complexities in trapping the COMPASS complexes during tumor development and unmasking unexplored directions for new therapeutic opportunities.

Taken together, our data highlight this *in vivo* screening platform as a powerful new tool to identify genetic vulnerabilities using patient-derived tumor samples. This system can also be

enabled in syngeneic mouse models where one can probe the effects of target inhibition in the context of an intact immune response, and in the presence of immune checkpoint activators. As initial proof of concept of the validity of our approach, we identify WDR5 as a new regulator of PDAC development and maintenance, and illuminate therapeutic vulnerabilities that can be rapidly evaluated in the clinic through treatment with ATR inhibitors.

Materials and Methods

Antibodies, plasmids and chemical reagents. Primary antibodies used for flow cytometry, immunohistochemistry and immunoblotting were: Cytokeratin 19 (Proteintech, 14965-1-AP), Vimentin (Cell Signaling Technologies, D21H3), HLA A (Abcam, ab52922), phospho-ERK (Cell Signaling Technologies, D10), WDR5 (Cell Signaling Technologies, D9E1I), SMC2 (Cell Signaling Technologies, D91E3), PHF5A (Santa Cruz Biotechnology, V-12), Vinculin (Cell Signaling Technologies, E1E9V), β-Actin (sc-1615, Santa Cruz), HSP90 (Cell Signaling Technologies, C45G5), ASH2L (Cell Signaling Technologies, D93F6), RBBP5 (Cell Signaling Technologies, D3I6P), phospho-H2AX (Cell Signaling Technologies, 20E3), H432), p53 (Santa Cruz Biotechnology, FL-393), p16 (Proteintech, 10883-1-AP), SMAD4 (Abcam, ab40759), Ki67 (Thermo Scientific), Sox9 (Millipore), Pdx1 (Santa Cruz Biotechnology, A-17), H3K4me3 (Abcam, ab8580), H3 (Abcam, ab12079), GFP (Abcam, ab290), APC-conjugated BrdU (BD Biosciences, 552598), CDC45 (Cell Signaling Technologies, D7G6), c-Myc (Santa Cruz Biotechnology, 9E10), CDT1 (Cell Signaling Technologies, D10F11), MCM2 (Cell Signaling Technologies, D7G11), GAPDH (Cell Signaling Technologies, 14C10), APC-conjugated HLA-ABC (BD Biosciences, 555555), PE-conjugated H-2Kd (BD Biosciences, 553566). *Plasmids*: pLKO WDR5 shRNAs (Sigma, Human sh1 TRCN0000118047, sh2 TRCN0000118049; Mouse sh1 TRCN0000034415; sh2 TRCN0000034416), pLKO RBBP5 shRNAs (Sigma, Human sh1 TRCN0000166086, sh2 TRCN0000165777), pLKO ASH2L shRNAs (Sigma, Human sh1 TRCN0000019274, sh2 TRCN0000019275), pLKO PHF5A shRNAs (Sigma, Human sh1 TRCN0000286156, sh2 TRCN0000286158), pLKO SMC2 shRNAs (Sigma, Human sh1 TRCN0000291366, sh2 TRCN0000291367), pLKO Non-Targeting (Sigma, SHC007), pLX304 was a gift from D.Root (Addgene plasmid ♯25890), Tet-inducible pRSIT16 WDR5 shRNAs (Cellecta,, Human sh1 shhWDR5 251, sh2 shhWDR5 1000), Tet-inducible pRSIT16 Non-Targeting (Cellecta, shNT), CRISPR/Cas9 WDR5 constructs (Cellecta, pRSGC1). Full-length, sequence-verified cDNA for WDR5 (Clone♯IOH4895) was obtained from the Ultimate ORF

collection (Invitrogen) and transferred by Gateway cloning into a modified pHAGE-EF1a-IRES-EGFP lentiviral vector (generously provided by D.Kotton, BU School of Medicine). *Chemical reagents:* VE-821 (Selleck Chemicals), Hydroxyurea (Sigma Aldrich), Doxycycline Hydrochloride (RPI Corp.)

Somatic lentiviral vectors and other plasmids. pLSM5: a synthetic cassette (Geneart, Life Technologies) containing the U6 promoter and the Cre recombinase sequence under the human Keratin 19 promoter (-1114, +141) flanked by 2 TATA-Frt sites (XbaI-U6-TATA-Frt-EcoRI-hKrt19-NheI-Cre-TATA-Frt-HpaI) was cloned into the XbaI/HpaI site of the pSICO vector. A DNA fragment was liberated by XbaI/KpnI digestion and cloned into the XbaI KpnI sites of the pLB vector 32. The introduction of the TATA box into the Frt sites was designed according to Ventura et al. All the constructs were verified by restriction digestion and sequencing. The pSICO and pSICO-Flpo were made by Dr. Tyler Jacks[227]. The pLB vector was created by Dr. Stephan Kissler. All plasmids were obtained through Addgene.

Mouse strains. $Kras^{LSL\_G12D/+}$ mice were generated by Dr. Tyler Jacks and obtained through the Jackson Laboratory.[228] The $p48^{Cre/+}$, the $Ink4a^{LoxP/LoxP}$ and the $Trp53^{LoxP/LoxP}$ strains were donated by Dr. Ronald DePinho. The $R26^{Cag\text{-}FlpoERT2}$ was generated by Dr. Alexandra Joyner and obtained from the Jackson Laboratory[229]. The Cag-FlpER allele was kindly donated by Dr. Susan Dymecki[230]. $R26^{mTmG}$ strain was generated by Dr. Liqun Luo and obtained through the Jackson Laboratory[13]. The strains were kept in a mix C57BL/6 and 129Sv/Jae. All animal studies and procedures were approved by the UTMDACC Institutional Animal Care and Use Committee.

Tumor cell isolation and culture from human PDXs. Tumors from human primary xenografts (Xenograft I) were harvested in HBSS (Gibco). Isolation of pancreatic adenocarcinoma tumor xenograft (PATX) cells was performed by a combination of enzymatic (Tumor Dissociation Kit, human, Miltenyi Biotec) and mechanical (mincing the tumor tissue in very small pieces with sterile scissors) dissociation protocols. Erythrocytes were removed through RBC Lysis Buffer 1X (eBioscience). The single cell populations were seeded at high-confluency on collagen IV-coated plates (Corning) in DMEM/F12 (Gibco) supplemented with 10% FBS (Gibco), 1% BSA (Fisher Scientific), 0.5 µM Hydrocortisone (Sigma Aldrich), 10 mM HEPES (Invitrogen), 100 ng/mL Cholera Toxin (Sigma Aldrich), 5 mL/L Insulin-Transferrin-Selenium (BD), 100IU/mL Penicillin (Gibco), 100 µg/mL Streptomycin (Gibco). In order to get rid of the murine fibroblasts in the culture we performed brief trypsinization cycles (0.25% Trypsin-EDTA, Gibco) before each round of splitting. The enrichment for human components was confirmed by flow cytometry comparing the percentage of cells expressing human (HLA-ABC) or mouse (H-2Kd) histocompatibility complex antigens. The isolated human cells were maintained in culture for a maximum of three passages before being transplanted in a secondary host (Xenograft II). Primary xenograft isolated cells were also kept in culture as spheres in semi solid media. Single cell suspensions were plated in DMEM (Gibco) supplemented with 2 mM Glutamine (Invitrogen), 10% FBS (Gibco), 40 ng/mL hEGF (PeproTech), 20 ng/mL hFGF (PeproTech), 5 µg/mL h-Insulin (Roche), 0.5 µM Hydrocortisone (Sigma Aldrich), 100 µM β-Mercaptoethanol (Sigma Aldrich), 4 µg/mL Heparin (Sigma Aldrich), 100IU/mL Penicillin (Gibco), 100 µg/mL Streptomycin (Gibco). Methocult (StemCell Technologies) was added to SCM (0.8% final) to keep cells growing as clonal spheres versus aggregates. Fully formed spheres were collected and digested with 0.25% trypsin-EDTA (Gibco) to single cells and re-plated[25].

Tumor cell isolation and culture from GEM models. For isolation of cells from primary pancreatic tumors see Viale et al.[211]. Single cells derived from GEM tumors were kept in culture as adherent cells or spheres in semi-solid media. Briefly, tumors were digested at 37°C for 1h with a Collagenase IV-Dispase mix (4 mg/mL, Invitrogen) soon after explant. Single cell suspensions were plated in DMEM (Gibco) supplemented with 2 mM Glutamine (Invitrogen), 10% FBS (Gibco), 40 ng/mL hEGF (PeproTech), 20 ng/mL hFGF (PeproTech), 5 µg/mL h-Insulin (Roche), 0.5 µM Hydrocortisone (Sigma Aldrich), 100 µM β-Mercaptoethanol (Sigma Aldrich), 4 µg/mL Heparin (Sigma Aldrich), 100IU/mL Penicillin (Gibco), 100 µg/mL Streptomycin (Gibco). Methocult (StemCell Technologies) was added to SCM (0.8% final) to keep cells growing as clonal spheres versus aggregates. Fully formed spheres were collected and digested with 0.25% trypsin-EDTA (Gibco) to single cells and re-plated. For 2-D tumor culture cells were maintained in DMEM supplemented with 10% FBS (Gibco), 100IU/mL Penicillin (Gibco), 100 µg/mL Streptomycin (Gibco).

Tumor transplantation, transplantation in limiting dilution and *in vivo* studies. Tumor cells were isolated from PDX tumors or GEMM models and stabilized in culture (see Tumor cell isolation and culture). Usually, 10^6 or 10^5 tumor cells were respectively used for routine transplantation from PDX or GEMM models. For the experiments with the Tet-inducible shRNAs when the tumors reached a volume of 100-130 mm^3, mice were put under a doxycycline-containing (1 mg/mL) drinking water regimen and tumor size was measured every 5 days. Tumor volume was calculated using the formula: $V=l^2*L/2$ (l length; L width). For transplantation in limiting dilution were used 10^4, 10^3, 10^2 or 10 cells. Tumor cells were suspended in PBS (Gibco) and Matrigel (BD Biosciences) (1:1 dilution) and injected subcutaneously into the flank of 4- to 6-week-old female immunodeficient mice (NSG, The Jackson Laboratory). Tumor-initiating cell (TIC) frequencies and 95% confidence intervals were determined by the Extreme Limiting Dilution Analysis (ELDA) software[202]. For orthotopic

end point survival studies, 6-9 weeks old ncr/Nude female mice were injected with $2x10^5$ PDX cells resuspended in a 1:1 solution of PBS (Gibco) and Matrigel (BD Biosciences). All manipulations were performed under IACUC-approved protocols.

Libraries design and construction. A custom library constituted by 2410 shRNAs focused on chromatin remodeling enzymes was constructed by using chip-based oligonucleotide synthesis and cloned into the pRSI16 lentiviral vector (Cellecta) as a pool. The shRNA library targeted 236 genes with a coverage of 10 shRNAs/gene. The shRNA includes 2 G/U mismatches in the passenger strand, a 7 nucleotides loop and a 21 nucleotides targeting sequence. Targeting sequences were designed using a proprietary algorithm (Cellecta). The oligo corresponding to each shRNA was synthesized with a unique molecular barcode (18 nucleotides) for measuring representation by NGS. The 12.5k barcoded library applied for the TIC covering studies was constructed using the same technology and cloned as a pool into the pRSI17 lentiviral plasmid (Cellecta).

***In vivo* TIC coverage studies.** The volume of virus required to give a percentage of infection around 30% or below was determined sample by samples using a 3 points dose response in the presence of 8 µg/mL polybrene (Millipore)**:** 0.15, 0.3 and 0.6 transducing unit (TU)/cell for the human PDX cells; 0.3, 0.6 and 1.2 TU/cell for the GEMM-derived cells (See Suppl.Fig.2). Infectivity was determined as the % of RFP positive cells 2 days after infection as measured by FACS analysis. In vivo TIC covering studies were performed al least in replicate. For large scale infection of human PDX cells, 60 million cells were plated in T-175 flasks (Corning) with fresh media containing 8 µg/mL polybrene and sufficient virus to guarantee a 25% infection rate based on precedent calculations. For infection of GEMM-derived cells, 20 million cells were plated in T-75 flasks (Corning) with fresh media containing 8 µg/mL polybrene and sufficient virus to guarantee a 15% infection rate based on precedent calculations. 24 hours

after infection, the culture media was replaced with fresh media containing puromycin (Gibco). For each cell line the optimal puromycin dose to achieve more than 95% cell killing in 72 hours was determined by measuring cell viability with a Cell Titer Glo assay (Promega) for a 6 points dose response ranging from 0 to 8 μg of puromycin. 72 hours following puromycin addiction, cells were trypsinized, pooled together and counted. A representative portion of the total cells (normally 1/3 or 1/4) was collected as Reference Cells and immediately frozen as pellet at -80°C. The cells for the in vivo studies were separated into independent tubes (replicates or triplicates), suspended in 200 μl of a PBS:Matrigel (1:1) solution and injected subcutaneously into the flank of 4- to 6-week-old female immunodeficient mice (NSG, The Jackson Laboratory). The experiments with the GEMM-derived cells were performed transplanting 10^6 cells per mouse ensuring an in vivo representation of ≈80 cells/barcode. For the human PDX experiments, each injection was performed with $5x10^6$ cells to guarantee an in vivo coverage of ≈400 cells/barcode. Specifically, the TIC in vivo study with the MDA-PATX53 to modulate the appropriate coverage in the human models was executed in triplicate with 10^6, $3x10^6$ and $5x10^6$ cells from the same infection. Mice were monitored every 5 days and euthanized when the tumors reached a volume around 750 mm^3 as determined by caliper measurement. Tumor volume was calculated using the formula: $V=l^2*L/2$ (l length; L width). The whole tumor was collected from each mouse under sterile conditions, weighed and snap-frozen in liquid nitrogen.

***In vivo* shRNA screens.** Infectivity was determined sample by sample as the % of GFP positive cells 2 days after infection as measured by FACS analysis (see *In vivo* TIC covering studies). *In vivo* shRNA screens were performed al least in replicate. For large scale infection of human PDX cells, 60 million cells were plated in T-175 flasks (Corning) with fresh media containing 8 μg/mL polybrene and sufficient virus to guarantee a 25% infection rate based on precedent calculations. For infection of GEMM-derived cells, 20 million cells were plated in T-75 flasks

(Corning) with fresh media containing 8 µg/mL polybrene and sufficient virus to guarantee a 15% infection rate based on precedent calculations. 24 hours after infection, the culture media was replaced with fresh media containing puromycin. 72 hours following puromycin addiction, cells were trypsinized, pooled together and counted. A representative portion of the total cells (normally 1/3 or 1/4) was collected as Reference Cells and immediately frozen as pellet at -80°C. The cells for the *in vivo* screens were separated into independent tubes (replicates or triplicates), suspended in 200 µl of a PBS:Matrigel (1:1) solution and injected subcutaneously into the flank of 4- to 6-week-old female immunodeficient mice (NSG, The Jackson Laboratory). The experiments with the GEMM-derived cells were performed transplanting 10^6 cells per mouse ensuring an *in vivo* representation of ≈400 cells/barcode. For the human PDX experiments, each injection was performed with $5x10^6$ cells to guarantee an *in vivo* coverage of ≈2000 cells/barcode. Mice were monitored every 5 days and euthanized when the tumors reached a volume around 750 mm^3 as determined by caliper measurement. Tumor volume was calculated using the formula: $V=l^2*L/2$ (l length; L width). The whole tumor was collected from each mouse under sterile conditions, weighed and snap-frozen in liquid nitrogen.

Genomic DNA extraction and PCR for NGS library production. Frozen tumors from *in vivo* experiments were minced to small pieces through mechanical procedure with sterile scalpels and suspended in buffer P1 (QIAGEN, 1 mL Buffer/100 mg tumor) supplemented with 100 µg/mL RNase A (Promega). The dissociation step was performed in disposable gentleMACS M tubes (Miltenyi Biotech) with the gentleMACS dissociator (Miltenyi Biotec). The cell pellet obtained from the Reference Cells was suspended in 1 mL buffer P1/RNAse A. Samples were transferred in a 15 mL polypropylene tube (Falcon) and lysed adding 1/20 volume of 10% SDS (Promega). After mixing, the Reference Cells lysates were incubated at RT for 5 minutes and the tumors for 20 minutes. Genomic DNA was sheared by passing the lysate 10-15 times

through a 22-gauge syringe needle. Then, a first genomic DNA extraction step was executed adding 1 volume of Phenol:Chloroform pH 8.0 (Sigma Aldrich). After centrifugation (12000 rpm, 12 minutes), the upper phase was moved to a new tube and a second extraction step with Chloroform (Sigma Aldrich) was performed. Again, the upper phase was transferred to a new tube and added with 0.1 volumes of 3M NaCl (Sigma Aldrich) and 0.8 volumes of isopropanol (Fisher Scientific) to precipitate the genomic DNA. Centrifugation of tumor samples was performed at 12000 rpm for 20 minutes, the samples from Reference Cells were stored over-night at -20°C before centrifugation. DNA pellet was washed once in 70% ethanol (Fisher Scientific) and centrifuged again for 5 minutes at 12000 rpm. The DNA pellet was finally air-dried and dissolved over-night in UltraPure distilled water (Invitrogen). The final DNA concentration was assessed by NanoDrop 2000 (Thermo Scientific) quantification. For NGS libraries generation, the barcodes were amplified starting from the total amount of genomic DNA in 2 rounds of PCR using the Titanium Taq DNA polymerase (Clontech-Takara) and pooling together the total material from the first PCR before proceeding with the second run. The first PCR reactions were performed for 16 cycles with the common primer 13K_R2 (5'- AGTAGCGTGAAGAGCAGAGAA-3') and the specific primers for *in vivo* TIC covering studies, FHTS3 (5'-TCGGATTCAAGCAAAAGACGGCATA-3') or *in vivo* screens, 13K_F2 (5'-TCGGATTCGCACCAGCACGCTA-3'). The second PCR reactions were performed for 12 cycles with the common primer P5_NR2 (5-AATGATACGGCGACCACCGAGACGAGCACCGACAACAACGCAGA-3') and the specific primers for *in vivo* TIC covering studies, Gx1_Bp (5'-TCAAGCAGAAGACGGCATACGAAGACA-3') or *in vivo* screens, P7_NF2 (5'-CAAGCAGAAGACGGCATACGATTCGCACCAGCACGCCTACGCA-3'). The primers for the second PCR reactions were optimized in order to introduce the required adapters for Illumina NGS technology. The PCR amplifications were analyzed by agarose gel electrophoresis (2.5%, Lonza) to check for the expected 262 bp (*in vivo* TIC covering studies) or 272 bp (*in vivo* screens) products. Amplified PCR products from 2 replicates of the second PCR reactions were pooled together and extracted from agarose gel with the QIAquick gel

purification kit (QIAGEN). The amount of purified PCR product was quantified using the High Sensitivity DNA Assay (Agilent Technologies) for the Agilent 2100 Bioanalyzer. Barcode representation was measured by Next Generation Sequencing on an Illumina HiSeq 2000 with a common sequencing primer for both the libraries, 13K_Seq (5'-AGAGGTTCAGAGTTCTACAGTCCGAA-3').

Bioinformatic data analysis. *Read Counting:* Illumina base calls were processed using CASAVA (version 1.8.2) and resulting reads were processed using our in-house pipeline. Raw FASTQ files are filtered for a 4 base pair spacer (CGAA) starting at 18th base allowing for one mismatch to account for sequencing errors, such that only reads amplified using above mentioned PCR reactions are used for further processing. We then extract 23-40bp of the above reads for targeting libraries, and 1-18 bp for non-targetting library. These are further aligned using Bowtie (2.0.2) to their respective libraries (2.4k mouse Epigenome, 2.4k human Epigenome and 12.5k non-targetting scramble)[231]. We then use samtools to count the number of reads aligned to each barcode[232].

Complexity Analysis: Read counts are normalized for the amount of sequencing reads retrieved for each sample, using library size normalization (to 100 million reads).

Hit Analysis: A similar approach was employed as with complexity analysis, described above. Using normalized counts, each sample is then compared with its respective reference and a Log_2 Fold Change is calculated. This is further normalizing using a robust Z-Score defined by: (FC_i – Median)/Median absolute deviation (MAD)[233]. To summarize the effect of knock-down at gene level we employed RSA, to score each gene[234].

Patient-derived samples. Patient-derived samples were obtained from consented patients under an Institutional Review Board (IRB)-approved protocol LAB07-0854 chaired by J.B.F. (UTMDACC).

shRNA expression and gene down-regulation. Human PDX or GEMM tumor cells were transduced with independent pLKO shRNA from concentrated or fresh virus (see Virus preparation) in the presence of 8 μg/mL polybrene. 24 hours after infection the culture media was replaced with a fresh media supplemented with puromycin. Down-regulation of the target was evaluated by western blot at 72hs after shRNA selection. The same infection procedure was applied for the Tet-inducible shRNAs. The top 20% RFP-expressing cells were sorted 72 hours after selection. 48 hours after reseeding, doxycycline was added to the culture to induce shRNA expression. Down regulation of the target was evaluated by western blot at 72 hours after shRNA induction. For the *in vivo* experiments with the Tet-inducible shRNAs, 72 hours after the initiation of the doxycycline-containing (1 mg/mL) drinking water regimen mice were euthanized and tumor collected to confirm shRNA induction by immunohistochemistry analysis.

Flow cytometry, cell sorting, immunohistochemistry, immunoblotting analysis and subcellular protein fractionation

Flow cytometry. Single cells isolated from PDX were stained with primary antibodies after blocking with 10% BSA and 5% rat serum. HLA-ABC (BD Biosciences) and H-2Kd (BD Biosciences) staining were performed according to manufacturer's instructions. To study the cell cycle of PDX tumor cells, BrdU Flow Kit (BD Pharmingen) was used according to datasheet specifications. DAPI (Invitrogen) was used to stain DNA content or to exclude dead cells depending on the experiment. After staining, samples were acquired using a BD FACSCantoII flow cytometer and BD LSRFortessa analyzer. Cell sorting of the top 20% RFP-expressing cells infected with Tet-inducible shRNAs was performed using BD FACSAria II cell sorter. Data were analyzed by BD FACSDiva or FlowJo (Tree Star).

ImmunoHistoChemistry. Tumor samples were fixed in 4% formaldehyde for 2 to 4 hours on ice, moved in 70% ethanol for 12 hours, and then embedded in paraffin (Leica ASP300S).

After cutting (Leica RM2235), baking and deparaffinization, slides were treated with Citra-Plus Solution (BioGenex) according to specifications. For IHC staining, endogenous peroxidases were inactivated by 3% hydrogen peroxide. Non-specific signals were blocked using 3% BSA, 10% goat serum and 0.1% triton. Tumor samples were stained with primary antibodies. For IHC, ImmPress and ImmPress-AP (Vector Lab) were used as secondary antibodies and Nova RED and DAB were used for detection (Vector Lab). Images (10X or 20X magnification) were captured with a Nikon DS-Fi1 digital camera using a wide-field Nikon EclipseCi microscope.

Immunoblotting. Protein lysates were resolved on 5-15% gradient polyacrylamide SDS gels and transferred onto Nitrocellulose membranes according to standard procedures. Membranes were incubated with indicated primary antibodies, washed, and probed with HRP-conjugated secondary antibodies. The detection of bands was carried out upon chemiluminescence reaction followed by film exposure.

Subcellular protein fractionation. Separation and preparation of cytoplasmic and chromatin-bound protein extracts were performed applying the Subcellular Protein Fractionation Kit (Thermo Scientific) according to vendor's specifications.

***In vitro* studies.** Cell viability was measured using MTT assay (Sigma Aldrich) or Cell Titer Glo (Promega) at various time points. Colony Formation Assay (CFA) was carried out seeding 2000/cells in a 6 well plate in replicate. After 10-15 days colonies were visualized by staining with Crystal Violet. For the experiment with the Tet-inducible shRNAs, cells were treated with 0 or 200 ng/mL of Dox after plating. The cells were incubated in the presence of the ATR inhibitor (VE-821) or Vehicle (0.1% DMSO) for the whole period of the experiment. Replication inhibitor Hydroxyurea (HU) was used at a concentration 2 mM and cells were harvested 1 hour after drug application. Experiments with the CRISPR/Cas9 system were performed using a single lentiviral vector (Cellecta, pRSGC1) carrying both the sgRNA for WDR5 and the Cas9 nuclease.

RNA-Seq and data analysis. RNA profiling was performed at the UTMDACC Sequencing and Microarray Facility. RNA was isolated 72 hours after puromycin-selection from PDX or GEMM tumor cells infected with constitutive WDR5 shRNAs or control (shNT) using RNeasy Mini Kit (Qiagen) according to technical specifications. Raw reads from PATX53 and PATC66 cells were aligned to the hg19 assembly of the Human genome using STAR aligner[235]. The raw reads from the p48-Cre, KP53 cells were aligned to XX assembly of Mouse genome using STAR aligner. Gene annotations from the ENSEMBL were used to quantify gene expression information using HTSeq pipeline[236]. The raw read matrix was filtered with median absolute deviation threshold and differential expression was carried out using edgeR Bioconductor package[237]. As the sequencing was carried out in two different batches, we checked the possibility of presence of the batch effect. No batch effect was found. We used ENSEMBL annotations to identify 1:1 orthologs from Human and Mouse using biomaRt package. Genes with at least 1.5 fold-change and 1% false discovery rate (FDR) were considered differentially expressed for overlap analysis. To identify functional enrichment, we used $-10*\log_{10}$(p-values) of differential expression and direction of fold change to generate ranked lists and pathway information from mSigDB with GSEA algorithm[238].

ChIP-Seq and data analysis. For characterization of genome-wide binding profiles, the raw reads were aligned to hg19 and mm9 genome assemblies using Bowtie2 aligner and the enriched peaks were identified using MACS algorithm[239,240]. The MACS algorithm was run with the combination of control and WDR5-shRNA treated samples that were enriched for H3K4me3. Therefore, the peaks identified in such analysis would be those lost upon WDR5 knock-down. Next, ChIPPeakAnno bioconductor package was used with ENSEMBL gene TSS annotations to assign target genes with a distance threshold of +/- 5kb to the enriched peak[241]. To integrate loss of genome wide binding of H3K4me3 with gene expression profiles, we define promoter enriched H3K4me3 peaks as a gene signature. We then checked

enrichment of these signatures with the ranked gene lists defined using RNA-seq profiles of control to WDR5 knock-down experiments in GSEA software.

Virus preparation, Lentiviral-based Somatic Mosaic model, Surgical procedures. *Virus preparation.* Viral particles were produced using 2nd generation packaging plasmids psPAX2 and pMD2.G generated by Didier Trono Lab and obtained by Addgene. 293T cells (ATCC) were cultured in DMEM supplemented with 10% FBS (Gibco), 100IU/mL Penicillin (Gibco), 100µg/mL Streptomycin (Gibco), 4mM Caffeine (Sigma Aldrich) and transfected using the Polyethylenimine (PEI). Supernatant was collected 48-72 hours after transfection, filtered through 0.45µm low-protein binding filters (Corning) and concentrated 100X in sterile PBS (Gibco) after ultracentrifugation at 23000 rpm for 2 hours. Concentrated virus was used fresh or stored at -80°C for further applications.

Lentiviral Somatic Mosaic GEM model. Embryonic liver progenitors (E12.5/E13) were isolated and cultured according to Zender et al.[125]. Cells were infected 24 hours after seeding with the pLSM5 virus and $2x10^5$ cells were injected in the tail of the pancreas of $Rag2^{-/-}$ mice. Before cell injection, $Rag2^{-/-}$ mice approximately 8 to 10 weeks old received 7 consecutive (one each day) intraperitoneal injections of 10µg/Kg caerulein. Animals were monitored weekly for tumor formation. Tumor bearing mice were treated with tamoxifen (5x 100ul injections, 15mg/ml) and tumor volume was measured by MRI (see below).

Orthotopic injections. Mice were anesthetized using a Ketamine/Xylazine solution (150 mg/Kg, 10 mg/Kg). Shaved skin was disinfected with Betadine and Ethanol. Incisions (1 cm) were performed through the skin/subcutaneous and muscular/peritoneal layers. The spleen and tail of the pancreas were exposed and cells were directly injected in the tail of the pancreas. The muscular/peritoneal planes were closed by continuous resorbable sutures. The

skin/subcutaneous planes were closed by interrupted resorbable sutures. Analgesia was achieved with Buprenorphine (0.1 mg/Kg BID).

MRI and Ivis imaging. Animals were imaged on a 4.7T Bruker Biospec (Bruker BioSpin, Billerica) equipped with 6 cm inner diameter gradients and 35 mm inner diameter volume coil. Multi-slice T2-weighted images were acquired in coronal and axial geometries using a rapid acquisition with relaxation enhancement (RARE) sequence with TR/TE of 2000/38 ms, matrix size 256x192, 0.75 mm slice thickness, 0.25 mm slice gap, 4x3 cm FOV, 101 kHz bandwidth, 3 NEX. Axial scan sequences were gated to reduce respiratory motion. Detection of luciferase activity was performed in an IVIS-100 imaging system. Mice were injected i.p. with the D-Luciferin bioluminescence substrate (Perkin Elmer) 5 minutes before the procedure, according to manufacturer's instruction. The Living Image 4.3 software (Perkin Elmer) was used for analysis of the images post-acquisition.

Statistical Analysis. *In vitro* and *in vivo* data are presented as the mean ± s.d. (standard deviation). Statistical analyses were performed using a two-tailed Student's t-test. Results from survival experiments were analyzed with a Log-rank (Mantel-Cox) test and expressed as Kaplan–Meier survival curves.

Bibliography

1. Rahib, L., *et al*. (2014) Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**:2913-21.

2. Faca, V.M., et al. (2008) A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med*. **5**:e123.

3. Ryan, D.P., Hong, T.S., and Bardeesy, N. (2014) Pancreatic adenocarcinoma. *New Engl. J. Med*. **371**:2140-1.

4. Garrido-Laguna, I., and Hidalgo, M. (2015) Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. *Nat. Rev. Clin. Oncol*. doi: 10.1038/nrclinonc.2015.53.

5. Zagouri, F., *et al*. (2013) Molecularly targeted therapies in metastatic pancreatic cancer: a systematic review. *Pancreas*. **42**:760-73.

6. Senderowicz, A.M., *et al.* (2007) Erlotinib/gemcitabine for first-line treatment of locally advanced or metastatic adenocarcinoma of the pancreas. *Oncology*. **21**:1696-706.

7. Von Hoff, D.D., *et al*. (2013) Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *New Engl. J. Med*. **369**:1691-703.

8. Conroy, T., *et al*. (2011) FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *New Engl. J. Med*. **364**:1817-25.

9. Oettle, H., *et al*. (2014) Second-line oxaliplatin, folinic acid, and fluorouracil versus folinic acid and fluorouracil alone for gemcitabine-refractory pancreatic cancer: outcomes from the CONKO-003 trial*. J. Clin. Oncol*. **32**: 2423-2429.

10. Teague, A., Kian-Huat, L., and Wang-Gillam, A. (2015) Advanced pancreatic adenocarcinoma: a review of current treatment strategies and developing therapies. *Ther. Adv. Med. Oncol*. **7**:68-84.

11. Moore, M.J., *et al.* (2007) Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advance pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group*. J. Clin. Oncol.* **25**:1960-1966.

12. Fjallskog, M.L., *et al.* (2003) Expression of molecular targets for tyrosine kinase receptor antagonists in malignant endocrine pancreatic tumors. *Clin. Cancer Res*. **9**:1469-1473.

13. Tobita, K., *et al*. (2003) Epidermal growth factor receptor expression in human pancreatic cancer: significance for liver metastasis. *Int. J. Mol. Med.* **11**:305-309.

14. Morris, J.P., Wang, S.C., and Hebrok, M. (2010) KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. *Nat Rev Cancer*. **10**:683-695.

15. Kanda, M., *et al*. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. *Gastroenterology*. **142**:730-733.

16. Collins, M.A., *et al.* (2012) Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. *J Clin Invest*. **122**:639-653.

17. Bryant, K.L., Mancias, J.D., Kimmelman, A.C., and Der, C.J. (2014) KRAS: feeding pancreatic cancer proliferation. *Trends in Biochemical Sciences.* **39**:91-100.

18. Macdonald, J.S., *et al*. (2005) A phase II study of farnesyl transferase inhibitor R11577 in pancreatic cancer: a Southwest Oncology Group (SWOG 9924) study. *Invest. New Drugs*. **23**:485-487.

19. Van Cutsen, E., *et al.* (2004) Phase III trial of gemcitabine plus tipifarnib compared with gemcitabine plus placebo in advanced pancreatic cancer. *J. Clin. Oncol*. **22**:1430-1438.

20. Collins, M., and Pasca di Magliano, M. (2014) Kras as a key oncogene and therapeutic target in pancreatic cancer. *Front Physiol*. **4**:407.

21. Shimizu, T., *et al*. (20120 The clinical effect of the dual-targeting strategy involving PI3K/AKT/mTOR and RAS/MEK/ERK pathways in patients with advanced cancer. *Clin. Cancer. Res*. **18**:2316-2325.

22. Hingorani, S.R., *et al*. (2003) Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*. **4**:437-450.

23. Aguirre, A.J., *et al*. (2003) Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev*. **17**:3112-26.

24. Bardeesy, N., *et al*. (2006) Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl. Acad. Sci. USA*. **103**:5947-5952.

25. Bardeesy, N., *et al*. (2006) Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev*. **20**:3130-3146.

26. Witkiewicz, A.K., *et al*. (2015) Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun*. **6**:6744.

27. Biankin, A.V., *et al*. (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. **491**:399-405.

28. Waddell, N., *et al.* (2015) Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. **518**:495-501.

29. Villarroel, M.C., *et al*. (2011) Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. *Mol. Cancer Ther.* **10**:3-8.

30. Van der heijden, M.S., *et al*. (2005) In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. *Clin. Cancer Res*. **11**:7508-7515.

31. Sonnenblick, A., *et al*. (2011) Complete remission, in BRCA2 mutation carrier with metastatic pancreatic adenocarcinoma, treated with cisplatin based therapy. *Cancer Biol. Ther*. **12**:165-168.

32. Dawson, M.A., and Kouzarides, T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell*. **150**: 12-27.

33. Sarkies, P., and Sale, J.E. (2011) Cellular epigenetic stability and cancer. *Trends Genet*. **28**:118-127.

34. Jones, P.A., and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet*. **3**:415-428

35. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*. **489**:519-525.

36. Cancer Genome Atlas Network (2012) Comprehensive genomic characterization of squamous cell lung cancer. *Nature*. **482**:519-525.

37. Laird, P.W., *et al*. (1995) Suppression of intestinal neoplasia by DNA hypomethylation. *Cell*. **81**:197-205.

38. Prokhorchouk, A., *et al*. (2006). Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell. Biol.* **26**:199-208.

39. Sansom, O.J., Berger, J., Bishop, S.M., Hendrich, B., Bird, A., and Clarke, A.R. (2003) Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nat. Genet*. **34**:145-147.

40. DeCarvalho, D.D., *et al*. (2012) DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell.* **21**:655-667.

41. Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*. **458**:719-724.

42. Miller, T., *et al*. (2001) "COMPASS: a complex of proteins associated with a trithorax-related SET domain protein." *Proc. Natl. Acad. Sci. USA.* **98**:12902-12907.

43. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*. **128**:693-705.

44. Smith, E., Lin C. and Shilatifard A. (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes Dev*. **25**:661-672.

45. Biankin, A. V., *et al*. (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. **491**:399-405.

46. Deb, M., *et al*. (2014) Chromatin dynamics: H3K4 methylation and H3 variant replacement during development and in cancer. *Cell. Mol. Life Sci.* **71**:3439-3463.

47. Yokoyama, A., *et al*. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell. Biol*. **24**:5639-5649.

48. Cho, Y. W., *et al*. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J. Biol. Chem*. **282**:20395-20406.

49. Patel, S. R., D. Kim, I. Levitan and G. R. Dressler (2007). The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev. Cell*. **13**:580-592.

50. Herz, H. M., D. Hu and A. Shilatifard (2014). Enhancer malfunction in cancer. *Mol. Cell*. **53**:859-866.

51. Dou, Y., *et al*. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*. **121**:873-885

52. Zhang, P., H. Lee, J. S. Brunzelle and J. F. Couture (2012). The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res*. **40**:4237-4246.

53. Milne, T. A., *et al*. (2005). Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc. Natl. Acad. Sci. USA*. **102**:749-754.

54. Liu, H., Cheng E.H., and Hsieh J.J. (2007). Bimodal degradation of MLL by SCFSkp2 and APCCdc20 assures cell cycle execution: a critical regulatory circuit lost in leukemogenic MLL fusions. *Genes Dev*. **21**:2385-2398

55. Liu, H., *et al.* (2010). Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature*. **467**:343-346.

56. Ali, A., Veeranki S.N., and Tyagi S. (2014). A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. *Nucleic Acids Res*. **42**:7611-7624.

57. Guo, S., and Kemphues, K. J. (1995) par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*. **81**: 611–620.

58. Fire, A., *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* **391**: 806–811.

59. Jorgensen, R. A., Cluster, P. D., English, J., Que, Q., and Napoli, C. A. (1996) Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. *Plant Mol. Biol.* **31**: 957–973.

60. Kumagai, M. H., *et al.* (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proc. Natl Acad. Sci. USA* **92**: 1679–1683.

61. Kennerdell, J. R., and Carthew, R. W. (1998) Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell.* **95**: 1017–1026.

62. Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. (1999) Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.* **13**: 3191–3197.

63. Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000) An RNA-directed nuclease mediates posttranscriptional gene silencing in *Drosophila* cells. *Nature.* **404**: 293–296.

64. Hamilton, A. J., and Baulcombe, D. C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science.* **286**: 950–952.

65. Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000) RNAi: double-stranded RNA directs the ATPdependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell.* **101**: 25–33.

66. Hannon, G. J. (2002) RNA interference. *Nature.* **418**: 244-251.

67. Filippov, V., Solovyev, V., Filippova, M., and Gill, S. S. (2000) A novel type of RNase III family proteins in eukaryotes. *Gene.* **245**: 213–221.

68. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature.* **409**: 363–366.

69. Ketting, R. F., *et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**: 2654–2659.

70. Blaszczyk, J., *et al.* (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure (Camb.)*. **9**: 1225–1236.

71. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**: 6877–6888.

72. Nykanen, A., Haley, B., and Zamore, P. D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell.* **107**: 309–321.

73. Mourrain, P., *et al.* (2000) *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell.* **101**: 533–542.

74. Bohmert, K., *et al.* (1998) *AGO1* defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* **17**: 170–180.

75. Ketting, R. F., *et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**: 2654–2659.

76. Knight, S. W., and Bass, B. L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science.* **293**: 2269–2271.

77. Grishok, A., *et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell.* **106**: 23–34.

78. Hutvagner, G, *et al.* (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science.* **293**: 834–838.

79. Olsen, P. H.,, and Ambros, V. (1999) The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**: 671–680.

80. Slack, F. J., *et al.* (2000) The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell.* **5**: 659–669.

81. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. and Conklin, D. S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**: 948–958.

82. Wianny, F., and Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development. *Nature Cell Biol.* **2**: 70–75.

83. Svoboda, P., Stein, P., Hayashi, H. and Schultz, R. M. (2000) Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development.* **127**: 4147–4156.

84. Elbashir, S. M., *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* **411**: 494–498.

85. Clarke, P. A. and Mathews, M. B. (1995) Interactions between the double-stranded RNA binding motif and RNA: definition of the binding site for the interferon-induced protein kinase DAI (PKR) on adenovirus VA RNA. *RNA.* **1**: 7–20.

86. Brummelkamp, T. R., Bernards, R. and Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science.* **21,** 21.

87. Sui, G., *et al.* (2002) A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl Acad. Sci. USA.* **99**: 5515–5520.

88. Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell.* **115**: 209−216.

89. Schwarz, D. S., *et al*. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell.* **115**: 199−208.

90 Reynolds, A., *et al*. (2004) Rational siRNA design for RNA interference. *Nature Biotechnol.* **22**: 326−330.

91. Silva, J. M., Sachidanandam, R., and Hannon, G. J. (2003) Free energy lights the path toward more effective RNAi. *Nature Genet.* **35**: 303−305.

92. Hsieh, A. C., *et al*. (2004) A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32**: 893−901.

93. Brummelkamp, T. R., Nijman, S. M., Dirac, A. M., and Bernards, R. (2003) Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-κB. *Nature.* **424**: 797−801.

94. Berns, K., *et al*. (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature.* **428**: 431−437.

95. Zheng, L., *et al*. (2004) An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA.* **101**: 135−140.

96. Luo, B., Heard, A. D., and Lodish, H. F. (2004) Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc. Natl Acad. Sci. USA.* **101**: 5494−5499.

97. Shirane, D., *et al*. (2004) Enzymatic production of RNAi libraries from cDNAs. *Nature Genet.* **36**: 190−196.

98. Sen, G., Wehrman, T. S., Myers, J. W., and Blau, H. M. (2004) Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nature Genet.* **36**: 183−189.

99. Silva, J. M., Mizuno, H., Brady, A., Lucito, R., and Hannon, G. J. (2004) RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**: 6548−6552.

100. Mullenders, J., and Bernards, R. (2009) Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. *Oncogene*. **28**: 4409-20.

101. Sigoillot, F. D., and King, R. W. (2011) Vigilance and validation: keys to success in RNAi screening. *ACS Chem. Biol.* **6**: 47–60.

102. Echeverri, C. J. *et al.* (2006) Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Methods.* **3**: 777–779.

103. Hu, Y. *et al.* (2013) UP-TORR: online tool for accurate and Up-to-Date annotation of RNAi Reagents. *Genetics.* **195**: 37–45.

104. Schmidt, E. E. *et al.* (2013) GenomeRNAi: a database for cell- based and *in vivo* RNAi phenotypes, 2013 update. *Nucleic Acids Res.* **41**: D1021–D1026.

105. Bassik, M. C. *et al.* (2009) Rapid creation and quantitative monitoring of high coverage shRNA libraries. *Nat Methods* **6**: 443–445.

106. Hoffman, G. R. *et al.* (2014) Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl Acad. Sci. USA.* **111**: 3128–3133.

107. Gazin, C., *et al*. (2007) An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature*. **449**: 1073-7.

108. Smolen, G.A., *et al.* (2010) A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration. *Genes Dev*. **24**:2654-65.

109. Sheng, Z., *et al*. (2010) A genome-wide RNA interference screen reveals an essential CREB3L2-ATF5-MCL1 survival pathway in malignant glioma with therapeutic implications. *Nat Med.* **16**: 671-7.

110. Coussens, M., *et al*. (2010) RNAi screen for telomerase reverse transcriptase transcriptional regulators identifies HIF1alpha as critical for telomerase function in murine embryonic stem cells. *Proc Natl Acad Sci USA*. **107**: 13842-7.

111. Popov, N., *et al*. (2007) The ubiquitin-specific protease USP28 is required for MYC stability. *Nat Cell Biol*. **9**:765-74.

112. Gumireddy, K., *et al*. (2009) KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nat Cell Biol*. **11**: 1297-304.

113. Boettcher, M., and Hoheisel, J.D. (2010) Pooled RNAi Screens - technical and biological aspects. *Curr Genomics.* **11**: 162-7.

114. Boettcher, M., *et al*. (2010) Decoding pooled RNAi screens by means of barcode tiling arrays. *BMC Genomics.* **11**: 7.

115. Ngo, V.N., *et al*. (2006) A loss-of-function RNA interference screen for molecular targets in cancer. *Nature*. **441**: 106-10.

116. Schlabach, M.R., *et al*. (2008) Cancer proliferation gene discovery through functional genomics. *Science*. **319**: 620-4.

117. Hartwell, L.H., *et al*. (1997) Integrating genetic approaches into the discovery of anticancer drugs. *Science*. **278**: 1064-8.

118. Luo, J., *et al*. (2009) A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*. **137**: 835-48.

119. Scholl, C., *et al.* (2009) Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell*. **137**: 821-34.

120. Lam, L.T., *et al*. (2008) Compensatory IKKalpha activation of classical NF-kappaB signaling during IKKbeta inhibition identified by an RNA interference sensitization screen. *Proc Natl Acad Sci USA.* **105**: 20798-803.

121. Wiltshire, T.D., *et al*. (2010) Sensitivity to poly(ADP-ribose) polymerase (PARP) inhibition identifies ubiquitin-specific peptidase 11 (USP11) as a regulator of DNA double-strand break repair. *J Biol Chem*. **285**: 14565-71.

122. Azorsa, D.O., *et al*. (2009) Synthetic lethal RNAi screening identifies sensitizing targets for gemcitabine therapy in pancreatic cancer. *J Transl Med.* **7**: 43.

123. Whitehurst, A.W., *et al*. (2007) Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature*. **446**: 815-9.

124. O'Connell, B.C., *et al.* (2010) A genome-wide camptothecin sensitivity screen identifies a mammalian MMS22L-NFKBIL2 complex required for genomic stability. *Mol Cell.* **40**: 645-57.

125. Zender, L., *et al*. (2008) An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell.* **135**: 852–864.

126. Bric, A., *et al.* (2009) Functional identification of tumor-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell.* **16**: 324–335.

127. Meacham, C.E., Ho, E.E., Dubrovsky, E., Gertler, F.B., and Hemann, M.T. (2009) *In vivo* RNAi screening identifies regulators of actin dynamics as key determinants of lymphoma progression. *Nat Genet.* **41**: 1133–1137.

128. Gargiulo, G., et al. (2013) *In vivo* RNAi screen for BMI1 targets identifies TGF- β /BMP-ER stress pathways as key regulators of neural- and malignant glioma–stem cell homeostasis. *Cancer Cell.* **23**: 660–676.

129. Xue, W., *et al.* (2012) A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proc. Natl. Acad. Sci. USA.* **109**: 8212–8217.

130. Iorns, E., *et al.* (2012) Whole-genome *in vivo* RNAi screening identifies the leukemia inhibitory factor receptor as a novel breast tumor suppressor. *Breast Cancer Res. Treat.* **135**: 79–91.

131. Gargiulo, G., *et al.* (2014) *In vivo* shRNA screens in solid tumors. *Nat Protoc*. **9**: 2880-2902.

132. Possemato, R., *et al.* (2011) Functional genomics reveal that serine synthesis pathway is essential in breast cancer. *Nature*. **476**: 346-350.

133. Possik, P.A., *et al.* (2014) Parallel in vivo and in vitro melanoma RNAi dropout screens reveal synthetic lethality between hypoxia and DNA damage response inhibition. *Cell Rep*. **9**: 1375-1386.

134. Johnson, J.I., *et al*. (2001) Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br. J. Cancer*. **84**: 1424–1431.

135. Daniel, V.C., *et al*. (2009) A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture *in vitro*. *Cancer Res*. **69**: 3364–3373.

136. Giovanella, B.C., *et al*. (1989) DNA topoisomerase I–targeted chemotherapy of human colon cancer in xenografts. *Science*. **246**: 1046–1048.

137. Houghton, J.A., Maroda, S.J. Jr., Phillips, J.O., and Houghton, P.J. (1981) Biochemical determinants of responsiveness to 5-fluorouracil and its derivatives in xenografts of human colorectal adenocarcinomas in mice. *Cancer Res*. **41**: 144–149.

138. Houghton, J.A., and Taylor, D.M. (1978) Growth characteristics of human colorectal tumours during serial passage in immune-deprived mice. *Br. J. Cancer*. **37**: 213–223.

139. Jin, K., *et al*. (2010) Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review. *Clin. Transl. Oncol*. **12**: 473–480.

140. Morton, C.L., and Houghton, P.J. (2007) Establishment of human tumor xenografts in immunodeficient mice. *Nat. Protoc*. **2**: 247–250.

141. Rubio-Viqueira, B., and Hidalgo, M. (2009) Direct *in vivo* xenograft tumor model for predicting chemotherapeutic drug response in cancer patients. *Clin. Pharmacol. Ther*. **85**: 217–221.

142. Sausville, E.A., and Burger, A.M. (2006) Contributions of human tumor xenografts to anticancer drug development. *Cancer Res*. **66**: 3351–3354.

143. Tentler, J.J., *et al.* (2012) Patient-derived xenografts as a model for oncology drug development. *Nat Rev Clin Oncol*. **9**: 338-350.

144. Rubio-Viqueira, B., *et al.* (2006) An *in vivo* platform for translational drug development in pancreatic cancer. *Clin. Cancer Res*. **12**: 4652–4661.

145. John, T., *et al*. (2011) The ability to form primary tumor xenografts is predictive of increased risk of disease recurrence in early-stage non-small cell lung cancer. *Clin. Cancer Res*. **17**: 134–141.

146. Merk, J., Rolff, J., Becker, M., Leschber, G., and Fichtner, I. (2009) Patient-derived xenografts of non-small-cell lung cancer: a pre-clinical model to evaluate adjuvant chemotherapy? *Eur. J. Cardiothorac. Surg*. **36**:454–459.

147. Shultz, L.D., *et al*. (2005) Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R γ null mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* **174**: 6477–6489.

148. Sanz, L., *et al*. (2009) Differential transplantability of human endothelial cells in colorectal cancer and renal cell carcinoma primary xenografts. *Lab. Invest*. **89**: 91–97.

149. Gray, D.R., *et al*. (2004) Short-term human prostate primary xenografts: an *in vivo* model of human prostate cancer vasculature and angiogenesis. *Cancer Res*. **64**: 1712–1721.

150. Garrido-Laguna, I., *et al*. (2011) Tumor engraftment in nude mice and enrichment in stroma- related gene pathways predict poor survival and resistance to gemcitabine in patients with pancreatic cancer. *Clin. Cancer Res*. **17**: 5793–5800.

151. Fichtner, I., *et al*. (2008) Establishment of patient-derived non-small cell lung cancer xenografts as models for the identification of predictive biomarkers. *Clin. Cancer Res.* **14**: 6456–6468.

152. Jones, S., *et al*. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. **321**: 1801–1806.

153. Kola, I., and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*. **3**: 711–5.

154. Johnson, J.I., et al. (2001) Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br J Cancer*. **84**: 1424–31.

155. Garrido-Laguna, I., *et al.* (2010) Integrated preclinical and clinical development of mTOR inhibitors in pancreatic cancer. *Br J Cancer*. **103**: 649–55.

156. Rajeshkumar, N.V., *et al*. (2009) Antitumor effects and biomarkers of activity of AZD0530, a Src inhibitor, in pancreatic cancer. *Clin Cancer Res*. **15**: 4138–46

157. Bertotti, A., *et al*. (2011) A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov*. **1**: 508–23.

158. Julien, S., *et al*. (2012) Characterization of a large panel of patient-derived tumor xenografts representing the clinical heterogeneity of human colorectal cancer. *Clin Cancer Res*. **18**: 5314–28.

159. Anido, J., *et al*. (2010) TGF-beta receptor inhibitors target the CD44(high)/Id1(high) glioma-initiating cell population in human glioblastoma. *Cancer Cell*. **18**: 655–68.

160. Lonardo, E., *et al*. (2011) Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell*. **9**: 433–46.

161. Mueller, M.T., *et al*. (2009) Combined targeted treatment to eliminate tumorigenic cancer stem cells in human pancreatic cancer. *Gastroenterology*. **137**: 1102–13.

162. Yen, W.C., *et al*. (2012) Anti-DLL4 has broad spectrum activity in pancreatic cancer dependent on targeting DLL4-Notch signaling in both tumor and vasculature cells. *Clin Cancer Res*. **18**: 5374–86.

163. Clarke, M.F., *et al*. (2006) Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res*. **66**: 9339–9344.

164. Valent, P., *et al*. (2012) Cancer stem cell definitions and terminology: the devil is in the details. *Nat Rev Cancer*. **12**: 767–775.

165. Al-Hajj, M., *et al*. (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA.* **100**: 3983–3988.

166. Singh, S.K., *et al*. (2004) Identification of human brain tumour initiating cells. *Nature*. **432**: 396–401.

167. Reya. T., *et al*. (2001) Stem cells, cancer, and cancer stem cells. *Nature*. **414**: 105–111.

168. Ishizawa, K., *et al*. (2010) Tumor-initiating cells are rare in many human tumors. *Cell Stem Cell*. **7**: 279-82

169. O'Brien, C.A., *et al*. (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*. **445**: 106–110.

170. Ricci-Vitiani, L., *et al*. (2007) Identification and expansion of human colon-cancer-initiating cells. *Nature*. **445**: 111–115.

171. Hermann, P.C., *et al*. (2007) Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell* **1**: 313–323.

172. Zhang, S., *et al*. (20080 Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Res*. **68**: 4311–4320.

173. Quintana, E., *et al*. (2008) Efficient tumour formation by single human melanoma cells. *Nature*. **456**: 593–598.

174. Lapidot, T., *et al*. (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. **367**:645–648.

175. Bonnet, D., and Dick, J.E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. **3**: 730–737.

176. Furth, J., and Kahn, M. (1937) The transmission of leukemia of mice with a single cell. *Am J Cancer*. **31**: 276–282.

177. Makino, S. (1956) Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Ann N Y Acad Sci* **63**: 818–830.

178. Hewitt, H. (1958) Studies of the dissemination and quantitative transplantation of a lymphocytic leukemia of CBA mice. *Br J Cancer*. **12**: 378–401.

179. Southam, C., Brunschwig, A., and Dizon, Q. (1962) Autologous and homologous transplantation of human cancer. In: Brennan M, Simpson W eds. *Biological Interactions in Normal and Neoplastic Growth: A Contribution to the Tumor-Host Problem. 9*. Little, Brown: Boston, MA, USA, p 723–738.

180. Pece, S., *et al.* (2010) Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell*. **140**: 62-73.

181. Taetle, R., *et al*. (1987) Use of nude mouse xenografts as preclinical screens. Characterization of xenograft-derived melanoma cell lines. *Cancer*. **60**: 1836–1841.

182. Fiebig, H.H, *et al*. (2007) Gene signatures developed from patient tumor explants grown in nude mice to predict tumor response to 11 cytotoxic drugs. *Cancer*

Genomics Proteomics. **4**: 197–209.

184. Schatton, T., *et al*. (2008) Identification of cells initiating human melanomas. *Nature*. **451**: 345–349.

185. Nemati, F., *et al*. (2010) Establishment and characterization of a panel of human uveal melanoma xenografts derived from primary and/or metastatic tumors. *Clin. Cancer Res*. **16**: 2352–2362.

186. Olive, K.P., *et al*. (2009) Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*. **324**: 1457–1461.

187. Kim, M.P., *et al*. (2009) Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. *Nat. Protoc*. **4**: 1670–1680.

188. Fu, X., Guadagni, F., and Hoffman, R.M. (1992) A metastatic nude-mouse model of human pancreatic cancer constructed orthotopically with histologically intact patient specimens. *Proc. Natl Acad. Sci. USA*. **89**: 5645–5649.

189. Garrido-Laguna, I., *et al*. (2010) Integrated preclinical and clinical development of mTOR inhibitors in pancreatic cancer. *Br. J. Cancer*. **103**: 649–655.

190. Rubio-Viqueira, B., *et al*. (2007) Optimizing the development of targeted agents in pancreatic cancer: tumor fine-needle aspiration biopsy as a platform for novel prospective *ex vivo* drug sensitivity assays. *Mol. Cancer Ther*. **6** :515–523.

191. Hidalgo, M., *et al*. (2011) A pilot clinical study of treatment guided by personalized tumorgrafts in patients with advanced cancer. *Mol. Cancer Ther*. **10**: 1311–1316.

192. Villarroel, M.C., *et al*. (2011) Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. *Mol. Cancer Ther*. **10**: 3–8.

193. Jones, S., *et al*. (2009) Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*. **324**: 217.

194. Shamir, E.R., and Ewald, A.J. (2014) Three-dimensional organotypic culture: experimental models of mammalian biology and disease. *Nat Rev Mol Cell Biol*. **15**: 647-664.

195. Krapp, A., *et al.* (1998) The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev*. **12**: 3752-3763.

196. Hingorani, S.R., *et al.* (2003) Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*. **4**: 437-450.

197. Guerra, C., and Barbacid, M. (2013) Genetically engineered mouse models of pancreatic adenocarcinoma. *Mol Oncol*.**7**: 232-247.

198. Bardeesy, N., *et al*. (2006) Smad is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreatic cancer. *Genes Dev*. **20**: 3130-3146.

199. Aguirre, A.J., *et al.* (2003) Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev*. **17**: 3112-3126.

200. Fehse, B., Kustikova, O.S., Bubenheim, M. and Baum, C. (2004). Pois(s)on – It's a question of dose. *Gene Therapy*. **11**: 879-881.

201. Salmon, P., and Trono, D. (2007) Production and titration of lentiviral vectors. *Current Protocols*. DOI: 10.1002/0471142905.hg1210s54

202. Hu, Y., and Smyth, G.K. (2009) ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of Immunological Methods.* **347**: 70-78.

203. Bonnefoix, T,. and Callanan, M. (2010) Revisiting the concept of phenotypically distinct malignant pancreatic stem-cell subsets based on limiting dilution transplantation assays. *JCO*. **28**: 89-90.

204. Hubert, C.G., *et al*. (2013) Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev*. **27**: 1032-1045.

205. Krattenmacher, A., *et al.* (2014) Synthetic lethality screen identifies SMC2 as a new target gene in pancreatic cancer. *Pancreatology*. **13**: S66.

206. Sahai, V., *et al.* (2014) BET bromodomain inhibitors block growth of pancreatic cancer cells in three-dimensional collagen. *Mol. Cancer Ther*. **13**: 1907-1917.

207. Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol.* **13**: 852-854.

208. Trievel, R.C., and Shilatifard, A. (2009) WDR5, a complexed protein. *Nat Struct Mol Biol.* **16**: 678-680.

209. Kim, J.Y., *et al.* (2014) A role for WDR5 in integrating threonine 11 phosphorylation to lysine 4 methylation on histone H3 during androgen signaling and in prostate cancer. *Mol. Cell*. **54**: 613-625.

210. Chen, X., *et al*. (2015) Upregulated WDR5 promotes proliferation, self-renewal and chemoresistance in bladder cancer via mediating H3K4 trimethylation. *Scientific Reports*. **5**: 8293.

211. Viale, A., *et al.* (2014) Oncogene ablation-resistant pancreatic cells depend on mitochondrial function. *Nature*. **514**: 628-632.

212. Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995) Altered Hox expression and segmental identity in Mll-mutant mice. *Nature.* **378**: 505-508.

213. Schuettengruber, B., Martinez, A., Iovino, N., and Cavalli, G. (2011) Trithorax group proteins: switching genes on and keeping them active. *Nature Reviews Molecular Cell Biology*. **12**: 799-814.

214. Wang, P., *et al.* (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol*. **29**: 6074-8532.

215. Nghiem, P., Park, P.K., Kim, Y., Vaziri, C., and Schreiber, S.L. (2001) ATR inhibition selectively sensitizes G1 checkpoint deficient cells to lethal premature chromatin condensation. *Proc Natl Acad Sci USA*. **98**: 9092-9097.

216. Branzei, D., and Foiani, M. (2010) Maintaining genome stability at the replication forks. *Nat Rev Mol Cell Biol*. **11**: 208-219.

217. Cimprich, K. A., and Cortez, D. ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol*. **9**: 616-627.

218. Thomas, L.R., *et al*. (2015) Interaction with WDR5 promotes target gene recognition and tumorigenesis by Myc. *Mol Cell*. **58**: 440-452.

219. Hofmann, G.R. (2014) Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl. Acad. Sci. USA.* **111**: 3128-3133.

220. Slany R.K. (2009) The molecular biology of Mixed Lineage Leukemia. *Haematologica*. **94**: 984-993

221. Balakrishnan, A., *et al.* (2007) Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. *Cancer Res*. **67**: 3545-3550.

222. Cao, F., *et al.* (2014) Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell*. **23**: 247-261.

223. Shlyueva, D., Stampfel, G., and Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet*. **15**: 272-286.

224. Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009) A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol Cell*. **36:** 970-983.

225. Latham, J.A., Chosed, R.J., Wang, S., and Dent, S.Y. (2011) Chromatin signaling to kinetochores: Trans-regulation of Dam1 methylation by histone H2B ubiquitination. *Cell*. **146**: 709-719.

226. Dominguez-Sola , D., *et al.* (2007) Non transcriptional control of DNA replication by c-Myc. *Nature*. **448**: 445-451.

227. Ventura, A., *et al.* (2004) Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci U S A.* **101**: 10380-10385.

228. Ventura, A., *et al.* (2007) Restoration of p53 function leads to tumour regression in vivo. *Nature.* **445**: 661-665.

229. Lao, Z., Raju, G. P., Bai, C., and Joyner, A. L. (2012) MASTR: a technique for mosaic mutant analysis with spatial and temporal control of recombination using conditional floxed alleles in mice. *Cell Reports.* **2**: 386-396.

230. Hunter, N. L., Awatramani, R. B., Farley, F. W. , and Dymecki, S. M. (2005) Ligand-activated Flpe for temporally regulated gene modifications. *Genesis*. **41**: 99-109.

231.Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.

232. Li, H., *et al*. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. **25**: 2078-9.

233. Konig, R., *et al.* (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods*. **4**: 847-849.

234. Birmingham, A., *et al*. (2009) Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods*. **6**: 569-575.

235. Dobin, A., *et al*., (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29v**: 15-21.

236. Anders, S., Pyl, P.T., and Huber, W. (2014) HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics*. doi: 10.1093/bioinformatics/btu638

237. Zhou, X., Lindsay, H., and Robinson, M.D. (2014). "Robustly detecting differential expression in RNA sequencing data using observation weights." *Nucleic Acids Research*, **42**: e91.

238. Subramanian, A., *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*. **102**: 15545-50

239. Langmead, B., and Salzberg, S. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**: 357-359.

240. Zhang, Y. *et al.* (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**: R137.

241. Zhu, L., *et al*. (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data". *BMC Bioinformatics*. **11**: 237.