

Evolutionary trajectories and *KRAS* gene dosage define pancreatic cancer phenotypes

Sebastian Mueller^{1,2*}, Thomas Engleitner^{1,2,3*}, Roman Maresch^{1,2,3*}, Magdalena Zukowska^{1,2}, Sebastian Lange^{1,2}, Thorsten Kaltenbacher^{1,2,3}, Björn Konukiewitz⁴, Rupert Öllinger^{1,2}, Maximilian Zwiebel², Alex Strong⁵, Hsi-Yu Yen^{3,6}, Ruby Banerjee⁵, Sandra Louzada⁵, Beiyuan Fu⁵, Barbara Seidler^{1,2}, Juliana Götzfried², Kathleen Schuck², Zonera Hassan², Andreas Arbeiter², Nina Schönhuber^{1,2}, Sabine Klein^{1,2}, Christian Veltkamp^{1,2}, Mathias Friedrich⁵, Lena Rad², Maxim Barenboim^{2,3}, Christoph Ziegenhain⁷, Julia Hess⁸, Oliver M. Dovey⁵, Stefan Eser², Swati Parekh⁷, Fernando Constantino-Casas¹⁰, Jorge de la Rosa^{5,11,12}, Marta I. Sierra⁹, Mario Fraga^{9,13}, Julia Mayerle¹⁴, Günter Klöppel⁴, Juan Cadiñanos^{5,11}, Pentao Liu⁵, George Vassiliou⁵, Wilko Weichert^{3,4}, Katja Steiger^{4,6}, Wolfgang Enard⁷, Roland M. Schmid^{2,3}, Fengtang Yang⁵, Kristian Unger⁸, Günter Schneider^{2,3}, Ignacio Varela¹⁵, Allan Bradley⁵, Dieter Saur^{1,2,3#}, Roland Rad^{1,2,3#§}

¹Center for Translational Cancer Research (TranslaTUM), Technische Universität München, 81675 Munich, Germany.

²Department of Medicine II, Klinikum rechts der Isar, Technische Universität München, 81675 Munich, Germany.

³German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

⁴Institute of Pathology, Technische Universität München, 81675 Munich, Germany.

⁵The Wellcome Trust Sanger Institute, Genome Campus, Hinxton/Cambridge, CB10 1SA, UK.

⁶Comparative Experimental Pathology, Technische Universität München, 81675 Munich, Germany.

⁷Anthropology & Human Genomics, Department of Biology II, Ludwig-Maximilians Universität, Großhaderner Straße 2, 82152 Martinsried, Germany.

⁸Helmholtz Zentrum München, Research Unit Radiation Cytogenetics, 85764 Neuherberg, Germany.

⁹Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo, Spain.

¹⁰Department of Veterinary Medicine, University of Cambridge, Cambridge, UK.

¹¹Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA), Oviedo, Spain.

¹²Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología (IUOPA), Universidad de Oviedo, Oviedo, Spain.

¹³Nanomaterials and Nanotechnology Research Center (CINN-CSIC)-Universidad de Oviedo-Principado de Asturias, Spain.

¹⁴Medizinische Klinik und Poliklinik II, Klinikum der LMU München-Grosshadern, 81377 Munich, Germany

¹⁵Instituto de Biomedicina y Biotecnología de Cantabria (UC-CSIC), 39011 Santander, Spain

*These authors contributed equally to this work.

#These authors jointly supervised this work.

§To whom correspondence should be addressed. E-mail: roland.rad@tum.de

ABSTRACT

The poor correlation of mutational landscapes with phenotypes limits our understanding of pancreatic ductal adenocarcinoma (PDAC) pathogenesis and metastasis. Here we show a critical role of oncogenic dosage-variation in PDAC biology and phenotypic diversification. We found gene-dosage increase of mutant *KRAS^{MUT}* in human PDAC precursors, driving both early tumorigenesis and metastasis, thus rationalizing early PDAC dissemination. To overcome limitations posed to gene-dosage studies by PDAC's stroma-richness we developed large cell culture resources of metastatic mouse PDAC. Integration of their genomes, transcriptomes and tumor phenotypes with functional studies and human data, revealed additional widespread effects of oncogenic dosage-variation on cell morphology/plasticity, histopathology and clinical outcome, with highest *Kras^{MUT}* levels underlying aggressive undifferentiated phenotypes. We also identify alternative oncogenic gains (*Myc*, *Yap1* or *Nfkb2*), which collaborate with heterozygous *Kras^{MUT}* in driving tumorigenesis, yet with lower metastatic potential. Mechanistically, different oncogenic gains and dosages evolve along distinct evolutionary routes, licensed by defined allelic states and/or combinations of hallmark tumor-suppressor alterations (*Cdkn2a*, *Trp53*, Tgf β -pathway). Thus, evolutionary constraints and contingencies direct oncogenic dosage gain and variation along defined routes to drive early progression and shape downstream PDAC biology. Our study uncovers universal principles in *Ras*-driven oncogenesis with potential relevance beyond pancreatic cancer.

Pancreatic ductal adenocarcinoma (PDAC) is the fourth leading cause of cancer-related death and is expected to become the second within the next decade¹. Whilst treatments constantly improved for many other cancer types, 5-year survival rates in PDAC stayed around 5%¹. Genome sequencing revealed extensive genetic heterogeneity beyond a few frequently mutated drivers²⁻⁸ like *KRAS*, *TP53*, *CDKN2A* or TGF β -pathway alterations. Disappointingly however, genomic changes could so far not be broadly linked to biologic, morphologic or clinical phenotypes. In addition, the molecular basis of cancer cell dissemination is poorly understood, and genetic comparisons of primary/metastasis pairs could not identify recurrent alterations linked to metastasis^{3,8}. Critical limitations to human PDACs (hPDAC) genomics are (i) the cancer genome complexity, posing challenges to their interpretation, (ii) the high (and variable) stromal content, which confounds particularly gene-dosage analyses and transcriptome interpretation, (iii) the limited availability of human cell culture-based resources to overcome this problem and (iv) the scarcity of paired primary/metastasis tissues, particularly of treatment-naïve ones, e.g. for evolutionary studies. Here we characterized large murine PDAC cell line resources and combined the results with cross-species comparisons and functional studies to unravel molecular principles underlying PDAC evolution and phenotypic diversification.

Genetic landscapes of mouse PDAC

We initially characterized primary PDAC cell cultures from 38 mice expressing *Kras*^{G12D} conditionally in the pancreas (PK mice)^{9,10} by multiplex FISH (M-FISH), whole-exome sequencing (WES) and array comparative genomic hybridization (aCGH). We developed a pipeline for WES data analysis allowing mouse/human comparisons using identical parameter settings. A WES study on microdissected human PDAC (reduced stromal “contamination”) served as the reference human data set⁶. Somatic mutation calling identified 318 synonymous and 606 non-synonymous mutations in 38 mPDACs (Extended Data Fig. 1a and Supplementary Table 1). Mutational patterns were similar, (Fig. 1a; Supplementary Table 2, Extended Data Fig. 1b-g) but mutational burdens were significantly higher in hPDAC than mPDAC, with 3.3 and 1.5 fold increased median SNV and indel numbers, respectively (Fig. 1b and Supplementary Tables 1, 3-5). Recurrently altered genes were infrequent in mice (Supplementary Table 1,4). A subset overlapped with recurrently mutated genes in human cancers and/or common insertion sites in pancreas-specific transposon screens¹¹⁻¹³ (Extended Data Fig. 1a).

Structural changes were also less common in mPDAC (Fig. 1b, Supplementary Tables 6-8). There was however substantial variation between cancers, with some mPDACs having only few focal alterations, but others showing extensive changes, including clustered intra-chromosomal alterations, aneuploidy and inter-chromosomal translocations (Fig. 1c and Extended Data Fig. 1h-l). Notably, 34% (14/38) tumors had complex rearrangements, with ten or more alterations per affected chromosome. The majority (12/14) of such events affect chr4, invariably involving *Cdkn2a*. One cancer showed massively rearranged chr15 with high-level *Myc* amplification and another tumor had clustered chr1 rearrangements (Extended Data Fig. 2a-n). These findings reflect selection of complex rearrangements that affect cancer drivers.

The regularity of oscillating copy number states in most cancers suggested chromothripsis as the predominant process underlying these complex alterations. Whole-genome sequencing, followed by rearrangement analysis and computational simulations confirmed all hallmarks defining the one-off nature of chromothripsis¹⁴, including clustering of breakpoints, regularity of oscillating copy number states, identical CNA and LOH patterns, randomness of DNA segment order/joints and alternating head-tail sequences (Fig. 1d and detailed analyses in Extended Data Fig. 2p-y. In addition, M-FISH confirmed chr4 content loss affecting only one haplotype (Fig. 1d).

Complex rearrangements were proposed to trigger accelerated evolution of human PDAC¹⁵. The mouse model allows experimental interrogation of this hypothesis due to the “synchronized” nature of tumor initiation (*Kras*^{G12D}-mutation). We found that time-to-tumor development was indeed shorter in animals with *Cdkn2a* loss through catastrophic events (Fig. 1e and Extended Data Fig. 2o). A subset (16%) of complex rearrangements in hPDAC disrupts multiple known tumor suppressors through translocations¹⁵. Chromothripsis-associated chr4 translocations were also frequent in mice (Fig. 1c), although no recurrent translocation partners were found.

***Kras*^{MUT-iGD} links early progression & metastasis**

The most common amplification affected the *Kras* locus (Extended Data Fig. 3a,b), which is also frequent in hPDAC^{16,17}. Combined analyses of M-FISH, aCGH and *Kras* mutant/wild-type (wt) allele frequencies revealed four different *Kras*^{G12D} gene dosage "states" (Fig. 2a, Extended Data Fig. 3c-h

and Supplementary Table 9): focal gain ($Kras^{G12D-FG}$, 7.9%), arm-level gain ($Kras^{G12D-AG}$, 23.7%), copy number neutral loss of wt $Kras$ ($Kras^{G12D-LOH}$, 36.8%) or no change ($Kras^{G12D-HET}$, 31.6%). Thus, two thirds of cancers had allelic imbalances causing increased $Kras^{G12D}$ gene dosage (hereafter designated $Kras^{G12D-iGD}$), suggesting strong selective pressure for its acquisition. In addition, two $Kras^{G12D-HET}$ tumors displayed loss of $Kras^{WT}$ -mRNA, but high $Kras^{G12D}$ expression (blue dots in Fig. 2b), suggesting additional non-genetic mechanisms. Of note, we observed similar $KRAS^{G12D-iGD}$ rates/types in human PDAC cell lines (Supplementary Table 10). Gene dosage increase affects transcriptional output, as $Kras^{G12D-iGD}$ mPDAC had higher $Kras^{G12D}$ mRNA expression than $Kras^{G12D-HET}$ cancers (Fig. 2b and Extended Data Fig. 3i).

Ras/Raf signaling amplification was observed at different stages of mammary, intestinal or lung tumorigenesis¹⁸⁻²¹. To identify the stage of $KRAS^{MUT-iGD}$ -acquisition in PDAC, we microdissected low-grade human pancreatic intraepithelial neoplasias (hPanIN) from 19 patients and performed amplicon-based deep sequencing of $KRAS$ exon-2 (Fig. 2c and Supplementary Table 11/12). hPanIN with exon-2-mutated $KRAS$ (20/40 hPanINs featured $KRAS^{G12*}$ mutations) displayed frequent $KRAS^{MUT-iGD}$: $KRAS^{MUT}$ allele frequencies >50% occurred in 50%, 38% and 67% of $KRAS$ exon-2-mutated hPanIN1a, hPanIN1b and hPanIN2, respectively. Given that healthy tissue "contamination" rates in microdissected PanINs ranged between 10-60%, $KRAS^{MUT-iGD}$ is likely to be even more frequent. In cases with close to 100% mutant read frequency, $KRAS$ interphase FISH excluded false-positive $KRAS^{G12D-iGD}$ arising through chr12 monosomy (Extended Data Fig. 3l-n). Moreover, false-positive $KRAS^{MUT-iGD}$ through cross-"contaminating" hPDAC is excluded due to either (i) the large distance of selected hPanINs to associated cancers, (ii) distinct $KRAS$ mutations in hPanINs and associated cancers, or (iii) $KRAS^{MUT-iGD}$ in IPMN-related hPanINs without invasive hPDAC (Supplementary Table 12). Altogether, these data suggest a critical role of $KRAS^{MUT-iGD}$ in early PDAC progression.

Looking at organ dissemination, we found a dramatically increased metastatic potential of $Kras^{G12D-iGD}$ cancers (OR 16.7; 95% CI 2.8-98.0; Fig. 2d): primary mPDACs with $Kras^{G12D-iGD}$ were mostly metastasized (20/26), whereas $Kras^{G12D-Het}$ mPDACs were predominantly non-metastatic (2/12). Thus, $Kras^{G12D-iGD}$ drives both early progression and metastasis. This dual role explains (i) early PDAC dissemination in humans and mice²² and (ii) the high incidence of human PDAC metastasis at diagnosis²³. We also mined published data^{8,24} and invariably found $KRAS^{MUT-iGD}$ in human PDAC

metastases. However, because *KRAS*^{MUT-iGD} is present in the primary (early acquisition), its contribution to metastasis could not be recognized by primary/metastasis comparisons⁸.

Alternative oncogenic gains in *Kras*^{MUT-HET} tumors

Among the 12 cancers without *Kras*^{G12D}-dosage gain, two cases had *Myc* amplifications and two had *Yap1* gains (Fig. 2e and Extended Data Fig. 4a-d). *MYC* and *YAP1* are known human oncogenes, amplified in 12% (13/109) and 1% (1/109) of hPDAC, respectively (Fig. 2f). In addition, chr19 gain occurred more frequently in *Kras*^{G12D-HET} (3/12) than *Kras*^{G12D-iGD} tumors (4/26), although this was not significant. A focal amplification on chr19 contained 20 genes (Extended Data Fig. 4e). Cross-species analyses revealed frequent gains of the syntenic region in hPDAC, with two genes in the minimal peak region: *NFKB2* and *PSD*, both amplified in 7% (8/109) of hPDAC (Extended Data Fig. 4f). *NFKB2* (but not *PSD*) is expressed in human pancreas and hPDAC (Extended Data Fig. 4g,h), suggesting *Nfkb2* to be the target proto-oncogene on murine chr19. *NFKB2* mediates non-canonical Nfkb signaling. It has not yet been associated with hPDAC, but promotes cell cycle progression *in vitro*²⁵, and knockout of its interaction partner *RelB* impairs PanIN formation in PK mice²⁶. Thus, upon *Kras* mutation, further amplification of partial aspects of *Kras* downstream signaling (*Myc*, *Yap1* or *Nfkb2*) seem sufficient to drive early PDAC progression, whereas strong metastatic potential is linked to amplification of the full *Kras*^{G12D} signaling program.

Evolutionary licensing of oncogenic dosages

The most frequent deletion in mPDAC affected *Cdkn2a* and/or the adjacent non-coding *Cdkn2a*-regulatory region *Ncruc*: 23 *Cdkn2a*^{ΔHOM}, 4 *Ncruc*^{ΔHOM}, 10 *Cdkn2a*^{ΔHET}, 1 *Cdkn2a*^{WT} (chr4 alteration types shown in Fig. 3a,b, Extended Data Fig. 5a-d and Supplementary Table 9). Notably, the majority of *Cdkn2a*/*Ncruc*^{ΔHOM} cancers were *Kras*^{G12D-iGD} (23/27) and had high *Kras*^{G12D} expression. In contrast, *Cdkn2a*^{ΔHET} or *Cdkn2a*^{WT} cancers were predominantly *Kras*^{G12D-HET} (8/11) with low *Kras*^{G12D} expression (Fig. 3c, Extended Data Fig. 3j and 5e,f). Accordingly, in microdissected human PDAC datasets⁶, *KRAS*^{MUT} variant allele frequencies were higher in *CDKN2A*^{ΔHOM} than in *CDKN2A*^{ΔHET/WT} tumors (Fig. 3d and Extended Data Fig. 5g). Thus, *CDKN2A*^{ΔHOM}-deletion and *KRAS*^{MUT-iGD} are linked, with two possible scenarios: (i) *KRAS*^{MUT-iGD} occurs first, but induces senescence that prevents progression

until *CDKN2A* is lost (as proposed in breast tumorigenesis¹⁸), or (ii) *KRAS*^{MUT-iGD} is only tolerated if preceded by *CDKN2A* deletion.

To resolve the sequence, we determined copy-number changes and copy-number neutral allelic imbalance at *Cdkn2a* and *Kras* in *Cdkn2a*^{ΔHOM};*Kras*^{G12D-iGD} mPDACs (n=13) and associated metastases (n=25). We found identical *Cdkn2a* deletions in all 13 primary/metastasis pairs, but discordant chr6 CNA/LOH phenotypes in 7/13 pairs (Fig. 3e). In 6/13 pairs the sequence of *Cdkn2a*-loss/*Kras*^{G12D-iGD} could not be reconstructed, either because the SNP density was too low (4 cases) or because chr6 profiles in primary/metastasis pairs were identical (2 cases). Thus, in all cases with reconstructable sequence, *Cdkn2a* deletion preceded *Kras*^{G12D-iGD} acquisition. For example, mPDAC-53704 (Extended Data Fig. 6) had two liver metastases with identical *Cdkn2a* deletions, but distinct chr6 SNP patterns: one with *Kras*^{G12D-LOH} at distal chr6 (through mitotic recombination) and another affecting the whole chromosome (likely through missegregation). This confirms clonal chr6 diversification and convergent evolution following *Cdkn2a* loss, and explains the primary's gradual chr6 SNP pattern (Extended Data Fig. 6). Fig. 3f shows another example: mPDAC 5320 and its three metastases had identical *Cdkn2a* deletions, but distinct chr6 patterns: while liver metastasis-1 had *Kras*^{G12D-AG} (combined interpretation of aCGH and SNP data), liver metastasis-3 and the lung-metastasis had distinct *Kras*^{G12D-LOH} events, again showing convergent evolution of *Kras*-allelic imbalance and explaining the primary's composite SNP pattern (Fig. 3f).

These results reveal several evolutionary principles in PDAC. First, *Kras*^{G12D-iGD} is contingent on *Cdkn2a*^{ΔHOM} inactivation. Second, *Myc*, *Yap1* or *Nfkb2* amplifications can occur on a *Cdkn2a*^{ΔHET} context, suggesting context-dependent *Cdkn2a* haploinsufficiency. Of note, only one cancer was *Cdkn2a*^{WT}. Third, evolution of multiple independent *Kras*^{G12D}-gains in *Cdkn2a*^{ΔHOM} cancers demonstrates functional convergence towards *Kras*^{G12D-iGD} acquisition upon homozygous *Cdkn2a* loss.

To provide *in vivo* evidence for functional convergence in *Cdkn2a*^{ΔHOM} contexts, we generated mice with pancreas-specific *Kras*^{G12D}-expression and *Cdkn2a*-deletion (PKC). We found *Kras*^{G12D-iGD} in 100% (16/16) of PKC tumors (Fig. 4 and Supplementary Table 9), confirming that *Kras*^{G12D-iGD} acquisition is the preferred evolutionary route upon homozygous *Cdkn2a* loss.

Another hPDAC hallmark is *TP53* inactivation²⁷. The analysis of cancers from *Kras*^{G12D-Panc};*Trp53*^{ΔPanc} (PKP) mice revealed ubiquitous *Kras*^{G12D-iGD} (16/16) (Fig. 4 and Supplementary Table 9). Thus, *Trp53*-loss (like *Cdkn2a*^{ΔHOM} alteration) predisposes to *Kras*^{G12D-iGD} acquisition (also reflected in hPDAC; Extended Data Fig. 5g). PK/PKC/PKP cross-comparisons revealed higher CNA numbers and a tendency to amplify *Kras*^{G12D} through arm-level gain (trisomy) in PKP, whereas copy-number neutral LOH predominates in PKC (Fig. 4 and Extended Data Fig. 5h). Of note, PKP or PKC tumors did not have complex chr4 rearrangements, confirming that chromothripsis in PK cancers results from natural selection for *Cdkn2a* inactivation (Fig. 4).

To address the role of the TGFβ-pathway, we characterized *Kras*^{G12D-Panc};*Tgfb2*^{ΔPanc} (PKT) mice (Fig. 4 and Supplementary Table 9). Strikingly, all PKT tumors (n=12) had *Cdkn2a* alterations: two cancers were *Cdkn2a*^{ΔHOM}/*Kras*^{G12D-iGD}, ten were *Cdkn2a*^{ΔHET} and predominantly *Kras*^{G12D-HET} (8/10). Overall, the prevalence of *Kras*^{G12D-iGD} is significantly lower in PKT (4/12) than PK mice (26/38) (P=0.04, Fisher's exact test, OR 0.23, 95% CI 0.06-0.92). *Kras*^{G12D-HET} cancers had frequent alternative oncogenic gains (*Nfkb2*/chr19 trisomy in 4/8 PKT mice), similarly to *Kras*^{G12D-HET} cancers in the PK cohort. Thus, contrary to *Trp53*^{ΔHOM} or *Cdkn2a*^{ΔHOM} alterations, which license *Kras*^{G12D-iGD}-acquisition, *Tgfb2* alterations facilitate the alternative route with *Cdkn2a* haploinsufficiency.

Altogether these data show that evolutionary contingencies and convergence shape early tumorigenesis: different tumor suppressor genes/pathways (*Cdkn2a*, *Trp53*, Tgfβ), their alteration types (ΔHOM/ΔHET) or their combinations (e.g. *Cdkn2a*^{ΔHET}/*Tgfb2*^{ΔHET}) direct evolution into different trajectories by licensing distinct types and extents of oncogenic dosage gains.

Integrating genomes, transcriptomes, phenotypes

Unbiased hierarchical clustering of RNA-Seq data from mPDAC cell cultures (PK cohort) revealed two clusters C1/C2, with 3 sub-clusters within C2 (Fig. 5a). Pathway analyses identified “epithelial cell differentiation” as the top C2 GO-term, whereas “mesenchymal cell differentiation” was defining C1 (Fig. 5a,b and Supplementary Table 13,14). Notably, all C1 cell lines show mesenchymal cell morphology, while C2 lines are invariably epithelial (Fig. 5a,c).

Previous studies classified human pancreatic cancer based on transcriptional profiles^{7,28,29}. Unbiased hierarchical clustering with published classifiers shows large overlaps of subtypes proposed by Bailey⁷ and Moffitt²⁹ to the initially proposed three Collisson²⁸ subtypes: classical, exocrine-like, quasimesenchymal (QM). One exception is the lacking exocrine-like signature in the Moffitt classification, which was proposed to be an artefact of acinar cell "contamination" (details in Extended Data Fig. 7a-d). The Collisson classifier²⁸ separates human PDAC cell lines into two subtypes (classical and QM; Extended Data Fig. 7e) and mouse PDAC cell lines into 3 subtypes: classical-equivalent, QM-equivalent (both in epithelial C2) and the mesenchymal M subtype (C1) (Extended Data Fig. 7f). The equivalent of the mouse M subtype with the strong EMT signature has not been described in human cell lines so far, reflecting underrepresentation of mesenchymal phenotypes in human cell line collections (see also Extended Data Fig. 7g,h). As described below, however, mesenchymal mPDACs in C1 represent undifferentiated pancreatic carcinomas with a pronounced EMT signature and human equivalent.

C1 shows strong gene set enrichment for Ras downstream signaling pathways (Fig. 5b and Supplementary Tables 13,14). This cannot be explained by the genetic *Kras* status alone: only C2a is *Kras*^{G12D-HET}, whereas C2b, C2c and C1 are mostly *Kras*^{G12D-IGD}. However, integration of *Kras*^{G12D} expression revealed its gradual increase from C2a to C2b/c and further substantial elevation in C1 (Fig. 5d and Extended Data Fig. 3k). Thus, the mesenchymal phenotype is associated with *Kras*^{G12D} expression above a certain threshold.

To study this association further, we induced clonal PDACs by CRISPR/Cas9 somatic mutagenesis³⁰ in PK mice (Fig. 5e), screened for simultaneous presence of epithelial and mesenchymal cells, and separated/enriched either phenotype by differential trypsinization. Two such cancers were identified. In each case, indel patterns of epithelial/mesenchymal pairs were identical (Extended Data Fig. 8a,b), showing (i) common clonal origin of epithelial/mesenchymal cells and (ii) independence of epithelial/mesenchymal phenotypes from CRISPR/Cas9-induced TSG alterations. Notably however, mPDAC 021 had *Kras*^{G12D-IGD}, elevated *Kras* expression and downstream pathway activation in mesenchymal, but not epithelial cells. In mPDAC 901 both clones were *Kras*^{G12D-HET}, but mesenchymal cells had increased *Kras* expression, supporting a role of *Kras*^{G12D} dosage-variation in shaping cellular phenotypes (Fig. 5e, Extended Data Fig. 8c,d and Supplementary Table 15). Moreover, *KRAS*^{G12D}

overexpression in hPDAC cell lines induced an EMT signature, with Vimentin upregulation and E-cadherin repression (Extended Data Fig. 8e-g and Supplementary Table 16).

PDAC histology revealed a striking association with transcriptome clusters (Fig. 5a,f). Histopathologic grade scores increased from C2a to C2b/c and C1, with C2a being well- or moderately differentiated (G1, G2) and C1 being almost exclusively undifferentiated. Undifferentiated cancers are typically advanced and therefore underrepresented (1-3%) in human surgical series or cell line collections, but autopsy series reported up to 16% hPDACs with at least focal undifferentiated components^{31,32}. Dedifferentiation can occur during disease progression or is triggered by treatment. It is associated with poor prognosis^{32,33}, which is also reflected in mice (Extended Data Fig. 9a). Our results link this aggressive PDAC subtype with highest *Kras*^{G12D} expression levels and Ras-related transcriptional programs (Fig. 5b,d and Supplementary Table 13). We also screened human transcriptome data (ICGC PACA-AU cohort) for undifferentiated pancreatic carcinomas and performed unbiased hierarchical clustering of differentially regulated genes in undifferentiated cancers (Extended Data Fig. 9b). Of note, undifferentiated human pancreatic carcinomas are characterized by reduced expression of genes involved in “epithelial” (cluster-2) or “squamous differentiation” (cluster-1), and a strong upregulation of genes in cluster-4, containing gene sets enriched for EMT and Ras downstream signaling (Extended Data Fig. 9b-d and Supplementary Table 17 and 18).

We exploited the mouse to address complex questions, including cell based resources (overcoming human PDAC’s stroma richness), primary/metastasis resources (phylogenetic tracking, evolution) and *in vivo* modelling (proof-of-concept functional studies). In addition, discoveries were facilitated by the relatively low complexity of mouse PDAC genomes (easier interpretability). Notably, a transposon-induced PDAC model¹³ showed that our findings are equally valid in contexts of excessive mutational loads (Extended Data Fig. 10 and Supplementary Table 19).

Conclusions

Our study proposes a novel comprehensive conceptual framework for molecular PDAC evolution and phenotypic diversification. It describes evolutionary trajectories, identifies their genetic hallmarks and shows how oncogenic dosage-variation is differentially licensed along individual routes by the three major PDAC tumor suppressive pathways to control critical disease characteristics, including early progression, histopathology, metastasis, cellular plasticity and clinical behavior (Fig. 5g). *RAS* gene

mutations affect more than 30% of human cancers, often involving their allelic imbalance. We therefore presume that the principles identified here are significant far beyond PDAC.

Figure 1 | Genetic landscape of mouse PDAC and comparison to the human disease. a, Trinucleotide context-dependent SNV frequencies in mouse (n=38 PK mice) and human PDAC (n=51 patients from⁶) derived from WES. **b,** SNV, indel, CNA and translocation burdens by WES, aCGH and M-FISH in PK mice (n=38) and human PDAC (n=51 patients for SNV, indel, CNA [data from⁶] and n=24 cell lines for translocations). **P=0.002, ***P≤0.001, two-sided Mann-Whitney test; bars, median. **c,** CNAs and ploidy/translocations in PK mice (n=38), detected by aCGH and M-FISH. Mixed ploidy, n≥3 diploid/tetraploid cells in 10 karyotypes. **d,** Rearrangement graph showing chr4 chromothripsis in mPDAC S821, based on WGS. Haplotype-specific chromosome content loss confirmed by M-FISH (n=10/10 karyotypes). **e,** Age at tumor diagnosis of mice having cancers with (n=14) or without (n=23) complex/clustered chromosomal rearrangements (n≥10 CNAs/chromosome). Two-sided log-rank test.

Figure 2 | Mutant *KRAS* gene dosage increase occurs early in PDAC evolution and drives metastasis. **a**, *Kras*^{G12D} gene dosage "states" defined by aCGH, WES and M-FISH (n=38 PK mice). Exemplary CNV-plot for each "state" on the right, y-axis, copy number **b**, Allele-specific *Kras*^{G12D} mRNA expression in *Kras*^{G12D-iGD} (n=26 mice) and *Kras*^{G12D-HET} mPDACs (n=12 mice) by combined amplicon-based RNA-Seq and qRT-PCR. *P=0.02, two-sided Mann-Whitney test; bars, median. **c**, Codon-12 variant allele frequency of microdissected *KRAS*^{G12} mutant hPanIN (n=20) by amplicon-based deep sequencing. H&E stains show histopathologic stages of microdissected hPanINs. Scale bars, 50 μ m. **d**, Macro-/micro-metastasis prevalence in *Kras*^{G12D-HET} (n=12) vs. *Kras*^{G12D-iGD} (n=26) mPDACs. (**P=0.001, two-sided Fisher's exact test). Liver metastasis, H&E. Scale bars, 150 μ m (top) and 50 μ m (bottom); square, zoom-in area. **e**, *Kras*^{G12D-HET} mPDAC amplify alternative oncogenes (*Myc*, *Nfkb2* or *Yap1*) to intensify partial aspects of Ras downstream signaling. Focal, focal amplification; Arm, arm-level amplification. **f**, Amplification of *MYC*, *NFKB2* or *YAP1* in *KRAS*^{MUT} human PDAC. Note, these amplified genes can not only collaborate with *KRAS*^{MUT-Het} but also with *KRAS*^{MUT-iGD}. Data from⁶.

Figure 3 | *Cdkn2a* alteration “states” dictate distinct evolutionary PDAC trajectories. **a**, Chr4 alteration types involving *Cdkn2a* by aCGH/M-FISH (n=38 PK mice). Complex rearrangements, n≥10 CNAs/chromosome. Exemplary CNV plots on the right; y-axis, copy number. **b**, Translocations affecting chr4/*Cdkn2a* in mPDAC-R1035 by M-FISH (10/10 karyotypes). **c**, Prevalence of *Kras*^{G12D-iGD} in mPDAC with homozygously (Δ HOM, n=27) vs. wild-type/heterozygously (Δ HET/WT, n=11) deleted *Cdkn2a/Ncruc*. ***P=0.001, two-sided Fisher’s exact test, OR 15.3, 95% CI 2.8-83.9. **d**, *KRAS* variant allele frequencies in human PDAC with wild-type/heterozygously (n=56) vs. homozygously deleted (n=38) *CDKN2A*. Data from⁶. ***P≤0.001, two-sided Mann-Whitney test; bars, median. **e**, Sequential order of *Cdkn2a* and *Kras*^{G12D} alterations. Chr4 and chr6 CNA/LOH patterns (based on aCGH,WES) of primary mPDACs (n=13 PK mice) and associated metastases (n=25). For seven mPDACs and 16 associated metastases the order of genetic events (dots) could be reconstructed. Bifurcations, divergent evolution of clones; lines, lengths do not represent evolutionary distances; P, primary tumor; Li/Lu/LN, liver/lung/lymph node metastasis. **f**, Detailed chr4/chr6 CNV/LOH profiles for mPDAC5320 primary/metastases. *Cdkn2a* deletions are identical in all lesions (y-axis, copy number). SNP frequency analysis by WES shows distinct chr6 SNP patterns in metastases and a composite picture in the primary, showing convergent evolution of different *Kras*^{G12D-iGD}-gains upon *Cdkn2a* ^{Δ HOM}. Scheme, combined interpretation of WES/aCGH data.

Figure 4 | Defined allelic states and/or combinations of hallmark PDAC tumor-suppressor alterations license oncogenic dosage variation. Types and frequencies of *Kras*^{G12D} gene dosage gains and *Cdkn2a* inactivations, defined by aCGH and amplicon-based *Kras*^{G12D} sequencing in PDAC mouse models expressing pancreas-specific *Kras*^{G12D} alone (PK) or in combination with engineered *Cdkn2a*^{ΔHOM} (PKC), *Trp53*^{ΔHOM} (PKP) or *Tgfbr2*^{ΔHET/HOM} (PKT) inactivation.

Figure 5 | Integrative analyses of PDAC genomics, transcriptomics, cellular phenotypes and histopathologies link molecular, morphologic and clinical disease characteristics. **a**, Unbiased hierarchical clustering of primary mPDAC culture transcriptomes (PK mice). Cell morphology, histopathological grading, *Kras*^{G12D} mRNA expression, genetic *Kras*^{G12D} status and presence/absence of metastasis integrated below. **b**, Selected gene sets from gene-set enrichment analysis of clusters C2 vs. C1. (full list in Supplementary Table 13,14). **c**, mPDAC cultures with mesenchymal/epithelial morphology from clusters C1/C2, respectively. 100x magnification; squares, zoom-in area. **d**, *Kras*^{G12D}-allele-specific mRNA levels in mPDAC transcriptional clusters, combined amplicon-based RNA-Seq and qRT-PCR (C2a/b/c/C1, n=5/7/6/15 mice). P=1.9*10⁻⁶, two-sided Pearson correlation; bars, median. **e**, CRISPR/Cas9-mediated multiplexed somatic inactivation of PDAC-relevant tumor suppressors by electroporation-based transfection to achieve low-frequency mosaicism and clonal tumor outgrowth. Differential trypsinization separates epithelial/mesenchymal cells in mPDACs with mixed morphologies (100x magnification; squares, zoom-in area). CRISPR/Cas9-induced indel signatures are identical in epithelial/mesenchymal pairs (Extended data Fig. 8), indicating common cell of origin. Total *Kras* mRNA levels in epithelial/mesenchymal pairs (qRT-PCR, normalized to *Gapdh*, n=2 technical replicates). Bars, mean; error bars, SEM. **f**, mPDAC histopathological grading in transcriptional clusters (C2a/b/c/C1, n=4/7/6/15, single section per mPDAC). Representative sections (H&E) shown. *Benjamini-Hochberg-adj. P≤0.05, **P=0.005; two-sided Fisher's exact test; scale bars, 150µm. **g**, Simplified model of PDAC evolution reconciling molecular, morphologic and clinical disease characteristics. *KRAS*^{G12D-iGD} gain or alternative oncogenic amplifications (*Myc/Yap1/Nfkb2*) are critical for early disease progression. Different oncogenic gains and dosages evolve along distinct evolutionary routes, licensed by defined allelic states (heterozygous/homozygous) and/or combinations of hallmark tumor-suppressor alterations. For simplicity, only the prototype tumor suppressor gene *CDKN2A* is shown. Not visualized: *TP53*^{ΔHOM} loss, also promoting *KRAS*^{MUT-iGD}, or *TGBFR2*^{ΔHET/HOM} inactivation, supporting evolution through *CDKN2A*^{HET}/*KRAS*^{MUT-HET} trajectories. Depicted trajectories are typical, but not completely exclusive, e.g. *MYC* or *NFKB2* amplifications, which drive *KRAS*^{MUT-HET} cancers, can also cooperate with *KRAS*^{MUT-iGD}. Major aspects of a cancer's biology/phenotype are linked to differential evolution.

Acknowledgements

We thank the comparative experimental pathology team for valuable discussions and Aaron Selmeier, Laura Dajka, Olga Seelbach, Petra Meyer, Tatjana Schmidt, Julia Eichinger and Teresa Stauber for excellent technical assistance as well as Maximilian Reichert for vector constructs. The work was supported by the German Cancer Consortium Joint Funding Program, the Helmholtz Gemeinschaft (PCCC Consortium), the German Research Foundation (SFB1243; A13/A14) and the European Research Council (ERC CoG No. 648521). The authors declare no competing financial interests.

Author Contributions

S.M., D.S., R.R. designed the study; S.M., T.E., R.M., D.S., R.R. interpreted and visualized data; T.E., S.La., M.Zw., M.B. conducted bioinformatic analyses. S.M., T.E., R.M., S.La., M.Zw., I.V. developed bioinformatic analyses strategies; S.M., R.M., M.Zu., T.K., A.S., B.S., J.G., K.Sc., Z.H., A.A., N.S., C.V. L.R. isolated mPDAC cell cultures; S.M., R.M., J.H., K.U. performed genomics with help from R.Ö.; R.Ö., C.Z. conducted RNA-Seq; R.B., S.Lo., B.F., S.K., K. St., F.Y. performed cytogenetics; B.K. performed microdissection; B.K., H-Y.Y., G.K., W.W., K.St. carried out pathological assessment; C.Z., S.P., W.E., K.U., I.V. contributed analytic tools; M.Fri., O.M.D., S.E., F.C-C., J.R., M.I.S., M.Fra., J.M., G.K., R.M.S., J.C., P.L., G.V., W.W., K.St., W.E., G.S., A.B., D.S., R.R. provided resources and critical input; D.S., R.R. supervised the study; S.M., R.R. wrote the manuscript; T.E., R.M. and D.S. edited the manuscript.

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R. R. (roland.rad@tum.de).

References

- 1 Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* **74**, 2913-2921, doi:10.1158/0008-5472.CAN-14-0155 (2014).
- 2 Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806, doi:10.1126/science.1164368 (2008).
- 3 Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109-1113, doi:10.1038/nature09460 (2010).
- 4 Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405, doi:10.1038/nature11547 (2012).

- 5 Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501, doi:10.1038/nature14169 (2015).
- 6 Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun* **6**, 6744, doi:10.1038/ncomms7744 (2015).
- 7 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52, doi:10.1038/nature16965 (2016).
- 8 Makohon-Moore, A. P. *et al.* Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet* **49**, 358-366, doi:10.1038/ng.3764 (2017).
- 9 Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* **15**, 3243-3248, doi:10.1101/gad.943001 (2001).
- 10 Schonhuber, N. *et al.* A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. *Nat Med* **20**, 1340-1347, doi:10.1038/nm.3646 (2014).
- 11 Perez-Mancera, P. A. *et al.* The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature* **486**, 266-270, doi:10.1038/nature11114 (2012).
- 12 Mann, K. M. *et al.* Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A* **109**, 5934-5941, doi:10.1073/pnas.1202490109 (2012).
- 13 Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat Genet* **47**, 47-56, doi:10.1038/ng.3164 (2015).
- 14 Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226-1236, doi:10.1016/j.cell.2013.02.023 (2013).
- 15 Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378-382, doi:10.1038/nature19823 (2016).
- 16 Yamada, H. *et al.* Amplifications of both c-Ki-ras with a point mutation and c-myc in a primary pancreatic cancer and its metastatic tumors in lymph nodes. *Jpn J Cancer Res* **77**, 370-375 (1986).
- 17 Heidenblad, M. *et al.* Detailed genomic mapping and expression analyses of 12p amplifications in pancreatic carcinomas reveal a 3.5-Mb target region for amplification. *Genes Chromosomes Cancer* **34**, 211-223, doi:10.1002/gcc.10063 (2002).
- 18 Sarkisian, C. J. *et al.* Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. *Nat Cell Biol* **9**, 493-505, doi:10.1038/ncb1567 (2007).
- 19 Junttila, M. R. *et al.* Selective activation of p53-mediated tumour suppression in high-grade tumours. *Nature* **468**, 567-571, doi:10.1038/nature09526 (2010).
- 20 Feldser, D. M. *et al.* Stage-specific sensitivity to p53 restoration during lung cancer progression. *Nature* **468**, 572-575, doi:10.1038/nature09535 (2010).
- 21 Rad, R. *et al.* A genetic progression model of Braf(V600E)-induced intestinal tumorigenesis reveals targets for therapeutic intervention. *Cancer Cell* **24**, 15-29, doi:10.1016/j.ccr.2013.05.014 (2013).
- 22 Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349-361, doi:10.1016/j.cell.2011.11.025 (2012).
- 23 Stathis, A. & Moore, M. J. Advanced pancreatic carcinoma: current treatment and future challenges. *Nat Rev Clin Oncol* **7**, 163-172, doi:10.1038/nrclinonc.2009.236 (2010).
- 24 Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114-1117, doi:10.1038/nature09515 (2010).
- 25 Schneider, G. *et al.* IKKalpha controls p52/RelB at the *skp2* gene promoter to regulate G1- to S-phase progression. *EMBO J* **25**, 3801-3812, doi:10.1038/sj.emboj.7601259 (2006).
- 26 Hamidi, T. *et al.* Nuclear protein 1 promotes pancreatic cancer development and protects cells from stress by inhibiting apoptosis. *J Clin Invest* **122**, 2092-2103, doi:10.1172/JCI60144 (2012).
- 27 Redston, M. S. *et al.* p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res* **54**, 3025-3033 (1994).
- 28 Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* **17**, 500-503, doi:10.1038/nm.2344 (2011).
- 29 Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **47**, 1168-1178, doi:10.1038/ng.3398 (2015).

- 30 Maresch, R. *et al.* Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nat Commun* **7**, 10770, doi:10.1038/ncomms10770 (2016).
- 31 Morohoshi, T., Held, G. & Kloppel, G. Exocrine pancreatic tumours and their histological classification. A study based on 167 autopsy and 97 surgical cases. *Histopathology* **7**, 645-661 (1983).
- 32 Iacobuzio-Donahue, C. A. *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *J Clin Oncol* **27**, 1806-1813, doi:10.1200/JCO.2008.17.7188 (2009).
- 33 Winter, J. M. *et al.* Absence of E-cadherin expression distinguishes noncohesive from cohesive pancreatic cancer. *Clin Cancer Res* **14**, 412-418, doi:10.1158/1078-0432.CCR-07-0487 (2008).

METHODS

Primary mPDAC cultures preparation. For 2D primary cell culturing, primary tumor or metastasis were cut into small pieces and digested 1-2h in 200Units/mL collagenase II (Worthington) in DMEM medium (Thermo Fisher Scientific Inc.) containing 10% fetal calf serum (FCS, Merck) and 1x Penicillin/Streptomycin (P/S, Thermo Fisher Scientific Inc.). After short term expansion, primary cells were frozen in 10% dimethyl sulfoxide (Roth) and 50% FCS. For all primary culture experiments, culturing medium (DMEM supplemented with 10% FCS and 1x P/S) and cultures with less than 10 passages were used. Primary cultures were routinely tested for mycoplasma contamination by PCR and authenticated by re-genotyping of cell cultures and corresponding mice.

gDNA and RNA isolation. gDNA from murine primary cell culture pellets was isolated using the DNeasy Blood & Tissue Kit (Qiagen) according to manufacturer's instructions. Total RNA was isolated with the RNeasy kit (Qiagen) from 60-80% confluent primary cell lines cultured in a 10cm dish in culturing medium without P/S and immediately transferred into RLT buffer (Qiagen) containing β -mercaptoethanol.

Histology and micro-metastases screening. For histologic characterization of mPDACs, 2 μ m thick specimens from formalin-fixed paraffin embedded (FFPE) material were routinely H&E stained and submitted to two veterinary pathologists experienced in comparative pancreatic cancer pathology. Histopathologic grading was performed with respect to the most recent consensus report of genetically engineered mouse models³⁴. For histopathologic examination of micro-metastases, three H&E-stained liver sections (separated by 200 μ m) were screened for metastatic lesions by a veterinary pathologist.

Animal experiments. Mice were maintained on *C57Bl/6;129S6/SvEv* mixed background and housed under specific-pathogen-free conditions. Female and male mice were randomly submitted to respective tumor cohorts. For the generation of double- or triple-mutants, pancreas-specific Cre lines^{10,35,36} were intercrossed with *Kras*^{G12D-Panc} (PK mice)^{9,10} only, or in addition with *Cdkn2a* ^{Δ HOM-Panc} (PKC mice)³⁷, *Trp53* ^{Δ HOM-Panc} (PKP mice)^{38,39} or *Tgfbr2* ^{Δ HET-Panc} and *Tgfbr2* ^{Δ HOM-Panc} (PKT mice)⁴⁰ mice. Kaplan-Meier survival curves were generated using Prism (GraphPad Software v5.01). In case the animal presented a palpable abdominal mass above 1.5cm, ascites, signs of sickness or a weight loss of more than 15% of the body weight, mice were euthanized in compliance with the European guidelines for the care and use of laboratory animals. For necropsy of tumor-bearing mice, the

abdominal cavity was macroscopically checked for pancreatic cancer and for metastases at the main metastatic routes (liver, lung, lymph nodes). Animal studies were approved by the Institutional Animal Care and Use Committees (IACUC) of Technische Universität München (Regierung von Oberbayern, Munich, Germany).

Amplicon-based deep sequencing at the *Kras* locus or of *Kras* mRNA. Fifty ng of high-quality genomic DNA or reversely transcribed mRNA (cDNA) were subjected to amplicon-based deep sequencing. Briefly, the *Kras*^{G12D}-mutated locus was amplified using Q5® High-Fidelity DNA Polymerase (New England Biolabs, 40 cycles) and primers with Nextera adapter overhangs (Supplementary Table 20). In a second Q5® PCR step (15 cycles), Nextera index primers (Illumina) were added. After each PCR step, solid phase reversible immobilization cleanup (0.8x) was performed using the Agencourt® AMPure® XP kit (Beckman Coulter GmbH). The pooled library was quantified by SYBR® Green qPCR (Thermo Fisher Scientific Inc.) and the Kapa Biosystems library quantification kit. In total, 8pM of denatured library (20% spiked PhiX DNA) was sequenced in 300bp paired end mode using a MiSeq system (Illumina). Raw reads were mapped to *Kras* reference sequence (Ensemble release GRCm38p4, Genome Reference Consortium). Variant allele frequencies on chr6 at position 145246771 were calculated.

Microdissection of hPanIN and *KRAS*^{G12} status analysis. Nineteen patients (Supplementary Table 11) with or without a history of pancreatic cancer were included into hPanIN lesion analysis, according to approval by the Ethics Committee of the Faculty of Medicine of the Technische Universität München. Patients were classified using WHO recommendations and the TNM staging system. Serially cut 10µm thick specimens from FFPE material were air-dried overnight. Paraffin was removed through short incubation with xylene. Specimens were briefly stained with hematoxylin and kept wet for the micro-dissection procedure. Individually diagnosed samples were microdissected under the Axio Imager microscope (Zeiss) using 20 gauge cannula. Pre- and post-sampling microscopic pictures were taken to (i) document dissection performance and (ii) re-identify each specimen on the corresponding H&E-stained slide. gDNA was extracted as described above using MinElute spin columns (Qiagen) for higher sample concentration. Five µl of eluted hPanIN gDNA were submitted to amplicon-based deep sequencing of *KRAS* exon-2 for detection of *KRAS*^{G12} hotspot mutations. Briefly, 2 pairs of custom *KRAS* primers (Supplementary Table 20) were used for nested PCR amplification of the corresponding *KRAS* region. Illumina Nextera primer pairs were used to add sequencing adapters

and indices. PCR steps, library quantification and sequencing were performed as described above. Raw reads were mapped to *KRAS* reference sequence (GRCh38.p10). Variant allele frequencies were calculated for *KRAS*^{G12} hotspot mutations (positions 25398284 and 25398285 on chr12).

Whole genome sequencing (WGS). One µg of high-quality gDNA extracted from primary tumor cell line and corresponding tail were sheared on a Covaris M220 focused ultrasonicator (Covaris, Inc.) to an approximate fragment size of 500bp. Library was prepared from 500ng of fragmented gDNA using the NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs, Inc.) in combination with the adapter/primer sequences and PCR conditions published before⁴¹. The final library was quantified by qPCR using the Kapa Biosystems library quantification kit. Equimolar amounts of indexed libraries were denatured and diluted to a final concentration of 1.8pM and sequenced in 300bp paired end mode on a NextSeq 500 system (Illumina) to ~20x coverage. Sequencing raw data were converted to fastq format using the bcl2fastq software v2.18.0.12 (Illumina). Raw reads were trimmed with Trimmomatic v0.36 to preserve an average base Phred quality of 25, mapped using the BWA-MEM algorithm v0.7.12 with alternative contig handling and mapped to GRCm38.p5 reference genome.

Inference of chromothripsis. For estimation of copy number states, the Bioconductor HMMcopy package 1.16.0 was used followed by segmentation with the Bioconductor DNACopy package 1.48.0. For LOH analysis variant positions in control and tumor were computed with samtools mpileup v1.3.1. Only positions in regions with mapping quality of 60 and an average phredscore of 20 were considered for further analysis. Furthermore, positions harboring strand bias and variant allele frequencies less than 20% and above 85% in the control were excluded as they are likely homozygous in the germline. The minimal cutoff coverage for a given polymorphic position in the control was set to eight reads. Segmental duplications (UCSC Genome Browser) and regions with mouse line specific variation (Mouse Genomes Project, REL-1505) were excluded. For this set of somatic nucleotide polymorphisms (SNPs) the difference of frequencies between tumor and control samples were calculated. DELLY v0.7.6 was used for calling structural variations (SVs). SV-classes were defined according to DELLY callings: Deletion-type (3to5), duplication-type (5to3) and inversion-type (5to5 and 3to3). The predicted rearrangements were merged and filtered based on variant frequency, mapping quality and the distance between two connected breakpoints. The existence of chromothripsis was tested by applying the six hallmark criteria proposed by Korbelt *et al.*¹⁴. Clustering of SV breakpoints was tested using a χ^2 -goodness-of-fit test. Regularity of oscillating copy number

states in the chromothriptic model was compared to a virtual chromosome generated by a Monte Carlo simulation, as described in⁴². For each distinct number of breakpoints, 100 simulation runs were completed and mean values as well as 95% CI were calculated. Interspersed loss and retention of heterozygosity was analyzed by calculating the Jaccard index between heterozygously deleted segments and regions comprising LOH and SNP information. Randomness of observed DNA segment order was tested using a Monte Carlo simulation as described in¹⁴. The uniform distribution of SV-types was tested using a χ^2 -goodness-of-fit test. The Wald-Wolfowitz runs test as implemented in R package *randtests* 1.0 was performed for testing right-sided against the null hypothesis of randomly distributed 5'-to-3' breakpoint joints sequence.

FISH analyses. For the analysis of copy number status or large structural alterations of human and murine primary cell lines, multicolor fluorescence in situ hybridization (M-FISH) was performed as described before⁴³. For *KRAS* gene detection in hPanIN specimens, the *ZytoLight*® SPEC KRAS/CEN12 Dual Color Probe kit was used according to manufacturer's instructions (ZytoVision GmbH). In brief, 2 μ m FFPE specimens were deparaffinized, pre-incubated in CC 2 buffer (at 95°C for 24min), treated with pepsin solution (at 37°C for 8min) and denatured by a heat treatment step at 80°C for 8min on an automated Discovery XT system (Ventana Medical Systems, Inc.). KRAS/CEN12 dual color probe hybridization was performed by co-denaturing at 75°C for 10min and by incubating at 37°C overnight in a ThermoBrite system (Abbott Laboratories). Slides were washed, nuclei stained with 4',6-diamidino-2-phenylindole (DAPI), covered in an antifade mounting medium and by a coverslip and stored for confocal laser scanning microscopy (LSM) analyses at 4°C in the dark. *ZytoLight*® SPEC KRAS/CEN12 Dual Color Probes and DAPI nuclear stain (excitation/emission: DAPI 405nm/415-490nm; ZyGreen, 503nm/510-540nm; ZyOrange, 547nm/560-650nm) were detected by confocal LSM using a Leica TCS SP8; DMI8 CS microscope equipped with a 63x/1.4 oil immersion objective (Leica). Images (z-stacks, covering the whole nucleus) with a magnification factor of 3 and a frame size of 2048 x 2048 pixels were collected. Generated images were processed using the Huygens Essential software (Scientific Volume Imaging) for deconvolution, then merged and maximum projections were converted with Leica LAS X software.

aCGH analysis. Agilent oligonucleotide aCGH (SurePrint G3 Mouse CGH 240K or custom 60K microarray) was performed according to manufactures instructions. Agilent Genomic Workbench software v7.0.4.0 was used for aCGH data preprocessing. Legacy centralization option was used for

re-centralization of raw log ratios to the most common ploidy state. ADM-2 algorithm was applied for aberration calling. Abberations on chromosome 6 between positions 148719747 and 149503634 were excluded in further downstream analysis as this region resulted likely from an artefact. Normalized and curated data was imported into R.

Whole-exome sequencing (WES) analysis. Coding exons were enriched via whole-exome pull down using Agilent SureSelect Mouse Exon Kit according to manufacturer's instructions and sequenced on the Illumina HiSeq2000 system. Prior to mapping, raw sequencing reads were trimmed using Trimmomatic v0.33. Leading and trailing bases with phred scores below 25 and reads shorter 50 nucleotides were removed. In addition the average base quality within a sliding window of 10 nucleotides should be above 25 to keep the read for further downstream analysis. Reads were aligned to the GRCm38.p3 reference genome using BWA-MEM 0.7.12 with default settings. PCR duplicates were marked with Picard tools v1.130 and realignment around indels was performed with GATK toolkit v 3.4.46. Mutect v 1.1.7 was used for calling somatic mutations with default settings. Potential somatic events were filtered for SNPs by excluding SNVs which were listed in in release 1505 of the Mouse Genome Project SNP database⁴⁴. Somatic point mutations were included in the final list, if the read coverage for each position was ≥ 10 in both control and tumor, variant frequency was $\geq 10\%$ and read count supporting the variant nucleotide is ≥ 3 in the tumor sample and $= 0$ in the control. Further, SNVs marked as strand or PCR bias artifacts by "DKFZBiasFilter" (<https://github.com/eislslabs/DKFZBiasFilter>, using default settings) or with a FOXOG-Score of 1 were excluded. Annotation of somatic events was conducted with SNPeff v4.1. SNVs causing variation in splice sites or upstream/downstream of genes were excluded from further analysis. Indels were detected with Pindel⁴⁵. For each potential indel the read coverage was re-calculated using bedtools v2.17.0. Criteria for further downstream processing were: Variant frequency $\geq 10\%$ in tumor and $= 0\%$ in control; and total coverage at the altered position in both control and tumor ≥ 20 . LOH analysis was conducted as described in chapter Inference of Chromothripsis.

WES data analysis from hPDAC. Mapped BAM files from Witkiewicz *et al.*⁶ were downloaded from the Sequence Read Archive (accession number PRJNA278883) approved by the Ethics Committee of the Faculty of Medicine of the Technische Universität München. Further downstream analysis was performed as described above. SNPs were filtered by excluding variants with an alternate allele frequency $\geq 1\%$ in the 1000 Genomes Project, as listed in dbSNP build 146. All available VCF files

from the TCGA-PAAD cohort generated by Mutect2 were downloaded from the NIH Genomic Data Commons data. Downstream processing was performed as described above (PCR and strand bias marking by DKFZBiasFilter was not possible using VCF files). SNPs were filtered by excluding variants with an alternate allele frequency $\geq 1\%$ in the 1000 Genomes Project, as listed in dbSNP build 142. MAF files from other human pancreatic cancer cohorts were downloaded and included in our analysis: All samples, for which whole-exome sequencing data was available, as provided by Bailey *et al.*⁷; pancreatic cancer cell lines from the Cancer Cell Line Encyclopedia⁴⁶ and SNV data from pancreatic cancers as analyzed by Alexandrov *et al.*⁴⁷. In these cohorts, SNPs were filtered by excluding variants with an alternate allele frequency $\geq 1\%$ in the 1000 Genomes Project, as listed in dbSNP Build 146. Remaining SNVs were annotated and filtered with SNPeff as described above.

Analysis of mutational signatures. Mutation spectra for each cohort were compared to a list of 21 signatures previously described by Alexandrov *et al.*⁴⁷; Signature 1B was excluded from further analyses because of presumed biological similarity to Signature 1A. The contribution of each individual signature to the mutation spectrum of each cohort was analyzed by using “deconstructSigs” v1.8.0.

qRT-PCR analysis. Reverse transcription was performed with random hexamers using 1 μ g of total RNA following instructions of the SuperScript II protocol (Thermo Fisher Scientific Inc.). Real-time qPCR was performed either with the TaqMan qPCR chemistry (Thermo Fisher Scientific Inc.) for mouse using *Kras*-specific primers and probes or with the SYBR® Green master mix (Thermo Fisher Scientific Inc.) using primers for human target genes *VIM*, *CDH1* and *MMP1* (Supplementary Table 20). *Gapdh* or *GAPDH* in combination with *PPIA* were used as housekeeping genes for normalization (Supplementary Table 20). qPCR was conducted on a StepOnePlus system (Applied Biosystems). For analyses of mutant *Kras*^{G12D} mRNA levels in mPDACs, first total (wild-type plus mutant) *Kras* mRNA levels were determined using qRT-PCR. Second, the identical cDNA was used for amplicon-based deep sequencing to detect the proportion of mutant to wild-type *Kras* mRNA. Third, the mutant to wild-type *Kras* mRNA ratio was multiplied with the total *Kras* mRNA level to calculate the mutant *Kras*^{G12D}-specific mRNA level.

RNA-Seq analysis. Bulk 3'-prime transcript end RNA-Seq (SCRBS-Seq) libraries were prepared as described previously⁴⁸. Briefly, RNA was reversely transcribed using oligo-dT primers decorated with

sample barcodes, unique molecular identifiers (UMIs) and adapters (Integrated DNA Technologies). cDNA from all samples was pooled and un-incorporated primers digested using ExonucleaseI (New England Biolabs). Next, the cDNA pool was amplified with KAPA HiFi ReadyMix (KAPA Biosystems). To obtain sequencing libraries, 0.8ng of cDNA was tagged and 3' ends amplified with the Nextera XT Kit (Illumina) using a specific primer for the adapter on the 3'-end. The library was paired-end sequenced on a HiSeq1500 with 16 cycles for read 1 to decode sample barcodes and UMIs and 51 cycles on read 2 into the cDNA fragment. For the preparation of the human pancreatic cancer cell line samples, the flow cell binding sites P5 and P7 were exchanged to allow sequencing of the cDNA in read1 and barcodes and UMIs in read2. Data was processed using the published Drop-seq pipeline (v1.0)⁴⁹ to generate sample- and gene-wise UMI tables. Reference genome (GRCm38) was used for alignment. Transcript and gene definitions were used according to the ENSEMBL annotation release 75. Further analyses were performed with R version 3.2.2. Initial hierarchical clustering (method: Complete linkage, distance measure: Euclidian) of samples was performed for the top 10% variable genes. Bootstrapping was performed to assess cluster stability with the pvclust package v2.0. The 4 most prominent clusters were selected and differential expression between these clusters was calculated with DEseq2⁵⁰. A gene was considered to be differentially regulated if the absolute log2-foldchange was above 0.8 and the adjusted P-value was ≤ 0.05 . Gene set enrichment testing was performed with DAVID 6.8⁵¹ or the hypergeometric test as implemented on the "Molecular Signature Database" (MSigDB) v6.0 homepage (<http://software.broadinstitute.org/gsea/msigdb/annotate>). For all MSigDB analyses, top100 enriched terms with a false discovery rate (FDR) of $P \leq 10^{-4}$ were included. Published PDAC classifier genes²⁸ and the hallmark EMT gene set (downloaded from MSigDB v5.2⁵²) were used for sample clusterings (method: Ward, distance measure: Euclidian).

hPDAC subtyping. Normalized RNAseq data was derived from Bailey *et al.*⁷. Samples that were histologically classified as "PDA-Adenosquamous carcinoma" and "Pancreatic Ductal Adenocarcinoma" were used for hierarchical clustering (method: Ward, distance measure: Euclidian) with classifier gene lists published elsewhere^{28,29}.

Microarray data analysis. Affymetrix-based CCLE raw data set was downloaded from (Broad-Novartis Cancer Cell Line Encyclopedia, Version 2.17). Hematopoietic or lymphoid neoplasms were excluded since (1) the primary interest of our study were solid tumors and (2) the overall gene expression signature of these samples was shown to be very distinct from all other samples in the study⁴⁶. Normalization of the data was performed⁴⁶ with RMA. In general, if genes were represented by

two or more probe sets, the probe set with the highest mean expression was used for all further microarray data analyses. Mapping between probeset and genes were conducted with the appropriate Bioconductor packages. Target genes for the *TP63ΔN* network were downloaded from the “Pathway Interaction Database” (PID)⁵³ and hierarchically clustered (method: Ward, distance measure: Euclidian). Gene set enrichment analysis was conducted with DAVID or MSigDB v6.0. All following microarray data sets are Illumina-based and were VST-transformed followed by quantile normalization as implemented in lumi⁵⁴. Microarray data set of hPDAC cell lines (accession number GSE17891) was downloaded from Gene Omnibus Expression database. PDAC classifier genes and EMT hallmark gene set were used as described above. For the comparison of human wild-type pancreatic tissue and hPDAC cell lines, limma⁵⁵ was used for detection of differential expression between groups. Differentially expressed genes were determined with an alpha level threshold of 5%. The PACA-AU ICGC data set was downloaded from the repository <https://dcc.icgc.org/repositories>. Samples that met the following criteria were selected for further analyses: (i) Bailey HistoSubtypes “PDA-Adenosquamous carcinoma” or “Pancreatic Ductal Adenocarcinoma” with available subtype information from Bailey *et al.*⁷ and (ii) ICGC WHO Grading “Undifferentiated carcinoma”. Only representative samples, as judged by cluster analysis, from this group were selected for downstream analysis. ANOVA was performed across six defined subgroups of pancreatic cancer: (i) undifferentiated pancreatic carcinoma, (ii) adenosquamous pancreatic carcinoma and (iii-vi) PDAC sub-stratified in pancreatic progenitor, immunogenic, squamous and ADEX subtypes. Genes with an adjusted P-value ≤ 0.05 were hierarchically clustered (method: Ward, distance measure: Manhattan) and the resulting cluster tree was computationally stratified into five sub-clusters. Genes within subclusters were used for gene enrichment analysis as described above. Seventeen PK-PB primary cultures established elsewhere¹³ were submitted to RNA extraction and subsequent gene expression profiling analysis on a MouseWG-6 v2.0 Expression BeadChip (Illumina). The 5% of genes with the highest variability across all samples were used for hierarchical clustering using the ward method for aggregation of samples. Limma was used as described above. A gene was called differentially expressed if the adjusted P-value was ≤ 0.05 and the log₂-fold was at least 0.8.

Quantitative transposon insertion site sequencing (QiSeq). Aforementioned gDNA samples of the PK-PB pancreatic cancer cell cultures¹³ were sequenced for transposon integration sites and bioinformatics analyses were performed as described elsewhere⁵⁶. Transposon integration sites that are supported by at least 20 reads and reside in intragenic regions were counted for the computation

of the mutational burden. For the assessment of the *Cdkn2a/Ncruc*^{Δ^{HO}} status caused by transposon insertional mutagenesis, only the top hit of each tumor was considered.

***Kras*^{G12D} induction after lentiviral transduction of hPDAC cell lines.** The pINDUCER20⁵⁷ vector system comprising a puromycin resistance gene was used for doxycycline-inducible *KRAS*^{G12D} overexpression. Briefly, cDNA of oncogenic *KRAS*^{G12D} (CCDS 8702.1, 35G>A) and *GFP* were cloned into the pINDUCER20 lentiviral vector. Stbl3™ (Thermo Fisher Scientific Inc.) were chemically transformed and pDNA sequence was verified. For lentivirus production, HEK293FT cells were transfected using *TransIT*®-LT1 (Mirus Bio LLC) with standard virus packaging plasmids and respective pINDUCER20 vectors by following manufacturer's recommendations. Virus-containing supernatant was pooled 48h and 72h post transfection, concentrated by polyethylene glycol 6000 precipitation⁵⁸ and stored at -80°C after shock-freezing. 1x10⁵ HUPT3 (COSMIC ID: COSS907285) and PANC0327 (COSMIC ID: COSS925346) hPDAC cells were transduced in presence of 1μg/mL polybrene and selected with puromycin antibiotic. Target gene expression was induced for stated time points by the addition of 100ng/μL doxycycline into P/S-free culturing medium. RNA isolation, qRT-PCR and SCRB-Seq were performed as described above. For differential gene expression analysis, raw sequencing data were mapped to the human reference genome (GRCh 38p10). Transcript and gene definitions were used according to the ENSEMBL annotation release 87. Group comparisons (*KRAS*^{G12D} vs *GFP*) were conducted with DESeq2.

Somatic CRISPR/Cas9 gene-editing for tumor clone tracking in mice. Multiplexed gene editing of tumor suppressor genes using CRISPR/Cas9 in the pancreas of PK mice was performed as described elsewhere³⁰. Primary cultures of induced mPDACs were isolated as described above and monitored for the simultaneous presence of epithelial and mesenchymal phenotypes. Enrichment of epithelial and mesenchymal cell morphologies was achieved by differential exposition times to trypsin (Thermo Fisher Scientific Inc.). Short-term incubation (2-3 min) at room temperature induced detachment of mesenchymal cells, while epithelial colonies remained adherent. Both cell fractions were subsequently grown to 80% confluency in new flasks. This process was repeated for 3-6 times until homogenous epithelial and mesenchymal cell fractions were enriched. Clonal origin of both phenotypes was confirmed by targeted amplicon-based next generation sequencing of CRISPR/Cas9-edited loci as described earlier^{30,59}. Analyses of the *Kras* allelic status and mRNA expression were carried out as described above.

Statistics and reproducibility. For each experiment, all statistics were performed as indicated in respective Figure legends and Extended Data Figure legends. Statistical testing across all classes was performed to account for multiple testing. Continuous variables were tested for normal distribution. Non-parametric tests were used for non-normally distributed data. Complex statistical techniques are explained in detail in the Methods section. No animals were excluded from any of the cohorts. The veterinarian pathologists were blinded during histological grading of primary tumors and metastasis screening. The study was of explorative nature. Due to this study design prior knowledge of the expected effect-size was not available and no power calculations were conducted.

Code availability. Source code is available from the authors upon reasonable request.

Data Availability. Sequence data have been deposited at EBI European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>) under accession numbers PRJEB23116, PRJEB23110, PRJEB23183, PRJEB23378 and PRJEB23378. Microarray data have been deposited in the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number XXX. All data are available from the corresponding author upon reasonable request.

- 34 Hruban, R. H. *et al.* Pathology of genetically engineered mouse models of pancreatic exocrine cancer: consensus report and recommendations. *Cancer Res* **66**, 95-106, doi:10.1158/0008-5472.CAN-05-2168 (2006).
- 35 Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437-450 (2003).
- 36 Nakhai, H. *et al.* Ptf1a is essential for the differentiation of GABAergic and glycinergic amacrine cells and horizontal cells in the mouse retina. *Development* **134**, 1151-1160, doi:10.1242/dev.02781 (2007).
- 37 Aguirre, A. J. *et al.* Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev* **17**, 3112-3126, doi:10.1101/gad.1158703 (2003).
- 38 Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet* **29**, 418-425, doi:10.1038/ng747 (2001).
- 39 Lee, C. L. *et al.* Generation of primary tumors with Flp recombinase in FRT-flanked p53 mice. *Dis Model Mech* **5**, 397-402, doi:10.1242/dmm.009084 (2012).
- 40 Chytil, A., Magnuson, M. A., Wright, C. V. & Moses, H. L. Conditional inactivation of the TGF-beta type II receptor using Cre:Lox. *Genesis* **32**, 73-75 (2002).
- 41 Bronner, I. F., Quail, M. A., Turner, D. J. & Swerdlow, H. Improved Protocols for Illumina Sequencing. *Curr Protoc Hum Genet* **80**, 18 12 11-42, doi:10.1002/0471142905.hg1802s80 (2014).
- 42 Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27-40, doi:10.1016/j.cell.2010.11.055 (2011).
- 43 Jentsch, I., Adler, I. D., Carter, N. P. & Speicher, M. R. Karyotyping mouse chromosomes by multiplex-FISH (M-FISH). *Chromosome Res* **9**, 211-214 (2001).
- 44 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294, doi:10.1038/nature10413 (2011).
- 45 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).

- 46 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).
- 47 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
- 48 Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. The impact of amplification on differential expression analyses by RNA-seq. *Sci Rep* **6**, 25533, doi:10.1038/srep25533 (2016).
- 49 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).
- 50 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 51 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
- 52 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 53 Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674-679, doi:10.1093/nar/gkn653 (2009).
- 54 Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548, doi:10.1093/bioinformatics/btn224 (2008).
- 55 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 56 Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat Protoc* **12**, 289-309, doi:10.1038/nprot.2016.164 (2017).
- 57 Meerbrey, K. L. *et al.* The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc Natl Acad Sci U S A* **108**, 3665-3670, doi:10.1073/pnas.1019736108 (2011).
- 58 Kutner, R. H., Zhang, X. Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc* **4**, 495-505, doi:10.1038/nprot.2009.22 (2009).
- 59 Weber, J. *et al.* CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc Natl Acad Sci U S A* **112**, 13982-13987, doi:10.1073/pnas.1512392112 (2015).
- 60 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).
- 61 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).
- 62 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 63 Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929-944, doi:10.1016/j.cell.2014.06.049 (2014).

Extended Data Figure 1 | Mutational patterns, karyotype complexity and structural alterations in primary PDAC. **a**, Frequency of somatic base substitutions based on trinucleotide context in mouse (n=38 PK mice) and human PDAC (n=51 patients, data used for analysis from Witkiewicz *et al.*⁶). **b**, Frequency of somatic base substitutions based on trinucleotide context in mouse (n=38 PK mice) and human PDAC (n=51 patients, data used for analysis from ⁶). **b-f**, Mutation spectra defined by trinucleotide contexts around base substitutions as detected by whole-exome sequencing show similar patterns in PK mice (n=38) and in relevant human pancreatic cancer cohorts. Base substitutions were extracted from BAM, VCF or MAF files from: **b**, Witkiewicz *et al.*⁶, **c**, Bailey *et al.*⁷, **d**, TCGA-PAAD, **e**, Barretina *et al.*⁴⁶ and **f**, Alexandrov *et al.*⁴⁷. Additional information regarding the analysis of each cohort is provided in Supplementary Table 2. **g**, Mutational signatures in mouse and human pancreatic cancer cohorts. Information on mutational signatures was used from Alexandrov *et al.*⁴⁷, who identified 21 mutational signatures operative in human cancer. The „deconstructSigs“ tool was used to determine the composition of the given set of 21 mutational signatures in each pancreatic cancer cohort. Extraction of mutational signatures strongly depends on SNV load per tumor. Due to the low mutational burden of mPDACs from PK mice (median of 18 SNVs per tumor as detected by WES), the analyses of mutational signatures could not be performed at the level of individual tumors. We have therefore investigated the contribution of each of the 21 mutational signatures to the SNV spectrum at the cohort-level (see Methods). Signature 1, reflecting age-associated C>T transversions at NCG trinucleotides, was the only signature consistently identifiable in all cohorts of human and mouse pancreatic cancer. In comparison to human cohorts, PK mice show C>G substitutions at GCC trinucleotides that cannot be attributed to one of 21 mutational signatures. Note that mutations at the GCC motif are not a general phenomenon of PDAC from PK mice, since only 4 samples are predominantly contributing to this peak. **h-i**, Representative M-FISH karyotypes with no or few karyotypic changes are shown for a diploid (40 chromosomes) and tetraploid mouse PDAC (81 chromosomes). Tumor 9591 shows gain of chr14. **j**, Representative karyotype of a complex diploid mPDAC genome with aneuploidy and translocations (46 chromosomes). Both copies of chr4 are involved in translocations: der(4)t(4;10) and der(4)t(4;16); likely affecting *Cdkn2a*. Further structural alterations and copy number changes are: +5, der(5)t(4;5)*2, +6, +7, +8, del(9), +14, del(14), der(16)t(5;16), +17. **k**, Representative example of a complex tetraploid mPDAC karyotype (77 chromosomes). Structural alterations are: der(1)t(1;11), dic(9;9), der(11)t(1;11), and der(14)t(14;19). Single chromosomal copy number changes are: +2, -3, -9, -10, -11, -13, -14, +15 and +19. Del,

deletion; der, derivative chromosome; dic, dicentric chromosome; t, translocation; „-“, chromosome loss; „+“, chromosome gain. **I, (Extension to Fig. 1c.)** Circos plot shows CNAs assessed by aCGH as well as translocations and ploidy states detected by M-FISH in 38 primary PDACs derived from PK mice (n=38). CNAs for each mPDAC are displayed as log₂ difference from tail control. Frequencies of translocations per chromosome are indicated in green in the inner circle of the graph. Connecting lines indicate individual translocations and involved chromosomes. On chr4, genomic alterations frequently involve *Cdkn2a* or *Ncruc*, a Non-coding regulatory region upstream of *Cdkn2a* (27/38 cancers with homozygous and 10/38 with heterozygous inactivation of *Cdkn2a* and/or *Ncruc*). Only one cancer remained *Cdkn2a*^{WT}. The target of copy number changes on chr6 is *Kras*^{G12D}, either through arm level gain or focal amplification. In addition, primary mPDAC of PK mice exhibited recurrent genetic amplifications affecting other known oncogenes, such as *Myc* or *Yap1*, or *Nfkb2*, a novel oncogenic PDAC driver identified in this study (see also Fig. 2e,f and Extended Data Figure 4).

Extended Data Figure 2 | Characterization of complex rearrangements in PDAC from PK mice and statistical inference of chromothripsis based on whole genome sequencing (WGS). **a-n,** Copy number profiles of chromosomes with complex rearrangements (defined as 10 or more CNA segments per chromosome) from primary mPDAC cell cultures as detected by aCGH. A total of 14 mPDACs had chromosomes with such complex rearrangements. **a-i,** Nine primary mPDACs show copy number patterns characterized by heterozygous deletions and oscillation of copy number around few states, indicating chromothripsis as the underlying mechanism. **g,** mPDAC-S821 was also subjected to whole genome sequencing, which proved chromothripsis on chr4 by applying criteria for the inference of chromothripsis¹⁴ (see Fig. 1d and Extended Data Figure 2p-w). **j-m,** Four primary mPDACs feature complex rearrangements associated with multiple copy number states on chr4, likely acquired through progressive/sequential rearrangement cycles. **n,** Cancer 5671 carries a complex rearrangement on chr15 characterized by oscillating copy number states and 3 prominent focal amplifications, of which one contained the *Myc* oncogene. *Myc* amplification is most likely the result of double minute chromosome formation during chromothriptic rearrangement of chr15. **o,** Comparison of age at tumor diagnosis in *Cdkn2a*^{Δ^{HOM}}-deleted cancers with (n=10) or without (n=15) complex clustered chromosomal rearrangements (defined as 10 or more CNAs on individual chromosomes). Complex clustered rearrangements are associated with significantly shortened time to tumor diagnosis, indicating accelerated tumor evolution through genetic crisis. Two-sided log-rank test. **p,**

Criteria proposed by Korbel *et al.*¹⁴ were tested for the inference of chromothripsis. Circos plot displays SNP ratio (inner circle, red dashed line indicating heterozygosity), CNV (outer circle, blue area indicates deletion, red amplification) and structural variations (SVs, colors as in g) as detected by WGS. Chr4 shows a complex deletion pattern and massive rearrangements associated with loss of one copy of *Cdkn2a*. The second copy of *Cdkn2a* is focally deleted. In addition, a balanced translocation of a ~200Kb segment from trisomic chr6 to chr4 and a far smaller segment of chr4 into chr6 was detected. The *Kras* locus is not directly affected by this inter-chromosomal translocation. LOH, CNAs and rearrangements are not detected on other chromosomes. **q**, In a chromothriptic model, DNA breakpoints tend to cluster on a chromosome. Testing against an exponential distribution (parameter λ derived from mean of observed distance between adjacent breakpoints), revealed significantly shorter distances than expected in a progressive model (n=146 breakpoints). $P < 10^{-12}$; χ^2 -goodness-of-fit test. **r**, In a progressive model of acquisition of massive rearrangements, copy number states tend to be more complex than in the chromothriptic. Monte Carlo simulations were used to generate a progressive evolution model with sequential accumulation of observed rearrangements (n=100 simulations per number of SVs). mPDAC S821 showed fewer copy number states on chr4 than expected in the progressive model. Mean is indicated as a black point and lines represent the 95% CI. **s**, Chromothriptic tumors typically feature interspersed loss and retention of heterozygosity. Accordingly, there was a high overlap between deleted regions and LOH segments on chr4 (Jaccard index (J) = 0.99). **t**, In a chromothriptic model, DNA shattering typically occurs on a single haplotype. M-FISH showed that significant loss of chromosomal content occurred on only one copy of chr4. **u**, To show random chromothriptic DNA shattering and re-joining, observed segments (n=73) were re-ordered by running Monte Carlo simulations (n=10³) that generate a background probability distribution. S821 segment order lies within the chromothriptic null model. Two-sided P=0.78. **v**, All 4 SV-types are uniformly distributed in a chromothriptic tumor model. P=0.43; χ^2 -goodness-of-fit test. **w**, In a chromothriptic model, paired end connection types (as given by the SV-type) induce an alternating sequence of DNA segment ends when ordered according to the genomic position on the original chromosome. Tendency towards this alternating 3'-to-5' pattern of rearranged DNA segment ends (n=146) was tested by using right-sided Wald-Wolfowitz runs test. $P < 10^{-12}$. **x**, Mutation clusters in relation to breakpoint junctions involved in chromothripsis are shown as rainfall plot for primary PDAC from PK mouse S821. Each dot represents a single somatic nucleotide variation (SNV) and is ordered on the x-axis according to its position in the mouse genome. The distance of each SNV to the

previous SNV in the genome is shown on the y-axis. The coloring of individual SNV dots indicates the type of nucleotide substitution. **y**, Chr4 “zoom-in” from (x). Breakpoint junctions are shown according to their genomic position on chr4. No mutation clusters - neither in absence nor in combination with breakpoint junctions - were detected, consistent with chromothripsis involving end joining DNA repair mechanisms. This is in contrast to other complex rearrangement types, such as chromoanasythesis, which arise through replication-based mechanisms with breakpoint-associated high mutation rates (e.g. kataegis).

Extended Data Figure 3 | Specificity, timing, mechanisms and impact of *Kras*^{G12D} gene dosage alterations on gene expression in pancreatic tumorigenesis. **a**, Overlay of copy number profiles of primary mPDAC cell cultures from PK mice (n=38) as determined by aCGH. The frequency of a certain region to be amplified (up) or deleted (down) within the cohort is shown on the y-axis (percent of tumors showing alterations). The genomic loci of *Cdkn2a* and *Kras* are the most frequently deleted or amplified regions, respectively. **b**, LOH frequency across the cohort. LOH was analyzed using SNP information extracted from whole exome sequencing (WES) data from primary mPDAC cell cultures of PK mice (n=38). A chromosome was considered to be affected by LOH if the SNP frequency was shifted to ≤ 0.1 or ≥ 0.9 in a segment with a size of ≥ 200 kb. Chr4 (home of *Cdkn2a*) and chr6 (home of *Kras*) show markedly increased rates of LOH as compared to all other chromosomes. LOH on chr4 is frequently the consequence of heterozygous deletions involving the *Cdkn2a* locus. By contrast, LOH on chr6 is predominantly copy number neutral and results from an increase in *Kras*^{G12D} gene dosage. Overall, the frequencies of CNV and LOH events are strongly increased on chromosomes carrying *Cdkn2a* or *Kras*^{G12D}, supporting the functional importance of gene dosage variation during PDAC evolution. **c-h**, Genetic mechanisms of *Kras*^{G12D} gene dosage alterations as identified by aCGH, M-FISH and whole exome sequencing (WES) in pancreatic cancers from PK mice. *General notes related to captions c-h*: The observed types of increased *Kras*^{G12D} gene dosage were: (i) focal gain (defined as gain affecting less than 50% of the chromosome length), arising either through replication-based mechanisms (2 cases, one with high-level *Kras*^{G12D} amplification shown in [c] and one with low level amplification) or translocation and subsequent amplification of the translocated chromosome (one case, shown in [d]), (ii) arm-level gain (defined as whole-chromosome gain or gain affecting more than 50% of the chromosome length) arising through mitotic errors (7 cases of whole-chromosome gain [example shown in e], occasionally [2 cases] with concomitant intra-chromosomal deletions or

translocations that do not affect *Kras* [example shown in f]) and (iii) copy-number neutral LOH (*Kras*^{G12D} homozygosity, acquired uniparental disomy), which can occur through either mitotic recombination (affects parts of the chromosome, shown in [g]) or chromosomal missegregation (duplication of the *Kras*^{G12D}-mutant chr6 and loss of wild-type chr6, shown in [h]). *Specific description of captions c-h:* **c**, Primary mouse PDAC S134 shows a higher-order focal amplification of *Kras*^{G12D} in aCGH. The sharp borders and small size of the amplification (600kb) as well as the strong increase in copy number (4x) indicate that *Kras*^{G12D} was amplified through multiple cycles of repeated template-switching of a replication-based DNA repair mechanism. The mutant *Kras*^{G12D} allele frequency detected by amplicon-based deep sequencing of the *Kras* locus is 89.1%. **d**, Tumor 4706 carries a focal amplification of the *Kras*^{G12D} locus. M-FISH analysis revealed an involvement of chr4 in *Kras*^{G12D} gain. The mutant *Kras*^{G12D} allele (chr6) was likely first affected by a reciprocal translocation of chr4 and chr6, resulting in two rearranged chromosomes: Der(4)T(4;6) and Der(6)T(4;6). In a subsequent event, Der(4)T(4;6) was missegregated through mitotic error resulting in focal gain of distal chromosome 6 carrying the *Kras*^{G12D} locus. The mutant *Kras*^{G12D} allele frequency is 72.2%. **e**, Primary mouse pancreatic cancer R1035 shows a 'classical' whole chromosome gain (trisomy) in aCGH and M-FISH, which was likely generated through mitotic error/missegregation. The mutant *Kras*^{G12D} allele frequency is 69.8%. **f**, In tumor 8442 arm-level gain of *Kras*^{G12D} was likely generated through mitotic missegregation of chr6. The intra-chromosomal deletion of the 19.6Mb on one of the three chromosomes does not affect the *Kras* locus. Note the reduced length of the chr6 with the intra-chromosomal deletion (asterisk) in M-FISH. The mutant *Kras*^{G12D} allele frequency detected by amplicon-based deep sequencing of the *Kras* locus is 66.4%. **g-h**, Primary mPDAC 16992 and B590 display copy-number neutral LOH (homozygosity, acquired uniparental disomy) leading to increased *Kras*^{G12D} gene dosage. The mutant *Kras*^{G12D} allele frequency for both cases is 99.2% and 96.3%, respectively. The SNP pattern of chr6 in mPDAC 16992 reveals that the whole chromosome is affected by CN-neutral LOH, indicating chromosome missegregation (duplication of the *Kras*^{G12D}-mutant chr6 and loss of wild-type chr6) as the underlying mechanism. By contrast, the SNP pattern of chr6 in primary tumor B590 reveals that only a part of the chromosome is subject to CN-neutral LOH and therefore probably resulting from mitotic recombination. **i**, Allele-specific *Kras*^{G12D} mRNA expression in *Kras*^{G12D-HET} (n=12) and *Kras*^{G12D-IGD} (n=26) primary PDAC cell cultures from PK mice as detected by combined analysis of amplicon-based RNA-Seq (determining the proportion of mutant and wild-type *Kras* mRNA) and 3'-prime pA RNA-Seq (detecting the amount of total *Kras* mRNA, but not

the proportion of mutant and wild-type *Kras* mRNA due to sequencing of 3'-prime transcript ends). This figure is related to Fig. 2b. Calculation of expression levels is described in the methods section. *** $P \leq 0.001$, two-tailed Mann-Whitney test; bars, median. **j**, Expression of mutant *Kras*^{G12D} mRNA in *Cdkn2a/Ncruc*^{ΔHET/WT} (n=11) and *Cdkn2a/Ncruc*^{ΔHOM} (n=27) primary PDAC cell cultures from PK mice as detected by combined amplicon-based RNA-Seq and 3'-prime pA RNA-Seq. This figure is related to Extended Data Figure 5f. *** $P \leq 0.001$, two-tailed Mann-Whitney test, bars, median. **k**, Expression of mutant *Kras*^{G12D} mRNA in indicated groups of mPDAC from PK mice (C2a n=5, C2b n=7, C2c n=6, C1 n=15) as detected by combined amplicon-based RNA-Seq and 3'-prime pA RNA-Seq. This figure is related to Fig. 5d. Expression levels of mutant *Kras*^{G12D} mRNA show positive correlation with assigned clusters and are highest in C1 ($P=1.6 \times 10^{-5}$, Pearson correlation). Expression levels are calculated as described in the methods section. Bars, median. **l-n**, Interphase fluorescence in situ hybridization (FISH) for the analysis of copy-number and ploidy states at the *KRAS* locus on chromosome 12 in human pancreatic intra-epithelial neoplasia (PanIN). Interphase FISH for the analysis of chr12 (red) and the *KRAS* locus on chr12 (green) was performed on human PanIN samples with *KRAS*^{G12} variant allele frequencies (VAFs) of ~100%. VAFs at the *KRAS* locus were detected by amplicon-based deep sequencing and are indicated above each FISH profile for different patients/PanINs. All analyzed PanIN samples showed a diploid genome as suggested by CEN12 (two signals per nucleus). A VAF of ~100% can be caused either by loss of the wild-type *KRAS*-locus (hemizyosity of *KRAS*^{G12-MUT}: one *KRAS*^{G12-MUT} allele per cell) or by CN-neutral LOH (acquired uniparental disomy; homozygosity of *KRAS*^{G12-MUT}: two *KRAS*^{G12-MUT} alleles per cell). No loss of one *KRAS* allele nor monosomy of chr12 were observed. Thus, VAF frequencies of ~100% in human PanINs are the result of CN-LOH associated with an increased *KRAS*^{G12-MUT} gene dosage. Scale bars, 2.5μm; CEN12, centromere probe chr12.

Extended Data Figure 4 | Enrichment for amplification of alternative oncogenic drivers in mPDACs of PK mice with *Kras*^{G12D-HET} status. **a-b**, Two primary mPDACs with strong focal *Myc* amplification on chr15 are shown, as detected by aCGH. Red dashed line indicates no copy number change. **c-d**, Focal copy number gains targeting the *Yap1* locus on chr9 in primary mPDACs 4072 and 9203 as revealed by aCGH. **e**, Chr19 was also frequently subject to arm-level gain (see Fig. 1c and Extended Data Figure 1l). Primary mPDAC of PK mouse 4072 harbors a focal gain on chr19 containing 20 genes: *9130011E15Rik*, *Gm6813*, *Hps6*, *Ldb1*, *Pprc1*, *Nolc1*, *Elovl3*, *Pitx3*, *Gbf1*, *Nfkb2*,

Psd, *Fbxl15*, *Cuedc2*, *Tmem180*, *Actr1a*, *Sufu*, *Trim8*, *Arl3*, *Sfxn2*, *D19Wsu162e*. **f**, Cross-species analyses revealed that the orthologous region on human chr10 is also subject to recurrent amplifications in human PDAC (8 out of 109 hPDACs have focal amplifications; data from Witkiewicz *et al.*⁶). Of the 20 mouse genes, sixteen could be assigned to orthologues in humans. Further analyses revealed that only two genes, *NFKB2* and *PSD*, are within the minimal overlapping region of recurrent amplification (data from⁶ and oncoplot from cBioPortal^{60,61}). **g**, *NFKB2*, but not *PSD*, shows medium protein expression in exocrine glandular cells of normal pancreatic tissue, as detected by immunohistochemistry (IHC, data from TheHumanProteinAtlas⁶²). **h**, *NFKB2* is highly expressed in 17% (2/12) of stained hPDAC biopsies as shown by IHC. In contrast, there was no *PSD* expression in any of the analyzed pancreatic cancers (0/12). Protein expression data was used from TheHumanProteinAtlas⁶².

Extended Data Figure 5 | Characterization of *Cdkn2a* (chr4) alterations and correlation with *Kras*^{MUT} gene dosage variation and mRNA expression in mouse and human PDAC. **a-d, *Cdkn2a* alteration on mouse chr4 can occur through arm-level, complex or focal loss as well as uniparental disomy (see Figure 3). In addition, chr4 is frequently involved in inter-chromosomal translocations. Examples of representative karyotypes of primary pancreatic cancer cultures derived from PK mice with translocations involving chr4, likely affecting the *Cdkn2a* locus. In all 4 cases, chr4 translocations were found in all 10 metaphase spreads of each cancer, indicating their early acquisition during tumor evolution. **a**, mPDAC 4706 with diploid karyotype: 42, XX, del(X), +2, der(2)t(2;4)is(2;4), der(4)t(4;6)*2, +der(4)t(2;4), der(6)t(4;6). **b**, mPDAC 4900 also features a diploid karyotype: 41, XX, der(X)is(X;4), der(4)is(4;8), del(4), +6, der(8)t(4;8). **c**, mPDAC 5123 underwent polyploidization, after translocation of chr4 with chr1 and a deletion on the other copy: 78, XXXX, -1, del(1)*2, -2, +4*2, der(4)t(1;4)*3, del(4)*3, -5, -7, -9, +15, -17, +18 **d**, mPDAC 8349 shows a diploid karyotype: 40,XX, der(4)t(3;4), der(4)t(4;13), +del(4), der(13)t(4;13). **e**, *Kras*^{G12D} variant allele frequencies detected by amplicon-based deep sequencing of the *Kras* locus are higher in *Cdkn2a/Ncruc*^{ΔHOM} mPDAC (n=11) as compared to *Cdkn2a/Ncruc*^{ΔHET/WT} (n=27) pancreatic cancers. All cancers are from PK mice. Blue dots indicate tumors with complete *Ncruc* deletion. ***P≤0.001, two-tailed Mann-Whitney test; bars, median. **f**, Allele-specific expression of mutant *Kras*^{G12D} mRNA is increased in primary tumors from PK mice with *Cdkn2a/Ncruc*^{ΔHOM} (n=27) background in comparison to *Cdkn2a/Ncruc*^{ΔHET/WT} (n=11) cancers. Primary mPDACs with homozygous loss of *Ncruc* are highlighted in blue. *Kras*^{G12D}**

expression was analyzed by combining amplicon-based RNA-Seq and qRT-PCR (as described in the Methods section). **P=0.003, two-tailed Mann-Whitney test; bars, median. **g**, *KRAS*^{MUT} variant allele frequencies based on WES in a published dataset of microdissected human PDAC (Witkiewicz *et al.*, reduced stromal content) was analyzed with respect to *CDKN2A* and *TP53* status. *KRAS*^{MUT} allele frequency was higher in mutated/homozygous deleted *CDKN2A* and/or *TP53* (*CDKN2A*^{MUT/ΔHOM}/*TP53*^{MUT/ΔHOM}; hPDACs as compared to cancers with *CDKN2A*^{ΔHET/WT}/*TP53*^{ΔHET/WT} status (from left: n=28, n=14, n=28, n=30). Two-sided rank-based ANOVA (P=5.8*10⁻⁶); post hoc testing with two-sided Tukey honest significant difference test, *adj. P≤0.05, ***adj. P≤0.001; bars, median. **h**, Fraction of the genome altered by copy number changes detected by aCGH in primary mPDACs of PK (n=38), PKC (n=16) and PKP (n=16) mice. PKP mice show a significantly increased CNA load as compared to PKC mice. Two-sided rank-based ANOVA (P=0.01); post hoc testing with two-sided Tukey honest significant difference test, **adj. P=0.009, adj. P-values for group wise comparisons are shown; bars, median. Del, deletion; der, derivative chromosome; is, insertion; t, translocation; „-“, chromosome loss; „+“, chromosome gain.

Extended Data Figure 6 | Complete *Cdkn2a* barrier loss precedes *Kras*^{G12D-iGD} in primary mPDAC of PK mouse 53704. Copy number alterations at chr4 (*Cdkn2a*) and chr6 (*Kras*) in mPDAC 53704 and corresponding metastases, as detected by aCGH (top) and whole-exome sequencing based SNP pattern analysis (bottom). The primary cancer and both liver metastases display identical focal deletions of *Cdkn2a* and similar SNP patterns on chr4 revealing that all lesions share the same ancestor cell with complete *Cdkn2a* loss. By contrast, SNP analysis on chr6 revealed discordant patterns in the primary mPDAC and both metastases. Li2 shows partial LOH of a distal region on chr6 involving the *Kras* locus, while LOH in Li3 involves the whole chr6. This explains the step-wise LOH pattern observed on chr6 in the primary mPDAC. The graphic on the right shows the combined interpretation of CNV/LOH profiles, which suggests the following sequence of genetic events during tumor evolution: The initial *Kras*^{G12D} mutation was followed by focal deletion of one copy of *Cdkn2a*. In a subsequent genetic event, the second copy of *Cdkn2a* was lost by chr4 missegregation and copy-number neutral LOH. Complete barrier loss allowed for convergent evolution of increased *Kras*^{G12D} gene dosage through copy-number neutral LOH and gave rise to independent metastases in the liver. *Note*: A major obstacle for equivalent human studies is the limited availability of human matched primary/metastases samples, particularly of treatment naive ones. We performed cross-species

analyses using data from a recent study, which analyzed human treatment-naive metastatic PDACs by whole-genome sequencing⁸ and provided *CDKN2A* and *KRAS* copy number data for matched primaries/metastases from 3 patients. In one patient the sequential order of *CDKN2A* deletion and *KRAS* amplification could be reconstructed: homozygous *CDKN2A* deletions were identical in all primaries and metastases, whereas there were 5 different *KRAS* gains in the 6 metastases. This suggests convergent evolution of mutant *KRAS* gene dosage gain upon homozygous *CDKN2A* loss in this patient, in line with similar data in large series of mouse cancers and their metastases (see Figure 3e).

Extended Data Figure 7 | Transcriptome-based subtyping of human primary pancreatic cancer and classification of human PDAC cell lines and primary PDAC cell cultures from PK mice.

General background: Three studies have previously classified human primary pancreatic cancer based on transcriptional profiles (Collisson *et al.*²⁸, Moffitt *et al.*²⁹ and Bailey *et al.*⁷). Collisson *et al.* performed PDAC microdissection and defined 3 transcriptional subtypes: classical, quasimesenchymal (QM) and exocrine-like. Moffitt *et al.* defined 2 subtypes (classical, basal-like) and proposed that the exocrine-like signature stems from stromal cells (healthy exocrine pancreatic cells), rather than from the cancer cells. This was based on (i) virtual microdissection of PDAC (mathematical modelling; NMF-based separation of tumor, and non-tumor gene expression patterns), (ii) transplantation studies (lack of an exocrine-like signature in patient derived xenografts) and (iii) the observation that human PDAC cell lines do not show the exocrine-like signature (see also Extended Data Figure 7e-h). Bailey *et al.* used bulk tumors and defined 4 subtypes (pancreatic progenitor, immunogenic, squamous, ADEX). **Specific description of captions a-c:** To perform independent cross-comparisons of the three classification systems, we used RNA-Seq data from PDAC and adenosquamous pancreatic carcinoma from Bailey *et al.* Other histological subentities of pancreatic cancer were excluded (e.g. IPMN, MCN, acinar cell carcinoma). The Bailey subtyping for this dataset was available. The data were subjected to unbiased hierarchical clustering using classifier gene lists from either Collisson *et al.* (a) or Moffitt *et al.* (b). This allowed (in c) assignment of the Bailey subtype, Collisson subtype and Moffitt subtype for each cancer. **a**, Unbiased hierarchical clustering of primary pancreatic cancer samples (n=71) from Bailey *et al.* using Collisson classifier genes resulted in almost complete separation of classifier genes on the y-axis. On the x-axis three distinct sub-clusters could be identified with: quasimesenchymal and exocrine-like genes generating clusters, which map mostly

to Bailey's squamous and ADEX subtypes, respectively. Classical Collisson classifier genes formed a third distinct cluster that is enriched for Bailey subtypes pancreatic progenitor and immunogenic. **b**, Subtyping of primary pancreatic cancer samples (n=71) from Bailey *et al.* using classifier genes defined by Moffitt *et al.* resulted in perfect separation of classifier genes on the y-axis. Two clusters could be identified on the x-axis with: (i) upregulated expression of Moffitt classical genes enriched for in Bailey subtypes ADEX, pancreatic progenitor and immunogenic and (ii) increased expression of basal-like genes was associated with squamous Bailey pancreatic cancer subtypes. **c**, Consensus clustering showing the overlap of the three classification systems. Assignment of the Bailey subtype, Collisson subtype and Moffitt subtype was extracted for each cancer from the analyses of (a,b). Samples are sorted according to classification by Collisson *et al.* **Key conclusions:** (i) There is large overlap of Bailey subtypes to the initial classification proposed by Collisson *et al.* The Bailey classification (based on bulk tissue analyses) suggests that among Collisson classical cancers (which were defined using microdissected cancer tissue) there are some with and some without a strong immune cell infiltration. (ii) Moffitt *et al.* suggested that the Collisson exocrine-like signature (Bailey ADEX subtype) stems from "contaminating" healthy exocrine pancreatic cells, based on the evidence described above. Given that the Collisson exocrine-like signature was derived from microdissected PDAC, such "contamination" of a signature is only conceivable, if exocrine-like signature genes were dramatically higher expressed in pancreatic acinar cells as compared to PDAC cells. To explore this possibility, we compared gene expression profiles of human wild-type pancreas and human PDAC cell lines. hPDAC cell lines showed very low expression of the Collisson exocrine-like genes. **d**, Volcano plot shows strongly upregulated expression of exocrine-like genes in human wild-type pancreas samples (13 to 241 fold; median: 183 fold upregulation). Note that 15 out of 19 exocrine-like signature genes are represented within the top50 genes upregulated in human wild-type pancreas (n=3) as compared to hPDAC cell lines (n=30) (y axis is calculated on Benjamini-Hochberg adj. P-values derived from R package limma [see Methods section]). Although these data do not exclude the existence of exocrine-like pancreatic carcinomas, they do raise the possibility that "contamination" with few acinar cells could impose an exocrine-like signature on a cancer. This might explain why human PDAC cell line resources as well as our mouse PDAC cell culture collections do not cluster into the exocrine-like subtype (see also Extended Data Figure 7e-f). Thus, between the above studies, there is considerable overlap/consensus between at least two subtypes, which are in large parts captured by the initially proposed Collisson classical and quasimesenchymal (QM) signatures (which can also be

detected in mouse and human PDAC cell lines (see also Extended Data Figure 7e-h). **e**, Hierarchical clustering of microarray-based expression profiles using Collisson identifier genes²⁸ on human PDAC cell lines (n=19, GEO series GSE17891). As also described earlier by Collisson *et al.*, only two subtypes can be detected in human cell line collections: classical and quasimesenchymal (QM). Of note, the most prominent change in the QM cell lines is downregulation (extinction) of the classical assigner genes whereas the expression of QM classifier genes is quite variable. We therefore use here the terms classical and non-classical (QM) for subtypes hereafter. **f**, Projection of the Collisson classifiers on our mouse PDAC cell culture transcriptomes (n=33) also identified classical and non-classical (QM) subtypes. The non-classical (QM) subtype contained a subset of mPDAC cell cultures from cluster C2 (epithelial morphology) and all cluster C1 (mesenchymal morphology) mPDACs. **g**, Indeed, application of a human EMT hallmark gene set⁵² for hierarchical clustering of expression profiles of primary PDAC cultures from PK mice (n=33) resulted in a clear separation of samples from C1 (mesenchymal) and C2 (epithelial). The equivalent of murine C1 ("M") has not been described in human cell lines so far, reflecting the underrepresentation of mesenchymal phenotypes in available human cell line collections. **h**, Indeed projecting the EMT hallmark gene set on human PDAC cell line transcriptomes (n=19, GEO series GSE17891) did not result in a clear separation of samples, indicating that the full mesenchymal phenotype is not present in this set of hPDAC cell lines. As shown in Extended Data Figure 9b, however, the EMT signature is detectable in undifferentiated human pancreatic carcinoma, which is the human equivalent of the mesenchymal mouse PDACs in C1.

Extended Data Figure 8 | Functional analyses to study the role of *Kras*^{G12D} gene dosage increase in EMT. a-d, Multiplexed somatic CRISPR/Cas9 mutagenesis for phylogenetic tracking of epithelial/mesenchymal mPDAC clones *in vivo*. **a**, Graphic demonstrates major steps of multiplexed gene editing by pooled delivery of CRISPR/Cas9 vectors, each targeting a different tumor suppressor gene in the pancreas of PK mice. Electroporation-based transfection induces low-frequency mosaic vector delivery (average of 120 cells per pancreas are transfected) to induce clonal tumors. Primary tumor cell cultures were screened for the simultaneous presence of epithelial and mesenchymal cells. Two such cancers were identified (mPDACs from mouse 021 and mouse 901) and subjected to differential trypsinization in order to enrich for each morphology. **b**, Amplicon-based deep sequencing of all sgRNA-targeted loci revealed identical indel patterns in both epithelial/mesenchymal culture pairs. This shows (i) that epithelial and mesenchymal cells originate

from the same clone and (ii) that the CRISPR-induced mutations are not contributing to the differential phenotype. **c**, *Kras*^{G12D} variant allele frequencies in epithelial and mesenchymal cell cultures from mPDAC 021 and mPDAC 901, as detected by amplicon-based deep sequencing. Both cancers had increased *Kras*^{G12D} expression in mesenchymal cells (see Fig. 5e). In mPDAC 021 this is due to selective amplification of the *Kras*^{G12D} allele in mesenchymal cells. In mPDAC 901 genetic *Kras*^{G12D} amplification was not observed, suggesting induction of increased *Kras* expression in mesenchymal cells by other mechanisms. **d**, Gene set enrichment analysis using “Molecular Signatures Database” (MSigDB) of differentially regulated genes in mesenchymal versus epithelial mPDACs based on RNA-Seq. Mesenchymal clones of mPDAC 021 and mPDAC 901 show an upregulation of genes involved in “MAPK signaling pathway” and “EMT” as compared to the corresponding epithelial clones, in line with increased *Kras*^{G12D} gene dosage (a full list of enriched gene sets is provided for comparisons in Supplementary Table 15). FDR-adjusted P-values are shown on y axis. Representative data from one experiment are shown. **e-g**, induction of EMT-like transcriptional programs by *KRAS*^{G12D} overexpression in human PDAC cell lines. **e**, Graphic of experimental workflow. Two human PDAC cell lines (HUPT3 and PANC0327) with homozygous *CKDN2A* loss (*CDKN2A*^{ΔHOM}) and heterozygous *KRAS*^{MUT} (*KRAS*^{MUT-HET}) status were transduced with lentivirus carrying doxycycline-inducible *KRAS*^{G12D} or GFP-control expression constructs. *KRAS*^{G12D} or GFP expression was induced by adding doxycycline for 1, 3 and 5 days. **f**, Gene set enrichment analysis using “Molecular Signatures Database” (MSigDB) of differentially regulated genes in *KRAS*^{G12D}- versus GFP-induced hPDAC cell lines HUPT3 and PANC0327 based on RNA-Seq. Upon doxycycline treatment, both hPDAC cell lines showed a consistent upregulation of genes involved in “KRAS signaling up” and “EMT” (a full list of enriched gene sets is provided for both cell lines in Supplementary Table 16). FDR-adjusted P-values are shown on y axis. **g**, Expression of marker genes for epithelial (*CDH1*) or mesenchymal (*VIM*) cell differentiation and invasion/matrix disassembly (*MMP1*) was validated by qPCR (normalized to *GAPDH* and *PPIA*). In line with RNA-Seq data *KRAS*^{G12D}-induced cells show an increased expression of the mesenchymal marker gene *VIM*, increased expression of *MMP1* and reduced levels of epithelial marker gene *CDH1*. *P≤0.05, **P≤0.005, ns=not significant, two-tailed t-test; bars=mean; error bars=SEM.

Extended Data Figure 9 | Transcriptional profiles of human undifferentiated pancreatic carcinomas are enriched for signatures of oncogenic signaling intensification and EMT but not

for activation of *TP63* Δ *N* transcriptional network. *General background:* Mouse PDAC cell lines with mesenchymal morphology, undifferentiated histopathology, highest *Kras*^{G12D} expression and a „C1” transcriptome (EMT signature, increased Ras signaling; see Fig. 5) do not have a counterpart in the small set of publicly available human cell line collections (see Extended Data Figure 7h). Undifferentiated human pancreatic cancers are typically advanced and might therefore be underrepresented (1-3%) in human surgical series or cell line collections, whereas in autopsy series up to 16% of hPDACs were reported to have at least focal undifferentiated components^{31,32}. Dedifferentiation can occur during disease progression and is associated with worse prognosis^{32,33}, which is also reflected in mice (Extended Data Figure 9a). *Specific description of captions:* **a**, Undifferentiated/mesenchymal pancreatic carcinomas from PK mice show a tendency towards reduced survival. Primary pancreatic tumors from PK mice with a mesenchymal phenotype (C1 cluster, n=15) are almost exclusively classified as undifferentiated/sarcomatoid by histopathological evaluation and tend to have a reduced age at diagnosis when compared to epithelial (C2 cluster, n=18) tumors (histopathological grade 1 to 3 [G1-G3]). This aggressive behavior of undifferentiated pancreatic carcinoma is also observed in human patients and is associated with worse clinical outcome³³. P-value calculated by two-sided log-rank test. **b**, Analysis of expression profiles from human undifferentiated pancreatic carcinoma (n=4) and comparison to PDAC (WHO grade 1 to 3 [G1-G3], n=65) as well as adenosquamous pancreatic carcinoma (n=6). Samples with the above histopathological characteristics for which expression-based subtype information from Bailey *et al.*⁷ is available were used and complemented with available undifferentiated pancreatic carcinomas from the ICGC PACA-AU cohort (Supplementary Table 18). Other histological subentities of pancreatic cancer were excluded (e.g. IPMN, MCN, acinar cell carcinoma). ANOVA was performed to select genes which are differentially expressed in at least one of the six defined subgroups of pancreatic cancer: (i) undifferentiated, (ii) adenosquamous pancreatic carcinoma and (iii-vi) PDAC (G1-G3) sub-stratified in pancreatic progenitor, immunogenic, squamous and ADEX Bailey subtypes. Differentially regulated genes were used for unbiased hierarchical clustering of these pancreatic cancer transcriptional profiles. Five sub-clusters of co-regulated gene expression could be identified according to the cluster tree on the y-axis (separated by white horizontal bars in the heatmap). Gene set enrichment analysis using “Molecular Signatures Database” (MSigDB) was performed for individual sub-clusters and terms related to predominating gene sets/pathways are annotated for each cluster on the right (full list provided in Supplementary Table 17). Undifferentiated pancreatic carcinomas cluster together and are

associated with (i) upregulation of genes in cluster 3 (containing MAPK signaling pathway and gene sets relevant during embryonic development or EMT) and (ii) downregulation of genes in clusters 2 and 5, which contain gene sets related to epithelial cell differentiation, embryonic development or metabolic signatures. This reflects the pathway enrichment signature in the equivalent undifferentiated (mesenchymal) mouse PDACs (cluster C1 in PK mice) and provides further support for the link between *KRAS* signaling intensification, EMT and the undifferentiated tumor phenotype. The immunogenic PDAC subtype showed high expression of cluster 4 genes, which was also strong (even elevated) in undifferentiated pancreatic carcinomas, suggesting an increased immune cell infiltration in undifferentiated carcinomas. Cluster 1 contained gene sets related to cell proliferation/cell cycle, squamous differentiation and *TP63ΔN* transcriptional targets which were most highly overexpressed in pancreatic carcinomas with adenosquamous histology. Undifferentiated pancreatic carcinomas did not show activation of the *TP63ΔN* transcriptional targets. This suggests that activation of *TP63ΔN* transcriptional targets is not causally linked to *KRAS* signaling intensification and EMT (see also Extended Data Figure 9c-d, showing a lack of association of undifferentiated carcinomas with *TP63ΔN* transcriptional network activation). **c**, Unbiased hierarchical clustering of human pancreatic carcinomas with adenosquamous histology (n=7) as well as PDACs (WHO grade 1 to 3 [G1-G3], n=64) and undifferentiated pancreatic carcinomas (n=4) (sample set as in Extended Data Figure 9b) using a list of validated *TP63ΔN* transcriptional targets⁵³. Pancreatic cancers with adenosquamous differentiation were significantly enriched in a cluster showing increased *TP63ΔN* transcriptional network activity ($P \leq 0.001$, two-sided Fisher's exact test, OR 130, 95% CI 11.6-1452). Undifferentiated pancreatic carcinomas did not contribute to this cluster. In line, pancreatic cancers from PK mice did not show differential regulation of the *TP63ΔN* network, reflecting the lack of adenosquamous tumors in this cohort (not shown). **d**, Unbiased hierarchical clustering across solid cancers (Cancer Cell Line Encyclopaedia, n=856) using the same gene list showed a strong enrichment of tumors with squamous differentiation in the sub-cluster with highest *TP63ΔN* transcriptional network expression ($P \leq 0.001$, two-sided Fisher's exact test, OR 28.1, 95% CI 16.4-48.1), in line with the observation of Hoadley *et al.*⁶³ that *TP63ΔN* is a signature for squamous differentiation across cancers.

Extended Data Figure 10 | *Kras*^{G12D}-gene dosage is a critical determinant of PDAC biology in a mouse model with high mutational load. The mutational burden in primary PDAC cultures of PK mice was significantly lower as compared to human PDAC studies (see Fig. 1b). To account for this

potential confounding factor and to test if our discoveries in PK mice also apply in a setting of high mutational burden, we used a mouse model combining *Kras*^{G12D} mutation and *PiggyBac* transposon-based insertional mutagenesis (PK-PB mice¹³). PK-PB mice show accelerated tumorigenesis as compared to PK mice. PK-PB derived tumors had an extensive mutational burden (median of 494 transposon insertions per tumor). Primary cultures of PDAC from PK-PB mice (n=17) were subjected to comprehensive genetic characterization using aCGH, microarray-based gene expression profiling, quantitative transposon insertion site sequencing (QiSeq) and amplicon-based deep sequencing of the *Kras* locus. **a**, Transcriptome profiles of primary PDAC cultures from PK-PB mice (n=17) were used for unbiased hierarchical clustering that resulted in 2 major clusters (C1 and C2), like in PK mice. *Kras*^{G12D} gene dosage status (as determined by aCGH and amplicon-based deep sequencing of the *Kras* locus) and *Cdkn2a* status (as determined by aCGH and quantitative transposon insertion site sequencing [QiSeq]) are indicated below the cluster tree for each individual tumor. Similarly to PK mice, cluster C2a was characterized by *Kras*^{G12D-HET} and *Cdkn2a/Cdkn2a*^{AHET/WT} status, whereas mPDACs in clusters C2b/c and C1 had increased *Kras*^{G12D} gene dosage (*Kras*^{G12D-iGD}) and were *Cdkn2a/Ncruc*^{ΔHOM}. The genetic *Kras*^{G12D}-status was significantly associated with expression clusters (P=0.01, two-sided Fisher's exact test) providing further evidence that expression clusters are associated with *Kras*^{G12D} gene dosage. **b**, Prevalence of *Kras*^{G12D-iGD} in cultures of primary mPDAC (from PK-PB mice) with homozygous (n=12) or heterozygous/wild-type (n=5) *Cdkn2a/Ncruc* status. *P=0.03, two-sided Fisher's exact test, OR 20.0, 95% CI 1.4-287.8. **c**, Gene set enrichment analysis using DAVID of upregulated genes in cluster C1 (n=5) as compared to cluster C2 (n=12) of primary mPDAC cultures from PK-PB mice. As in PK mice, PK-PB tumors in C1 are characterized by upregulation of genes enriched in gene sets describing mesenchymal cell differentiation and revealed a strong enrichment for Ras downstream signaling pathways (full list in Supplementary Table 19). FDR-adjusted P-values are shown on y axis. Overall, these analyses show that the biological principles discovered in the PK model also apply to pancreatic cancers from PK-PB mice with high mutational load.

Figure 1

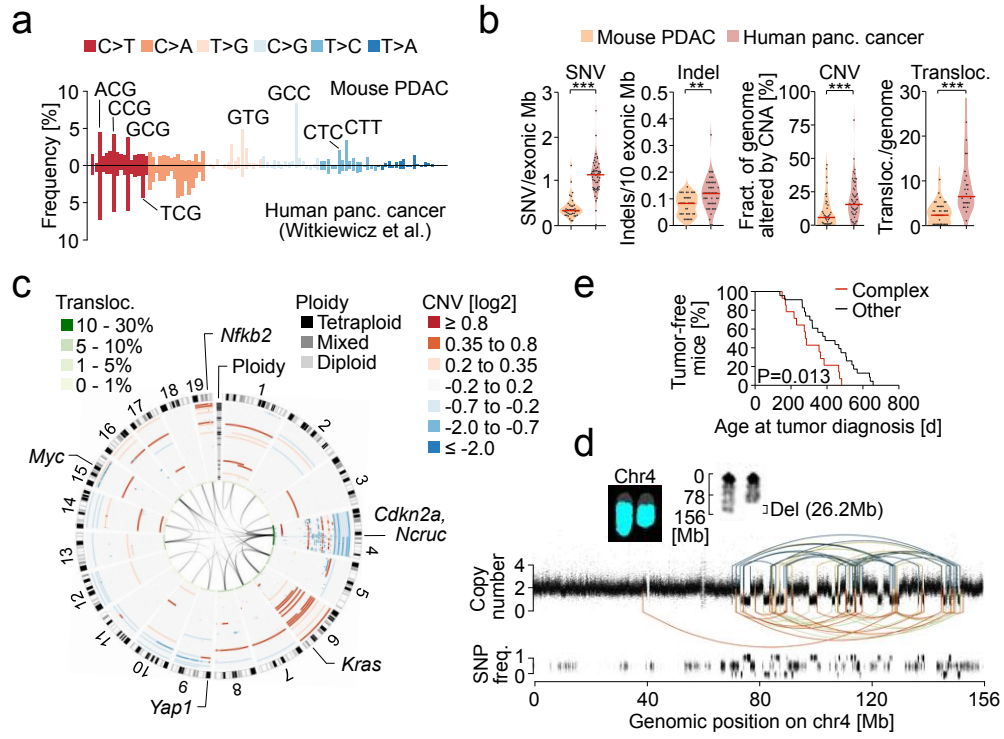


Figure 2

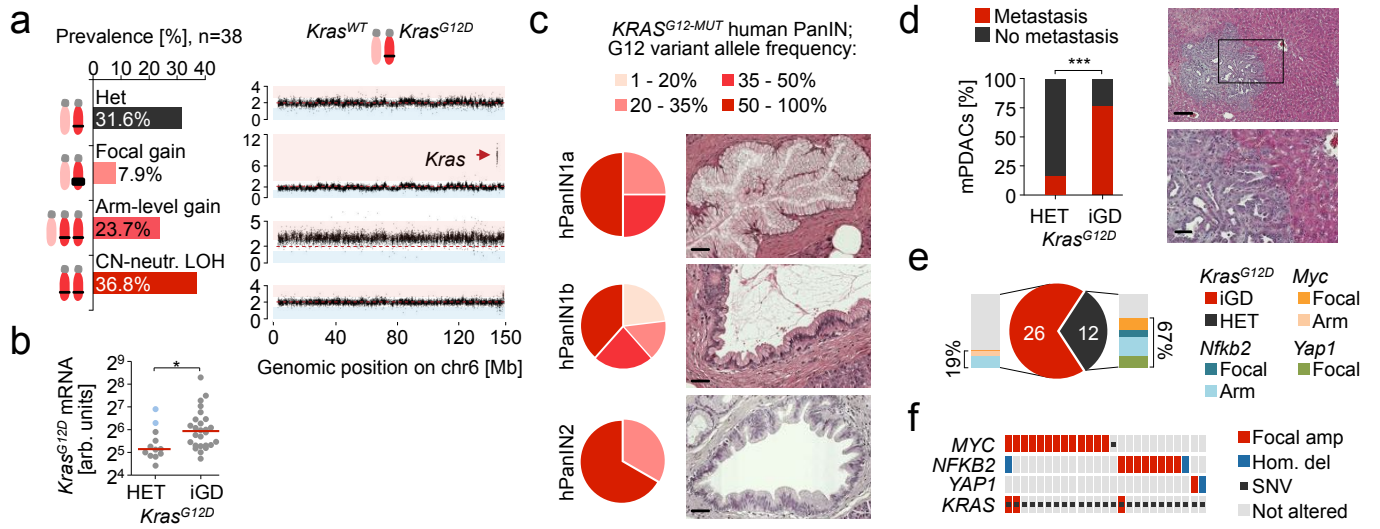


Figure 3

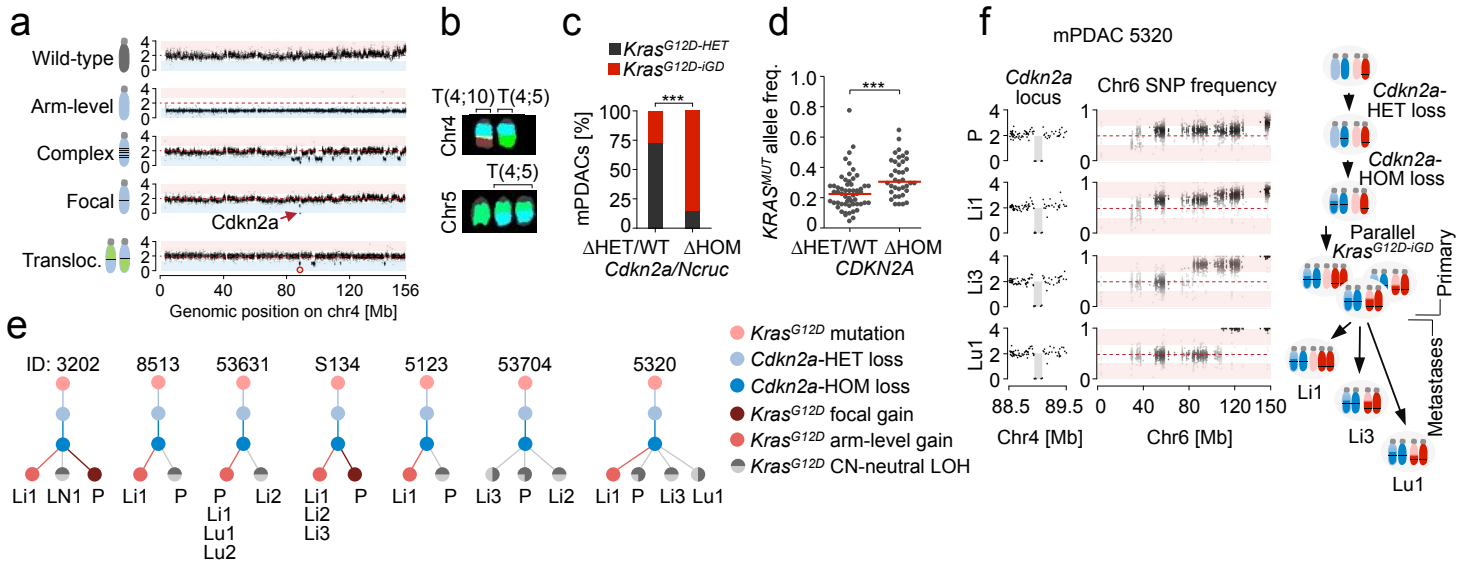


Figure 4

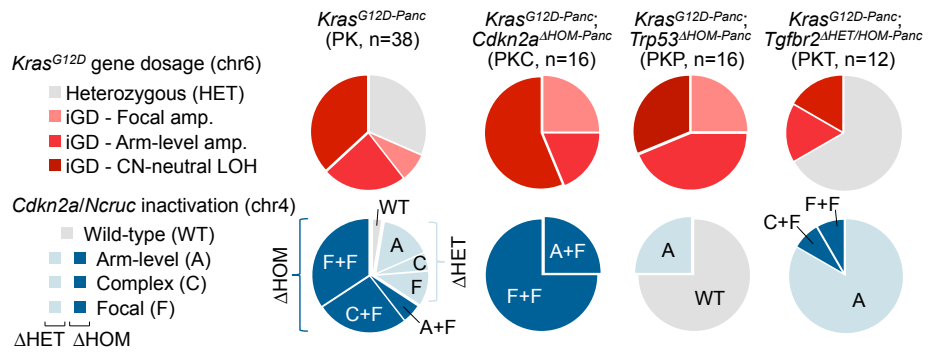
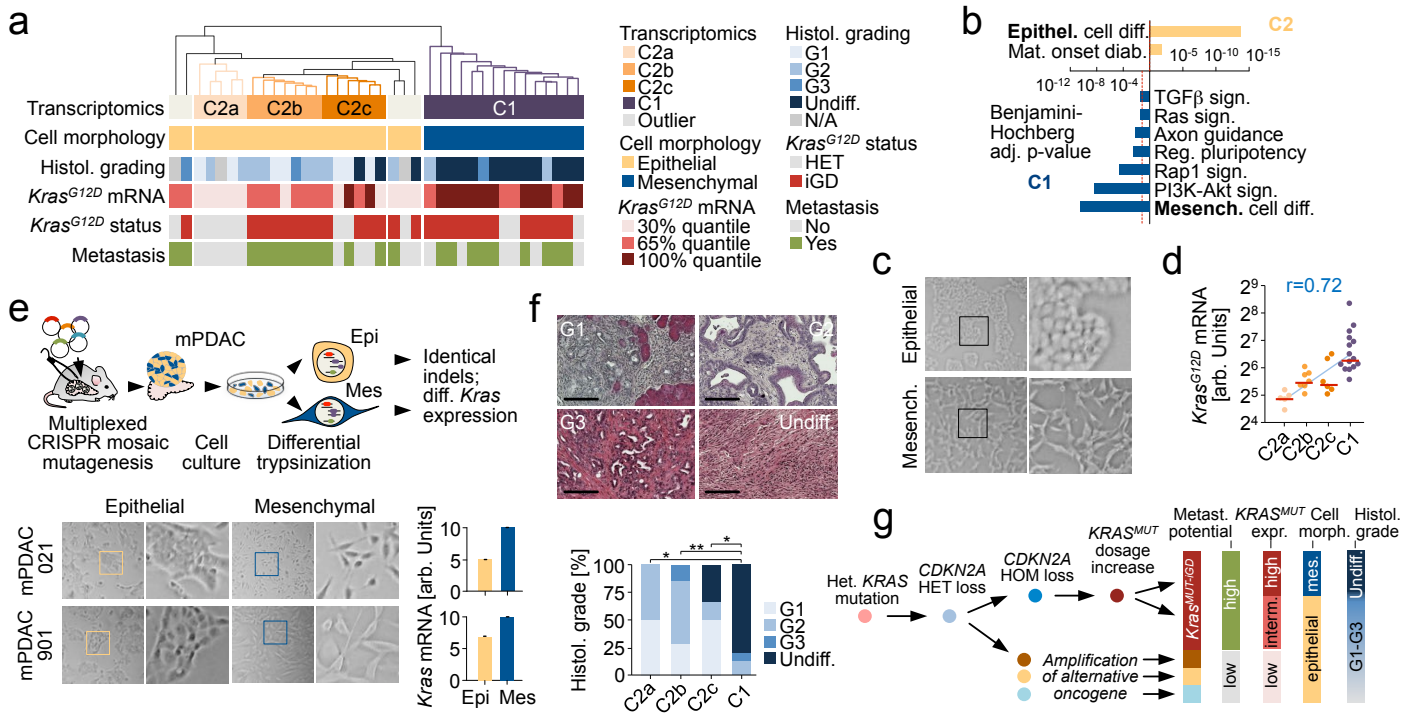
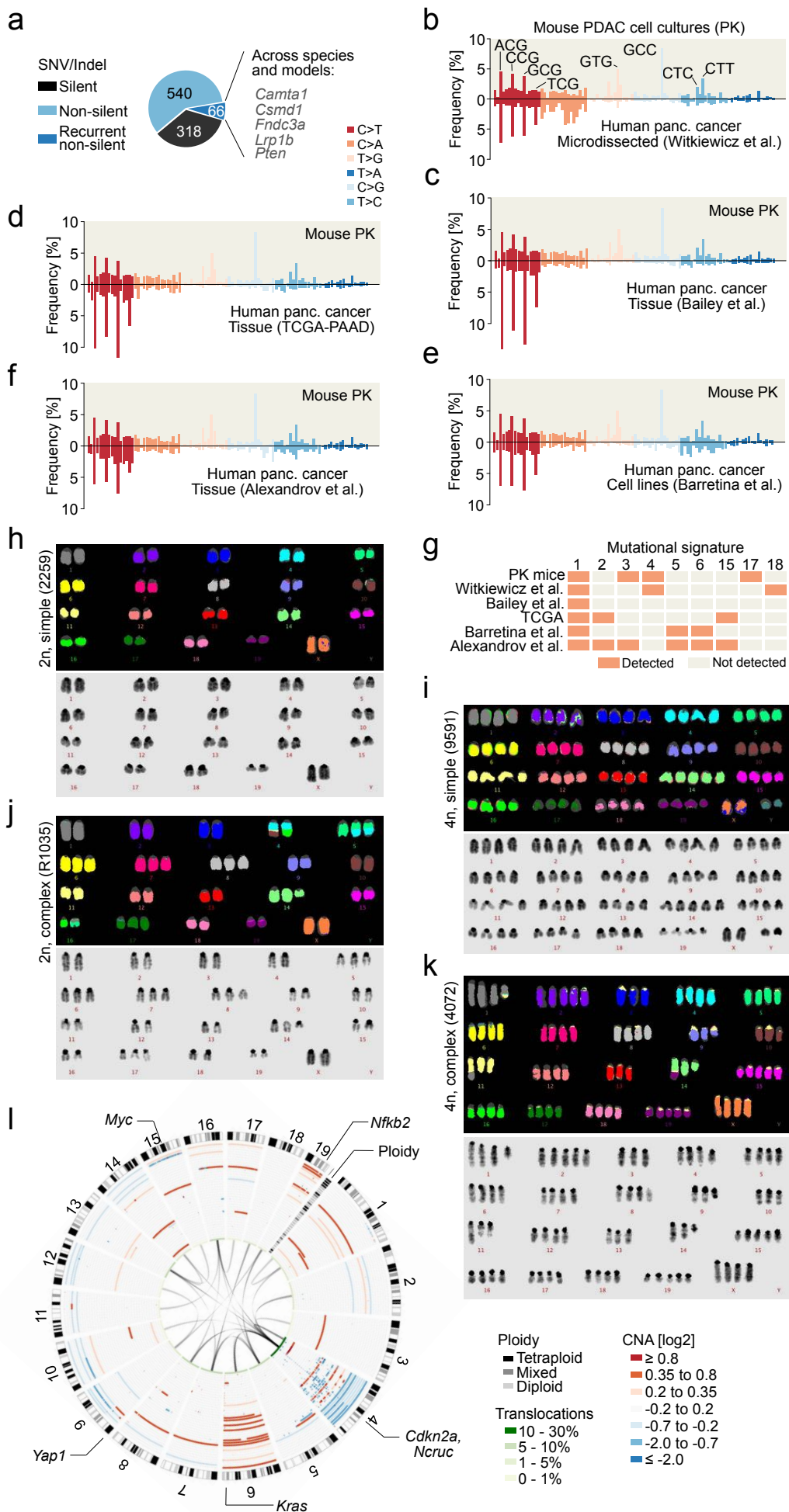


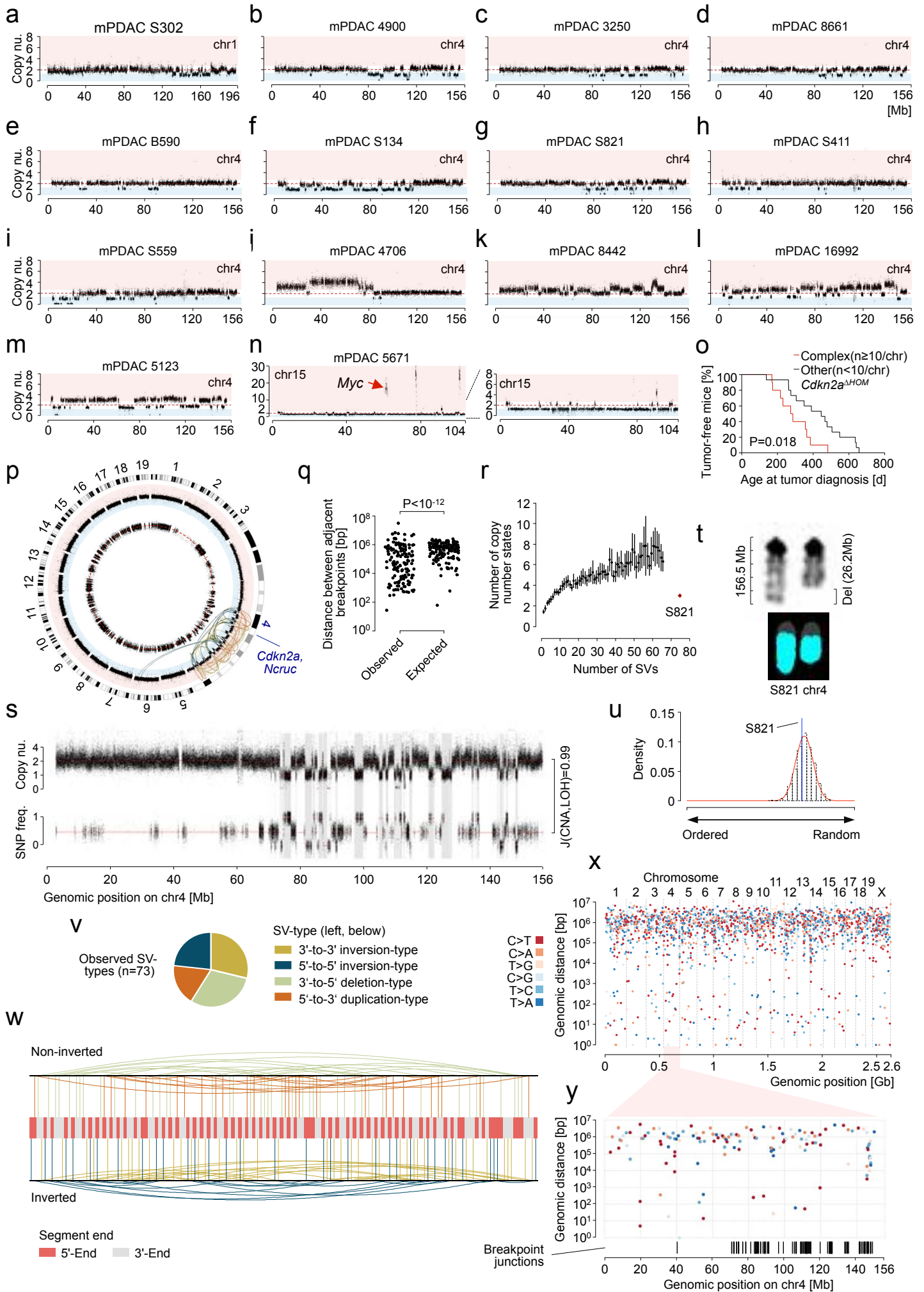
Figure 5



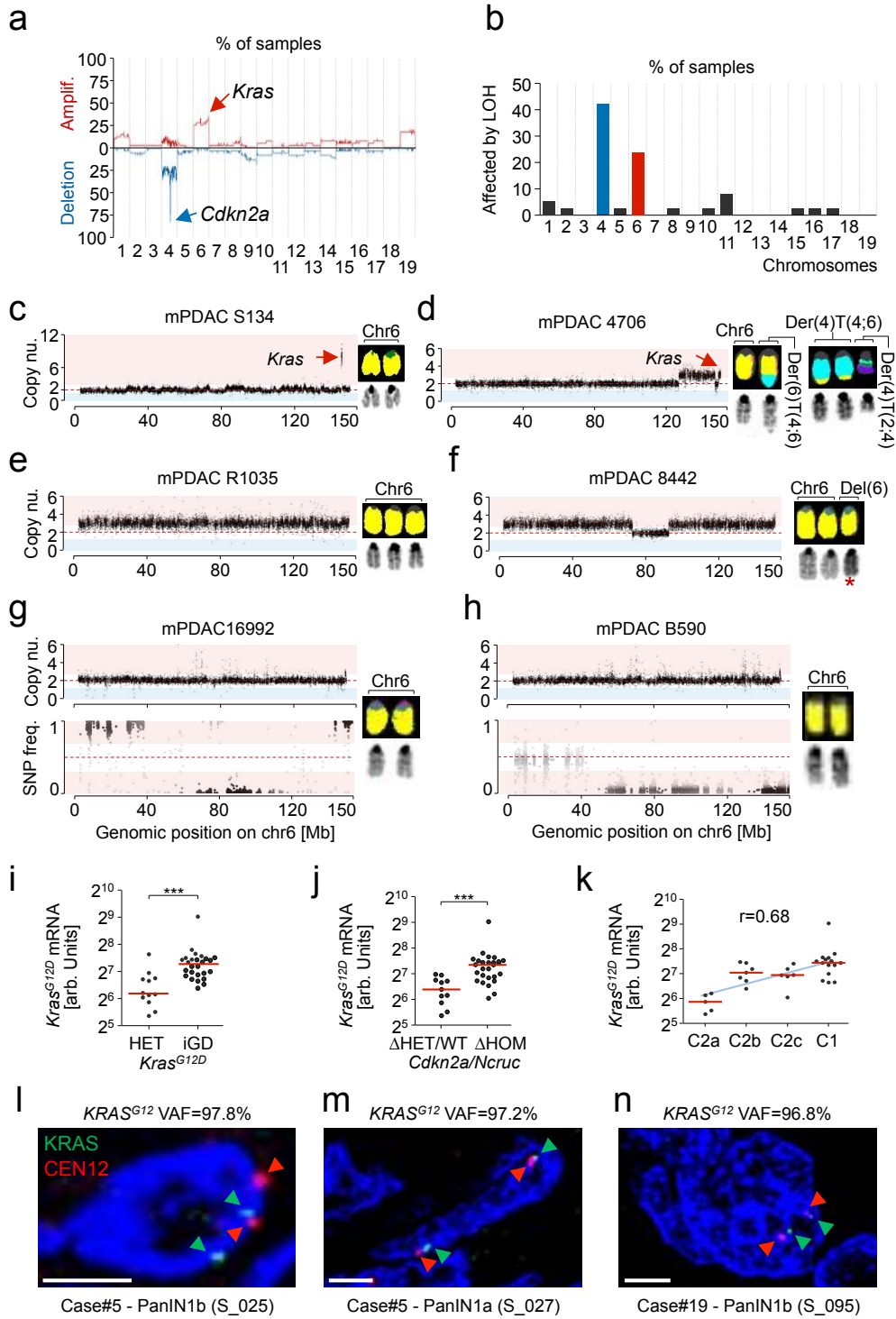
Extended Data Figure 1



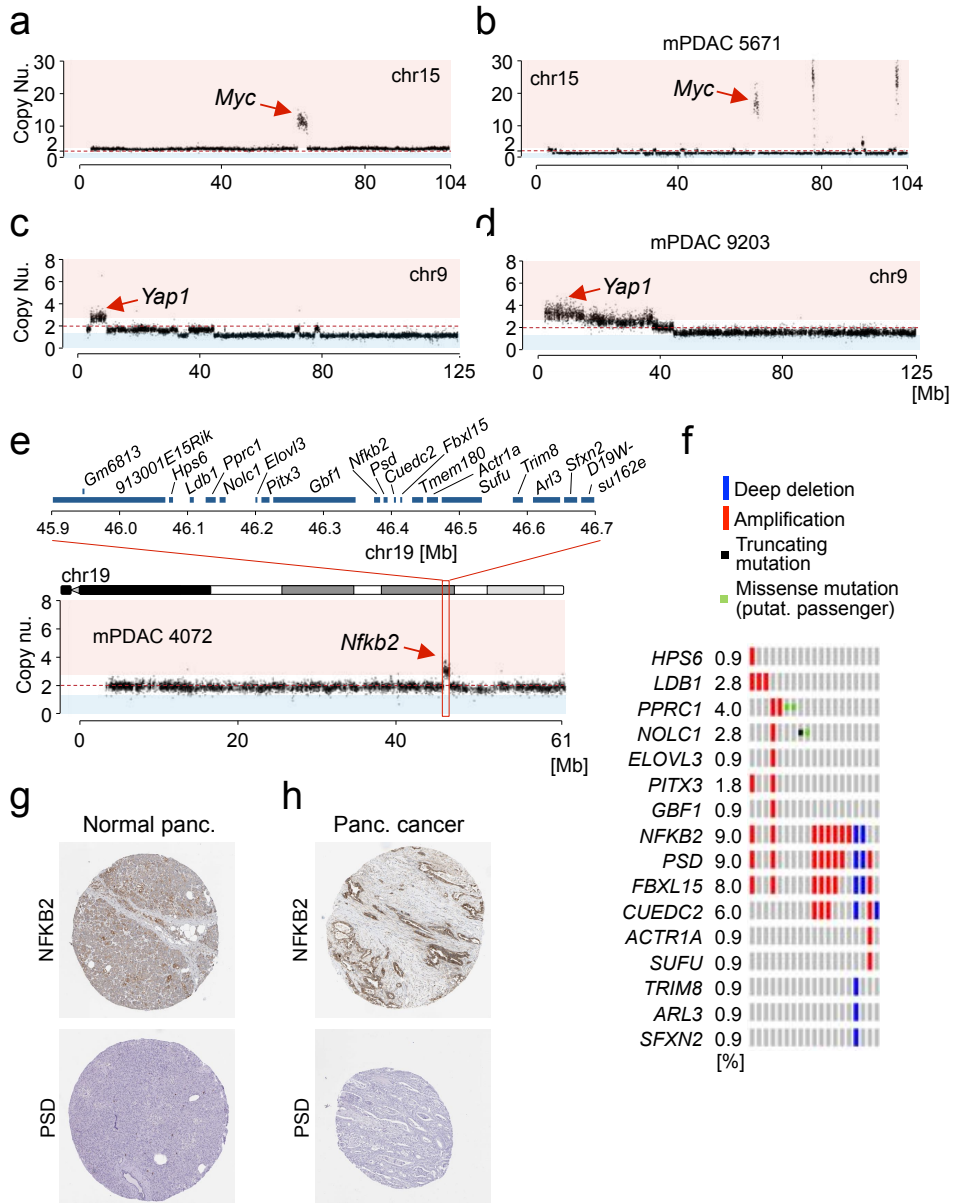
Extended Data Figure 2



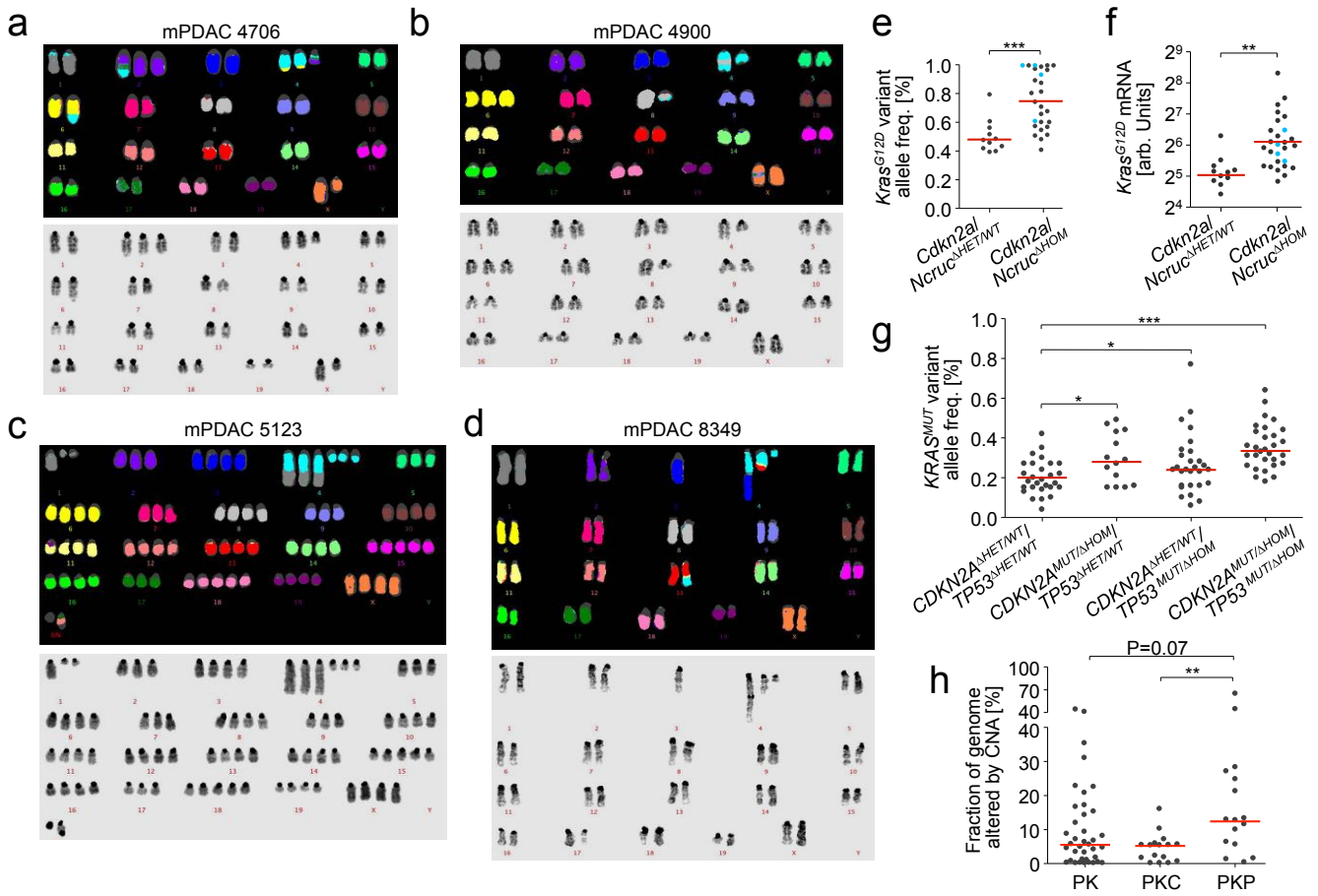
Extended Data Figure 3



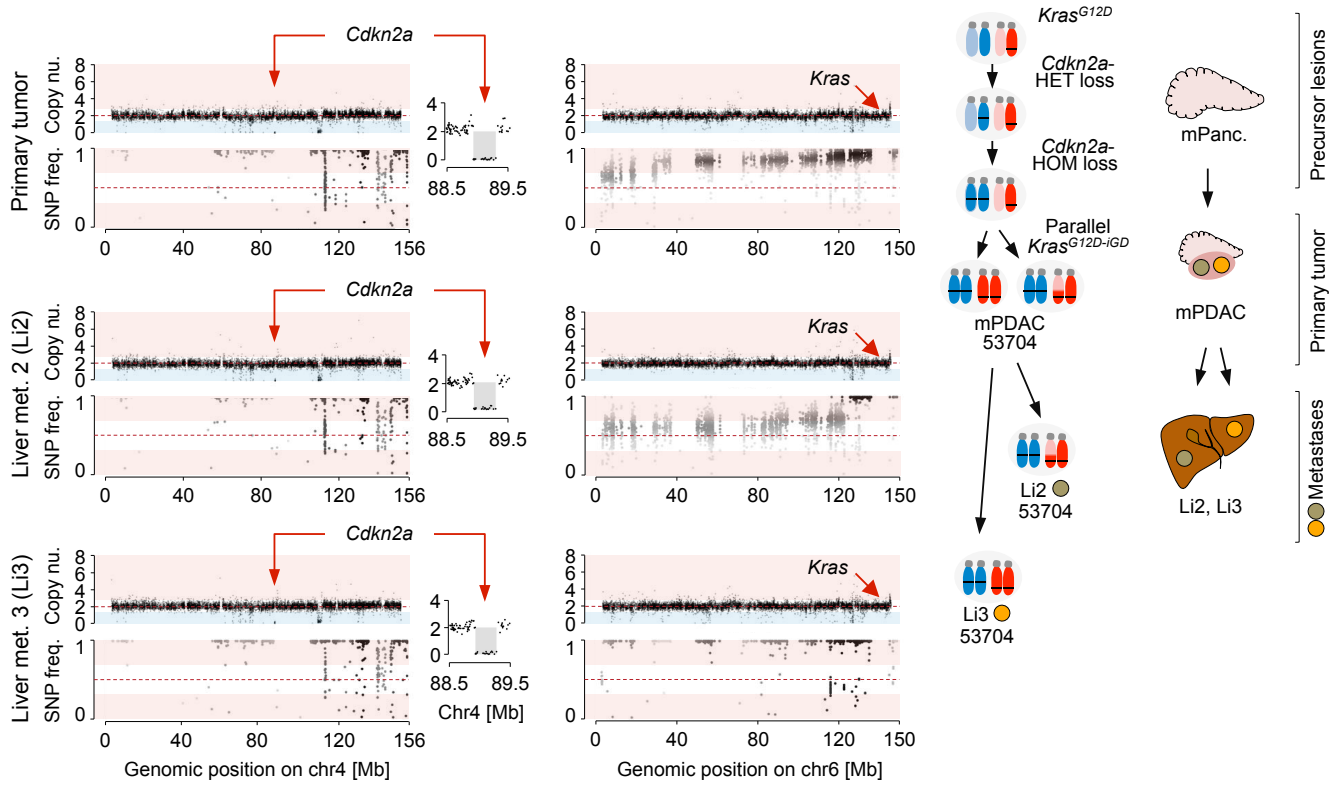
Extended Data Figure 4



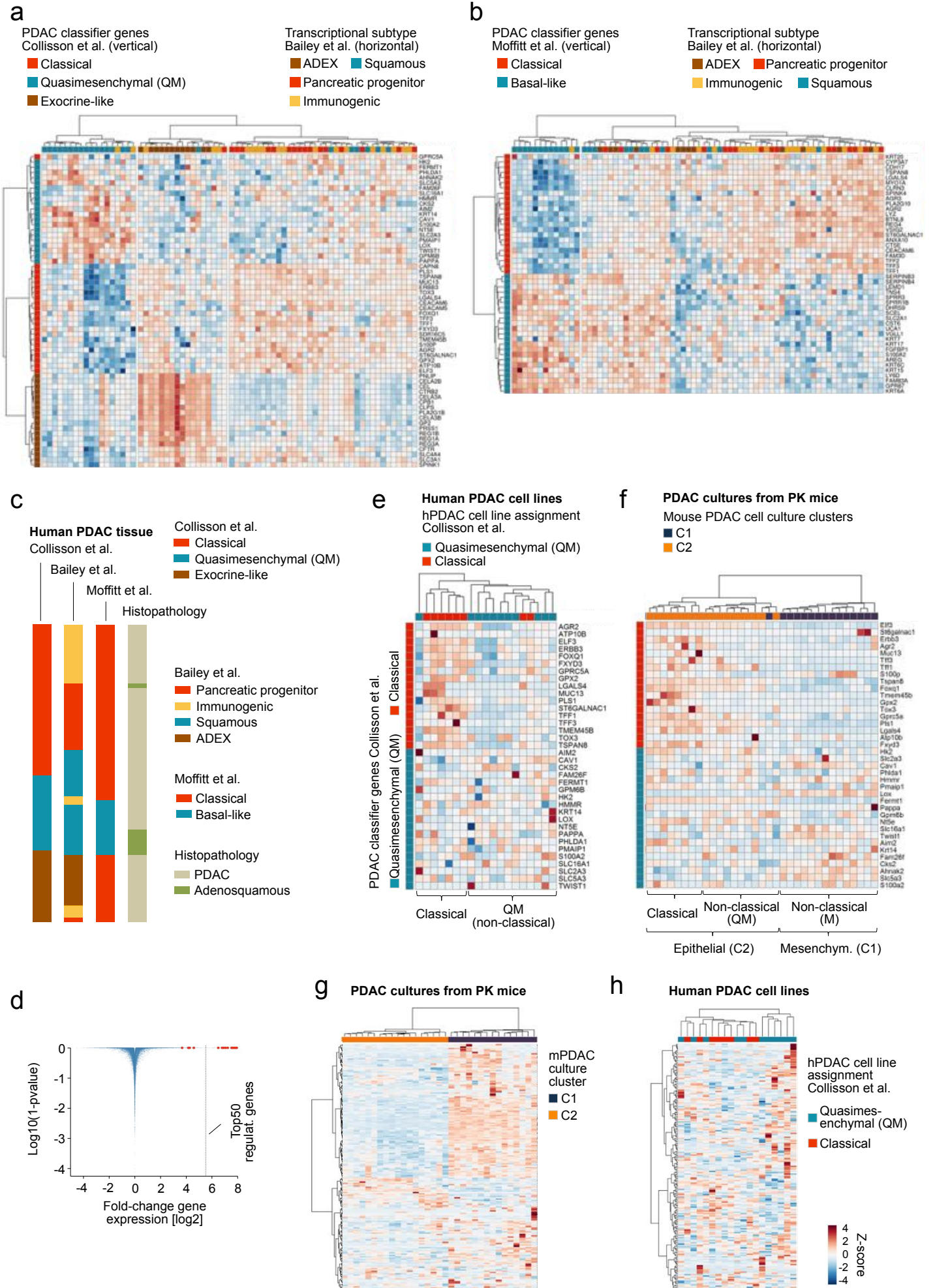
Extended Data Figure 5



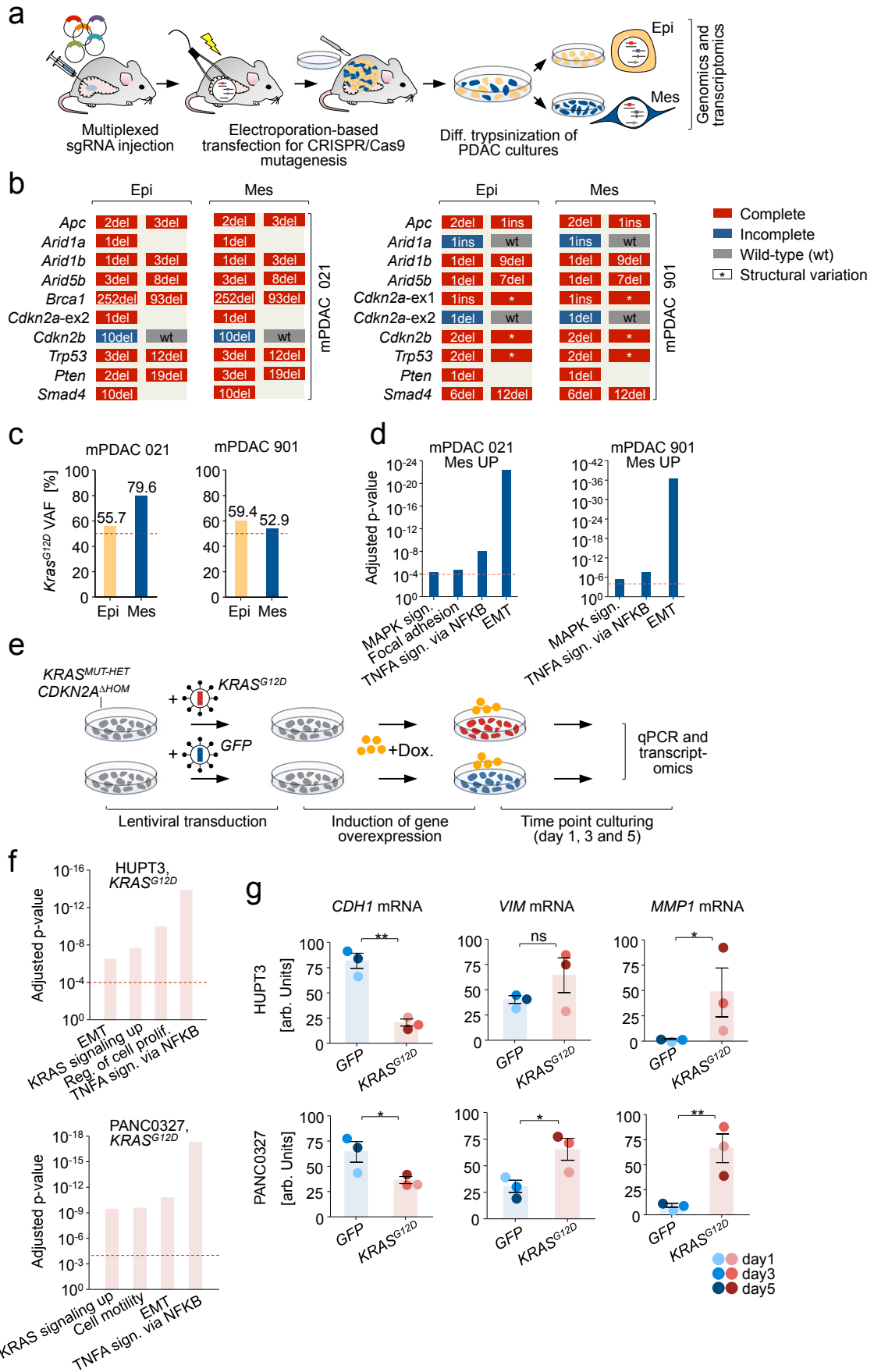
Extended Data Figure 6



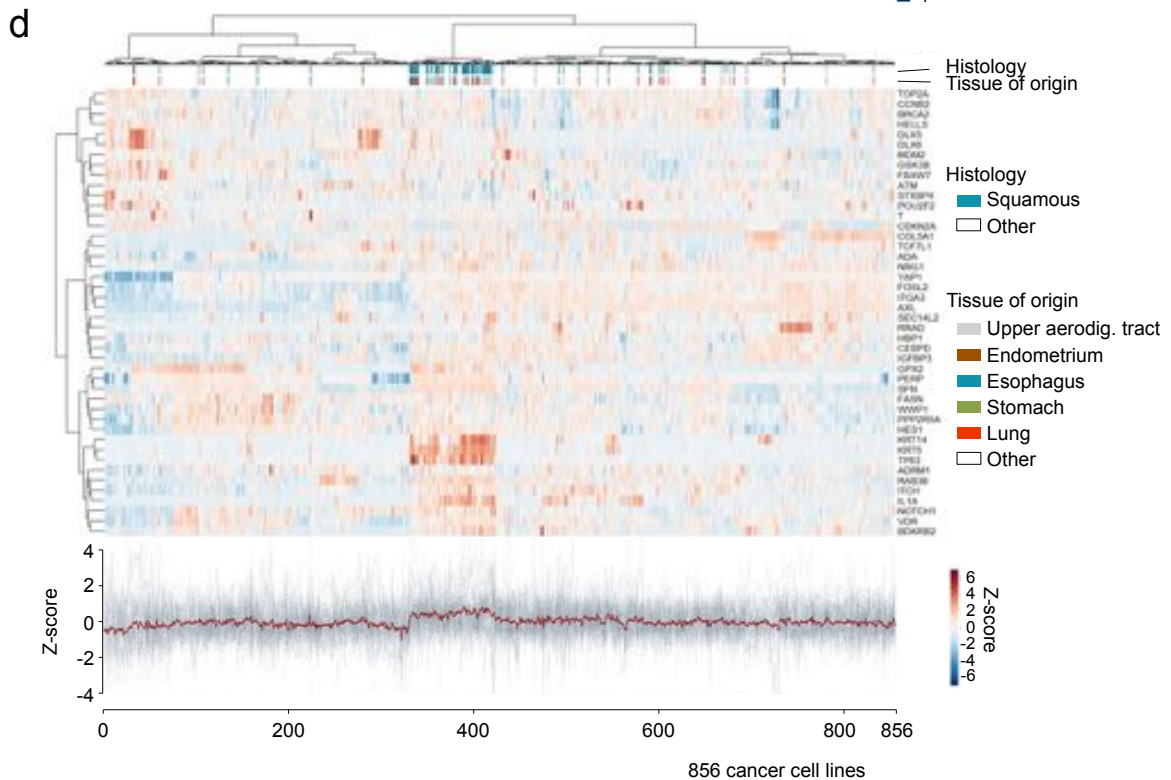
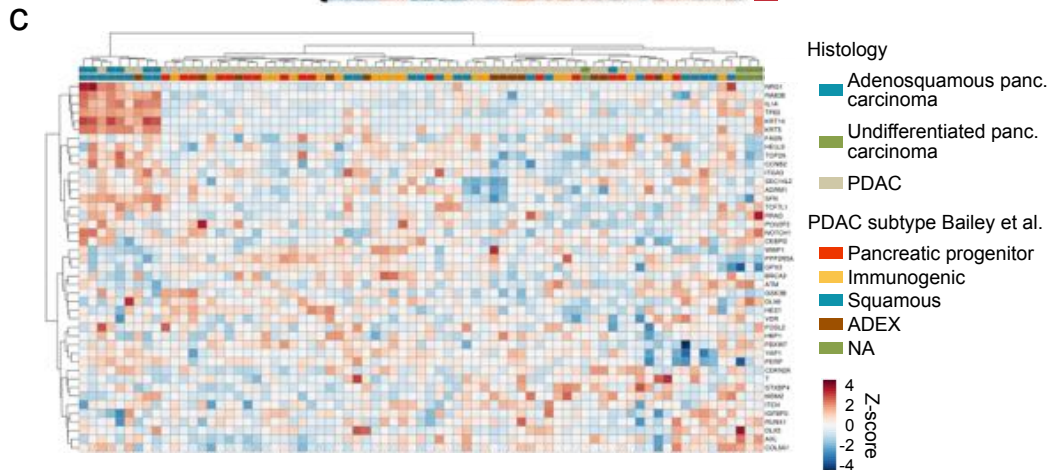
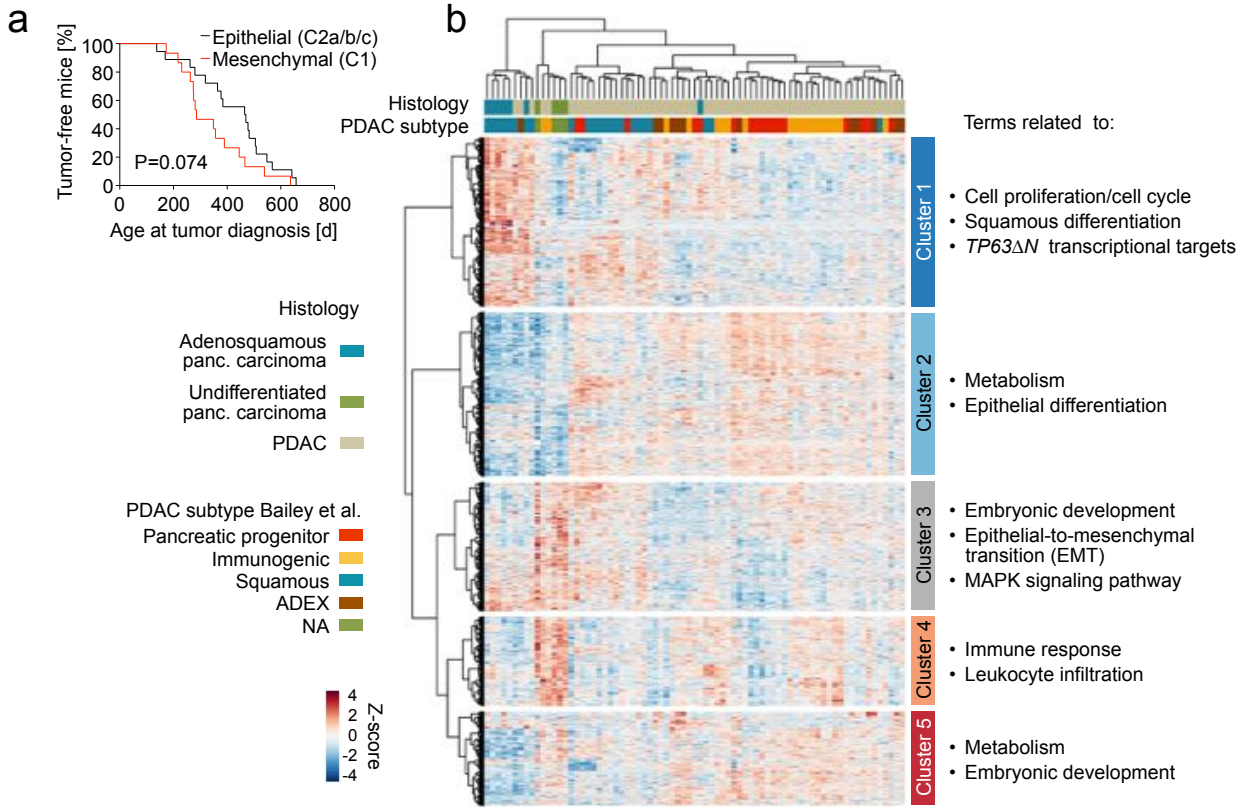
Extended Data Figure 7



Extended Data Figure 8



Extended Data Figure 9



Extended Data Figure 10

