Non-coding mutations at enhancer clusters contribute to pancreatic ductal adenocarcinoma

Minal B. Patel[1], Eleni Maniati[1], Santosh S. Atanur[2], Debosree Pal[3], Ana Rio-Machin[1], James Heward[1], Hemant M. Kocher[4], Jude Fitzgibbon[1], Madapura M. Pradeepa[3] & Jun Wang[1]

1. Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK, EC1M 6BQ
2. Section of Genetics and Genomics, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK
3. Centre for Genomics and Child Health, Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK
4. Centre for Tumour Biology, Barts Cancer Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK, EC1M 6BQ

[*] Corresponding authors
Jun Wang: j.a.wang@qmul.ac.uk
Madapura M Pradeepa: p.m.madapura@qmul.ac.uk

Abstract

Non-coding mutations (NCMs) that perturb the function of *cis*-regulatory elements (CRE, enhancers) contribute to cancer. Due to the vast search space, mutation abundance and indirect activity of non-coding sequences, it is challenging to identify which somatic NCMs are contributing to tumour development and progression. Here, we focus our investigation on the somatic NCMs that are enriched at enhancers from 659 pancreatic ductal adenocarcinoma (PDAC) tumours. We identify *cis*-regulatory NCMs within PDAC-specific enhancers derived from high and low-grade PDAC cell lines and patient derived organoids using two independent computational approaches. Five such CREs enriched for PDAC associated NCMs are also frequently mutated in other common solid tumours. Functional validation using STARR-seq reporter assays enables the prioritisation of 43 NCMs (7.3%) from a pool of 587 NCMs with 6,082 oligos, that significantly alter reporter enhancer activity compared to wild-type sequences. CRISPRi perturbation of an enhancer cluster harbouring NCMs over long non-coding RNA gene *MIR100HG*, which hosts a microRNA cluster (mir100-let7a-2-125b-1), leads to the downregulation of MIR100HG accompanied by a significant reduction in the TGF-β pathway (known to induce *MIR100HG*) and other PDAC critical pathways, including KRAS, p53, MTOR and TNFα signalling. Collectively, we have reported here *cis*-regulatory NCMs in PDAC proximal to many cancer-relevant genes, and our integrated approach paves way to explore CRE-associated NCMs in other human cancer genomes.

Introduction

Pancreatic cancer, ranking fourth in the cause of cancer death in developed countries, is an aggressive malignancy with a devastating five-year survival rate below 9% after diagnosis[1]. Pancreatic ductal adenocarcinoma (PDAC) is the predominant form of pancreatic cancer, encompassing approximately 90% of all cases[1]. Our understanding of the genomic landscape of PDAC is still mainly restricted to the somatic mutations within the coding regions of genes involved in PDAC[2-6]. Our knowledge of non-coding mutations (NCMs) and their functional consequences in the development and progression of PDAC is still limited. The availability of large-scale whole genome sequencing (WGS) projects, such as those by the International Cancer Genome Consortium (ICGC)[7], along with assays profiling chromatin modifications, accessibility and conformation, has allowed for a systematic search for functional NCMs in various cancer types[8-14].

Recent large-scale sequencing efforts by the Pan-Cancer Analysis of Whole Genomes (PCAWG) in over 2,600 primary tumours have identified several novel non-coding driver candidates, including NCMs in the 5' region of *TP53* and 3'UTR of *NFKBIZ* and *TOB1* using a statistically rigorous strategy for combining significance levels from multiple methods of driver discovery[14]. More recently, Dietlein et al., implemented a genome-wide, sliding-window approach to detect significantly recurrent mutated regions across the whole genomes of 3,949 patients and 19 cancer types, considering chromatin features, tissue specificity and background mutations. Using this approach, they identified NCMs in CREs near canonical cancer genes and tissue-specific genes, such as regulatory regions proximal to *HIST1H1B* and *TMEM151A* in PDAC genomes and pancreas tissue-specific genes *CPB1* and *PNLIP*[15]. Previously, Feigin et al., performed the PDAC-specific promoter-centric analysis and described Genomic Enrichment Computational Clustering Operation (GECCO) to uncover recurrent regulatory mutations in the *cis*-regulatory regions of 308 patient genomes. This method identified 16 genes with significant NCMs associated with promoter regions, and these genes were enriched for canonical PDAC pathways such as cell adhesion, axon guidance and Wnt signalling[16]. However, previous methods have not fully or effectively utilised PDAC-specific epigenomic data in the discovery analysis, particularly active enhancer regions, leaving a large number of putative gene regulatory NCMs unexplored and PDAC-specific enhancer drivers unidentified. Such active enhancer-centric methods have previously been implemented in T-cell acute lymphoblastic leukaemia using sequencing

reads derived from chromatin immunoprecipitation followed by sequencing (ChIP-seq) of histone H3 lysine 27 (H3K27ac) acetylation to identify recurrent enhancer associated variants[17]. Focusing on enhancer regions significantly reduces the non-coding genome search space to regions where non-coding variants are most likely to have potential functional activity at the gene control level[12,17,18].

To address these specific challenges, we integrated epigenomic datasets for histone modifications associated with enhancers to identify PDAC-specific active enhancers and promoter regions. Together with gene expression profiles (GEP) where available and simple somatic mutation (SSM) data of 659 PDAC patients from ICGC, to investigate NCMs associated with PDAC-specific *cis*-regulatory elements (CRE). We implemented a composite of two independent approaches to detect putative CRE drivers enriched for significant NCMs. We further tested the regulatory activity of NCMs within these CREs using the high-throughput functional screening approach STARR-seq, followed by the analysis of one enhancer cluster using CRISPR-interference (CRISPRi) of CREs with NCMs (Fig. 1a). Our study combines a systematic computational analysis and experimental validation, identifying important CRE drivers involving PDAC-relevant genes. It also demonstrates a versatile workflow to investigate CRE-associated NCMs in other disease genomes.

Results

The mutational burden within PDAC *cis*-regulatory regions

To identify likely pathogenic NCMs in PDAC, we retrieved SSM data from the ICGC Pancreatic Cancer Genome Project Australia (AU, n=391 patients) and the Canada (CA, n=268 patients) cohort. 1,379,638 and 2,211,000 somatic mutations were identified in the AU and CA cohort, respectively. After filtering out non-synonymous somatic mutations, 1,358,342 (98.5%) and 2,179,517 (98.6%) somatic NCMs were retained from the AU and CA cohort, respectively, for further analysis. This corresponds to an average of 3,701 (AU) and 8,132 (CA) NCMs per patient.

We wanted to focus on the NCM burden within CREs, specifically those enriched with H3K27ac, a chromatin feature associated with active enhancers[19]. We hypothesised that NCMs within these CREs may contribute to altering their function and target gene expression[17,20]. Using ChIP-seq datasets from seven PDAC cell lines[21] and two patient-

derived organoid samples[22], we identified 404,415 enriched H3K27ac peaks across all samples. To consolidate H3K27ac peaks across the nine samples into one representative consensus region per loci, we stitched together quality peaks residing within 2,000bp of another (inter-peak distance), resulting in a total of 65,168 H3K27ac consensus peaks for further analysis (average peak length = 4,639 and SD = 7,949). This allowed us to narrow the search space for potentially important NCMs to ~10% of the genome. Patient somatic NCMs were then mapped to the consensus H3K27ac coordinates to obtain a list of NCMs in PDAC-specific CREs. From the AU cohort of patients, 101,209 somatic mutations were observed within 36,409 (55.9%) consensus peaks and 166,541 CA cohort mutations within 43,002 (66.0%) consensus peaks. Therefore, capturing 7.45% and 7.64% of all AU and CA NCMs, respectively (Fig. 1b).

Prioritisation of *cis*-regulatory regions enriched with putative functional NCMs

We next aimed to interrogate somatic NCMs residing within consensus H3K27ac marked regions. To ensure the study of a significant proportion of PDAC patients, we retained CREs with a patient mutation incidence of 2% or above (n≥8), leaving 30,826 somatic mutations (AU cohort) across 1,711 consensus peaks/CREs and 64,867 somatic mutations (CA cohort) across 3,964 peaks (Fig. 1b). In total, 2.26% (AU) and 2.97% (CA) of the NCM burden remained to interrogate, similarly to observation in a previous study focusing on H3K27ac enriched elements[18].

To prioritise the remaining CREs, we utilised two independent approaches: one measuring the functional effect of each NCM within a CRE and ranking them based on the median functional score of all NCMs; the other identifying CREs with significantly recurrent NCMs accounting for local background mutation rate, and replication timing (Fig. 1a). We carried out the first approach using the IW-scoring algorithm[23], an integrative weighted scoring framework to score NCMs and prioritised elements with a median IW-score of two or above (corresponding to a p-value ≤ 0.1). From the remaining 1,711 (AU) and 3,964 (CA) peaks after filtering, we identified 14 CREs from the AU-cohort and 32 elements in the CA-cohort using the median threshold (Extended Data Fig. 1 and S1). This method prioritised CREs annotated to cancer-related genes such as the *AP-1* transcription factor (TF) *JUNB* expressed in low-grade PDAC cells[21,24], and *GATA2,* associated with high-grade PDAC[21]. Of the 46 prioritised CREs, five regions were shared between the AU and CA cohorts (Fig. 1c).

These five CREs reside within the introns of the oncogenic long non-coding RNAs (lnRNA) *MIR100HG*[25] and *HOTAIR*[26] from and including the *HOXC* cluster of homeobox genes[27], PDAC associated TFs *FOXA1*[28] and *FOXP1*[29] and ferroptosis related TF, *NFE2L1*[30] (Fig. 1c).

To further validate the putative significance of the NCMs within these five CREs, we compared the IW-score of NCMs residing within the H3K27ac positive regions to NCMs in immediate flanking sequences negative for H3K27ac marks. We observed a statistically significant higher IW-score of NCMs within H3K27ac enriched regions compared to those in flanking H3K27ac negative sequences (Extended Data S2), indicating the putative enhancer-associated NCMs have higher predicted functional consequences than mutations located outside these CREs. We also verified these findings with an independent scoring algorithm LINSIGHT, which scores variants on the likelihood of deleterious fitness consequences based on patterns of polymorphism and divergence from closely related species[31]. The LINSIGHT model demonstrated a significant increase in the selective constraint (i.e., more deleterious on fitness) of H3K27ac-associated NCMs compared to NCMs in nearby H3K27ac negative regions (Extended Data Fig. 2).

Using the second approach to identify significantly recurrently mutated CREs, we implemented LARVA[32]. The LARVA model yielded 68 (AU cohort) and 71 (CA cohort) candidate CREs which were significantly recurrently mutated in relation to nearby background sequences (Benjamini-Hochberg (BH) adjusted $p \leq 0.01$). These significant regions collectively harboured 1,842 and 2,258 NCMs in the AU and CA cohorts. Many NCMs were located proximally to several well-known genes implicated in PDAC, for example, an intergenic regulatory region in proximity to the miRNA: miR-21 and the Wnt/β-catenin signalling protein gene *WNT7b* (Extended Data S3). Nine significantly mutated CREs were shared between AU and CA cohorts. These recurrent CREs included regions proximal to the TF genes *TBX3* and *BNC1*, previously reported in PDAC[33,34]. NCMs were also located proximal to the adhesion molecule *PXDN*[35] and transmembrane protein TENM3[36], the lncRNA gene *TBX5-AS1*[37], and microRNA, miR-1305 (Fig. 1c).

Notably, the *MIR100HG* enhancer cluster was the only one prioritised in the two approaches, but consisting of two separate CREs (Fig. 1c, Extended Data Fig. 3). Overall, our computational strategy has revealed NCMs enriched within or proximal to PDAC or cancer-related genes, including candidates identified from a previous non-coding study in PDAC[16].

Proximal genes to enhancer NCMs are associated with transcription and PDAC-linked biological processes

We next performed pathway enrichment analysis based on the annotated genes proximal to CREs identified by the two *in-silico* approaches using the DAVID tool[38]. Inputting 95 annotated genes associated with 41 CREs identified by the IW-scoring approach, we observed significant enrichment in several gene families and regulatory processes, including homeobox genes, pattern specification, embryogenesis and transcriptional regulation pathways (Fig. 1d). Additional pathway analysis based on 212 genes annotated to the 130 recurrently mutated CREs identified significant enrichment in core molecular pathways including cell adhesion, epithelium development, cell proliferation, transcription, apoptotic processes and regulation of chemotaxis (Fig. 1d). The involvement of biological processes, such as embryogenesis, apoptosis and cell adhesion, has been reported in a previous genomic landscape study[39]. Furthermore, our findings complement Feign *et al.* in identifying NCMs significantly associated with homeobox genes and transcriptional regulation[16]. Our results suggest a convergent mode for CRE-associated NCMs in relation to biologically relevant coding genes in PDAC.

Enhancer NCMs show altered transcriptional reporter activity

To determine the effect of NCMs on the transcriptional regulatory activity, we performed luciferase-based enhancer reporter assays for a subset of NCMs. We selected twelve NCMs from two CREs identified from the first approach (IW score), comprising 11 single nucleotide variants (SNV) and a single 4bp deletion. Five SNVs were selected from the third intron of the *FOXP1* gene, and seven NCMs in the third intron of the lncRNA *MIR100HG* (Fig. 1e). Interestingly, the 2kb region surrounding the seven NCMs at the *MIR100HG* locus lack detectable H3K27ac and H3K4me1 marks in most of the cell lines, except those derived from high-grade PDAC cells PANC-1 and PT45P1, suggesting this putative active enhancer is specific to high-grade PDAC (Fig. 1e, Extended Data Fig. 3). Luciferase reporter assays were carried out in the high-grade PDAC cell line PANC-1 and easily transfectable cell line HEK293T. Within the *MIR100HG* CRE, NCMs (MUT 3, 6 and 7) and (MUT 1-3) showed significant increases in reporter activity in HEK293T and PANC-1 cells, respectively (Fig. 1f). Overall, all NCMs at this *MIR100HG* CRE showed an increase in luciferase activity compared to WT sequences, suggesting NCMs within this CRE are potentially gain-of-function, i.e.,

increase regulatory activity. The ~2kb regulatory element surrounding five selected NCMs within the third intron of *FOXP1* was positive for H3K27ac marks in six PDAC cell lines (except for MIA-PaCa2 cells), and two patient-derived organoid samples (Extended Data Fig. 4). Among the five NCMs tested, two NCMs in HEK293T cells and three in PANC-1 cells significantly altered luciferase expression. Most notably, mutation 3 (chr3:71104908:C>T, IW-score = 5.20, p = 0.006, LINSIGHT score = 97.2%) significantly decreased reporter gene expression in both cell lines (Fig. 1f). Interestingly, all five NCMs within the *FOXP1* putative enhancer demonstrated concordance in the overall transcriptional regulatory activity in both cell lines.

STARR-seq assays prioritise a subset of 43 NCM candidates for further validation

Next, we screened a larger set of NCMs within consensus CREs using the high-throughput approach, Self-Transcribing Active Regulatory Region sequencing (STARR-seq)[40]. To focus on NCMs with the strongest evidence of predicted function, we retained 504 NCMs with a variant allele frequency above 20% and strong TF binding strength as predicted by motifbreakR[41]. Of the 504 NCMs, binding motifs of 258 TFs were strongly predicted to occupy these mutation sites. Moreover, among the 73 NCMs identified by the first approach (IW score), 47 (64%) NCMs were predicted to cause TF-motif gain and 26 (36%) loss-of-motif (break). Among the 431 NCMs selected from the second approach (LARVA), 216 (50%) NCMs caused predictive gain and 215 (50%) loss of motif changes. We included 83 single base indels, resulting in 587 candidate NCMs in the final STARR-seq library (Fig. 2a).

We designed ten 230bp oligos per NCM, five for each NCM and five for the corresponding wild type (WT). One oligo represented the NCM in the middle and four oligos had a 10 bp sliding genomic window (SW) in either direction from the centre of the oligo (Fig. 2a, see Methods). A further 210 positive (PDAC enhancers) and negative (no enhancer features) control oligos were included in the library, resulting in a pool of 6,082 oligos. Sequencing and quality analysis of the cloned STARR-seq plasmid library demonstrated good complexity and accuracy (Extended data Fig. 5), with comparable outcomes to a previous MPRA study using synthetically designed oligos[42,43]. We performed two biological replicates of STARR-seq by transfecting the PANC-1 cell line (see methods)[44]. After filtering low-quality reads across samples, we observed a good concordance between replicates (Fig. 2b). As

expected, positive control sequences showed significantly higher reporter activity compared to negative controls (Fig. 2c).

We next tested the significance between mutant (MUT) and WT constructs on reporter gene expression across replicates. A total of 217 plasmids (representing 155 NCMs) showed significant differential enhancer activity (log_2 fold change -1.54 to 3.53, Student's t-test, p<0.05). 95 (61.3%) NMCs showed significantly increased enhancer activity, while 60 (38.7%) mutations showed a significant reduction in enhancer activity in comparison to WT sequences (Fig. 2c and 2d). Interestingly, 36 CREs harbouring indels showed significant fold changes at similar activity to SNVs (mean log_2 FC 1.07). Despite the differences in assays and genomic context, we observed concurrent directional changes in enhancer activity at NCMs assayed by luciferase reporter assays and sequencing-based high-throughput STARR-seq (Extended data Fig. 6).

Focusing on the most significant alterations between MUT and WT alleles (t-test, p<0.01), we highlighted 43 mutations, 33 of which demonstrated an increase in reporter activity and 10 with an observed reduction (Fig. 2e). Notably, the differential activity changes between MUT and WT in 13 NCMs were significantly altered in three or more independent STARRs-seq plasmids (p<0.05). Similarly, 31 NCMs were significantly altered in two independent SWs demonstrating concurrent directional activity changes. Eight of the 43 NCMs were located within an enhancer cluster (observed in low-grade and MiaPaCa2 cells) upstream of the *BNC1* gene (Extended data Fig. 7a). The NCMs proximal to *BNC1* significantly increased reporter gene expression in PANC-1 cells in comparison to WT sequences (Fig. 2d). Assessing the expression of genes within 1Mb of this consensus peak by comparing MUT and WT patient GEPs, we did not observe a difference in the expression of BNC1, previously reported to be methylated in early stage PDAC patients[45]. However, we observed a significant increase in the expression of nearby genes *BTBD1* (p = 0.003, ~234kb from the middle of the consensus peak to BTBD1 TSS), important in cell survival, the ubiquitin/proteosome degradation pathway and mesenchymal differentiation[46] and *FAM103A1* (p = 0.008, ~316kb) which encodes an important subunit for the 7-methylguanosine cap added to the 5' end of mRNA and an essential component for gene expression[47]. Patient GEP analysis also revealed a significant decrease in the transmembrane protein *TM6SF1* (frequently hypermethylated[48,49], p = 0.0003, ~137kb) between MUT and WT patients (AU cohort, Extended data Fig. 7b), overall suggesting these

NCMs may exert their regulatory potential in a more distal manner. Additional significant increases in reporter gene expression were observed proximal to the PDAC-associated TF *TBX*[50] (7 NCMs) and in the introns of lncRNA *MIR100HG* [25] (3 NCMS, Fig. 2d).

To assess the putative biological implications of these top-performing STARR-seq NCMs, we took a closer look at the TF-motif binding predictions. From the 35 NCMs in the top 43 STARR-seq performing mutations with TF-motif predictions, 21 were characterised as TF binding motif-gain (creating *de novo* TF binding motifs), while 14 were TF binding motif-loss. For example, one gain-of-function NCM proximal to *TBX3* (chr12:115067012:C>A) was predicted to create a binding motif for the oncogenic TF JUN (Fig. 2d and Extended Data Fig. 7c). This NCM led to a mean $\log_2$ fold-change of 3.69 in STARR-seq reporter gene expression across all five SWs (Mann Whitney U test, p=0.016). As expected, JUN was highly expressed in PDAC patients based on the patient GEP in the AU cohort (Fig. 2f)[51]. The most significant loss-of-function was observed in a NCM located in the intron of *FOXP1* (chr3:71123616:G>T) supported by three significant SWs (p <0.05, average log2 fold change across SWs = -1.36). At this site, the binding motif of an unfolded protein response (UPR) mediating TF, the activating TF-3 (ATF3)[52], was predicted to be disrupted (Extended Data Fig. 7d) and was found to be moderately expressed in the PDAC patient GEP (AU cohort, Fig. 2f). Furthermore, the top two NCMs located in the *MIR100HG* enhancer cluster also showed strong effects on TF binding: the first mutation (chr11:122010557:C>T) demonstrated a gain of TF motif, creating a *de novo* binding motif for *NR6A1*, a nuclear receptor family member; while the second mutation (chr11:122025440:G>C) was predicted to disrupt the binding motif for *SOX10* (Fig. 2g), a reported tumour suppressor through the suppression of the Wnt/β-catenin pathway in digestive cancers[53]. We observed that *NR6A1* and *SOX10* TFs were expressed at moderate levels in PDAC patients (AU cohort, Fig. 2f). Overall, using the STARR-seq assay enabled the prioritisation of CRE-associated NCMs for further investigation.

CRE cluster harbouring NCMs located at the *MIR100HG* locus regulates genes *in cis*

The two computational approaches used in this study identified the lncRNA MIR100HG locus as a significant candidate for harbouring NCMs in separate CREs in each approach. Notably, *MIR100HG* is host to the oncogenic miR-s pre-miR125b-1 and pre-miR-100, previously implicated in PDAC[25,54], and they modulate (including *MIR100HG*) in a pro or anti-

tumourigenic manner depending on the cancer[25,54-59]. It hosts the tumour suppressors pre-miR-Let7a-2[25] and the pro-apoptotic protein *BLID*[60], located within intron three of MIR100HG (Fig. 3a).

Next, we investigated the functionality of three CREs harbouring NCMs at the *MIR100HG* enhancer cluster using a CRISPRi approach recruiting the dCAS9/KRAB repressor to NCMs and CREs of interest[61] (Fig. 3a and b). The first region located ~2kb away from the hosted pre-miR-125b-1 in the third intron of *MIR100HG* harboured NCMs identified from the first *in silico* approach. CRISPRi with a pool of four independent lentiviral guide RNAs (G 1-4) were selected close to NCMs that were shown to alter enhancer activity in either luciferase or STARR-seq experiments (Fig. 1e, 3b). Two guides (G 5-6) were designed to target region two harbouring five NCMs (CRE-two), including the most significant NCM identified to drive reporter enhancer activity using STARR-seq (M20 in Fig. 2e and Fig. 3b). An additional two guides (G 7-8) were designed to target the third region harbouring six NCMs (CRE-three), including a gain-of-function NCM from the most significant STARR-seq candidates (M37 in Fig. 2e).

CRISPRi, followed by RT-qPCR, showed a significant reduction in MIR100HG expression in all three CREs in this enhancer cluster in comparison to dCAS9/KRAB negative controls (Fig. 3b). This data suggests that these CREs function as active enhancers to regulate the expression of *MIR100HG*. Analysis of looping interactions from the 4D genome[62] and integrated method for predicting enhancer targets (IM-PET)[63] data in PANC cells indicated interactions between CRE-two and the promoter of *UBASH3B* located upstream of *MIR100HG* (Extended Data Fig. 8a). *UBASH3B* has been reported to inhibit the endocytosis of the epidermal growth factor (EGFR), an essential component in the development of pancreatic precursor lesions[64-66]. RT-qPCR analysis demonstrated a significant decrease in *UBASH3B* expression with the CRE-two CRISPRi compared to controls (Fig. 3b). CRE-three shows interactions with the promoter of *ARHGEF12* (Extended Data Fig. 8b). *ARHGEF12,* a guanine nucleotide exchange factor (GEF), activates Rho A, a key regulator of cytoskeleton organisation and ROCK1/2 induced extracellular matrix remodelling, associated with poor outcomes in PDAC patients[67]. CRE-three CRISPRi resulted in a significant decrease in *ARHGEF12* levels compared to controls (Fig. 3b). These results suggest that CRISPRi-based perturbation of CRE-two and three leads to downregulation of

353 genes located *in cis*, although to a less extent compared to the reduction in *MIR100HG*
354 expression.
355
356 **CRISPRi perturbation of MIR100HG CREs alters core PDAC signalling pathways and cell**
357 **motility.**
358 We performed RNA-seq to evaluate the global mRNA changes in CRISPRi-targeted CRE-two
359 and -three clones (Fig. 3b). Principle component and correlation analyses showed CRISPRi of
360 CRE-two and -three shared similar gene expression programmes (Fig. 3c and 3d). Differential
361 expression (DE) analysis identified 98 and 102 significant genes in the perturbation of CRE-
362 two and -three clones compared to the control, respectively (FDR<0.05 and absolute $\log_2$ FC
363 >1). Of them, 59 DE genes were shared between the two clones (Fig. 3e). We also observed
364 a significant reduction in *MIR100HG* RNA-seq expression in both targeted CREs, consistent
365 with the qPCR data (Fig. 3f).
366 Gene set enrichment analysis (GSEA)[68] against the MsigDB Hallmark[69] and oncogenic
367 signature gene sets were then performed between the two CRISPRi groups and the dCas9-
368 KRAB control (Fig. 4a). In both CRISPRi perturbations; we observed a comparable and
369 significant downregulation of important PDAC hallmark gene sets involved in KRAS
370 signalling[70], UPR, reactive oxygen species (ROS)[71] and TNFα signalling[72] (Fig. 4a and 4b).
371 Oncogenic signatures associated with critical drivers KRAS[73], P53, epithelial-to-mesenchymal
372 transition (EMT) inducing TGF-β and cell survival and proliferation-related MTOR[73] pathway
373 genes were significantly reduced in both inhibited *cis*-regions. In contrast, migration
374 inhibiting cAMP[74] and interestingly pro-EMT related LEF1[75] signatures were significantly
375 upregulated (Fig. 4a). Collectively, the CRISPRi perturbation of two CREs at MIR100HG led to
376 a significant reduction in key oncogenic molecular mechanisms observed in PDAC, resulting
377 in a more favourable phenotype.
378 TGF-β regulates *MIR100HG* transcription and thus the release of its hosted miRs,
379 inducing EMT, encouraging cell motility and metastasis[25]. Here, we identified many TGF-β
380 related genes such as *FGF1*[76], *KDM6B*[77], *LIF*[78], *PIK3CD*[79], *PXDC1* and *TAGLN*[80] were
381 significantly downregulated in the two CRISPRi groups compared to the control (Extended
382 Data Fig. 9). Hence, we further aimed to validate the reduction in TGF-β signalling observed
383 with GSEA enrichment by using wound healing assays (Fig. 4c). Over 48-hours, the inhibition
384 of CRE-two (G 5-6) resulted in a significant reduction in cell motility in comparison to

385 controls, corroborating with a stronger gene enrichment reduction in TGF-β and EMT
386 signalling compared to CRE-three inhibition (Fig 4a). Similar but not significant changes in
387 cell motility were observed in PANC-1 cells inhibited at CRE-three (G 7-8) (Fig. 4d). These
388 results suggest a CRISPRi perturbation of CREs harbouring NCMs in the third intron of
389 *MIR100HG* can decrease the migration ability in PANC-1 cells.
390
391 **Mutation occurrence of functional CREs in other solid cancers**
392 Lastly, we explored the NCM burden of our top five prioritised regions (obtained from the
393 first approach) in other cancers. We analysed the mutational frequency of these CRE-
394 associated loci in seven other solid tumours using SSM data from the ICGC in oesophageal
395 (ESAD), liver (LIHC), breast (BRCA-UK), ovarian (OV), prostate (PRAD-CA, PRAD-UK),
396 colorectal (COAD) and gastric cancer (STAD) cohorts. The *HOTAIR*/*HOXC* CRE had the highest
397 mutation frequency of NCMs across oesophageal (16.6%), liver (13.2%), prostate (7.5%) and
398 ovarian (19.4%) cancers along with PDAC (5-12%, Extended data Fig. 10). However, a low
399 mutation frequency was observed in gastric, breast and colorectal cancers below 2%. The
400 *FOXA1* CRE was predominately mutated in prostate cancer at an incidence of ~16%,
401 followed by liver, ovarian and oesophageal cancers at a frequency of ~5%, higher than that
402 observed in PDAC (2%). Interestingly, this regulatory region and NCMs have been recently
403 reported in prostate cancer and are correlated with decreases in *FOXA1* expression and cell
404 growth[81]. For the *MIR100HG* CRE, oesophageal and prostate cancer (UK cohort) showed the
405 highest incidence at 14.2% and 5.7%, respectively, and liver and ovarian cancers showed a
406 similar mutational incidence to the PDAC cohorts (2-3%). Other cancer types, such as breast,
407 gastric and colorectal, had a very low to no mutational burden within this *MIR100HG* CRE
408 (Extended Data Fig. 10c). The *FOXP1* CRE had the highest mutation frequencies in the liver,
409 oesophageal and ovarian cancers (6-8%), but the *NFEL2* CRE generally had a much lower
410 mutation frequency across all cancers, with a mutation burden of 2-3% in liver and
411 oesophageal cancers, similar to that in PDAC. Our results suggest that several CREs
412 identified in this study were also frequently mutated in other cancers. NCMs within these
413 CREs may also play a functional role in contributing to these malignancies, as already
414 documented in prostate cancer[81].
415
416

Discussion

Our study combines a computational discovery strategy and experimental follow-up to assess the functional significance of NCMs associated with PDAC-specific CREs. We leverage NCMs from PDAC SSM data derived from the ICGC[7] and integrate with PDAC-specific CREs marked by H3K27ac in seven PDAC cell lines and two patient-derived organoid samples. Previous investigations have often relied on consensus regulatory regions defined by ENCODE cell lines or the Ensembl Regulatory Build[82], this is likely to miss many enhancers which regulate genes in a highly cell and tissue specific manner[83]. Our PDAC consensus peaks have incorporated high- and low-grade cell lines and patient derived organoids accounting for the tissue and stage specificity of regulatory elements associated with PDAC biology[21].

The non-coding genome comprises a diverse spectrum of elements, and the mutational patterns and consequences are highly heterogeneous, rendering one approach ineffective across all regions of the non-coding genome[84,85]. Thus, our pipeline incorporates an approach that directly estimates the functional consequence (i.e. deleteriousness) of each NCM and another that detects recurrently mutated CREs taking into consideration confounders such as replication timing and background mutation rates. Hence our combined approach identified a comprehensive, robust set of CREs subject to PDAC-relevant biological processes for *in vitro* validation.

High-throughput enhancer reporter assays are a powerful approach to screen the regulatory activity of a large number of NCMs in parallel[40,43,86,87]. Our STARR-seq data highlighted 43 NCMs from PDAC patients showing significant gene reporter activity in the PANC-1 cell line. Interestingly, we observed the largest number of NCMs upstream of the *BNC1* promoter, resulting in a significant increase in STARR-seq reporter gene expression (Fig. 2d). Assessing the GEP of patients with these NCMs compared to those without NCMs demonstrated significant expression changes in more distal genes *BTBD1*[46], *FAM103A1*[47] and *TM6SF1*[48]. These DE genes were also associated with poorer overall outcomes in PDAC patients with higher expression (in *BTBD1* and *FAM103A1* genes) and lower expression for *TM6SF1* expressing patients (data not shown). Additional interesting candidates, such as NCMs proximal to cancer and PDAC-related TF *TBX3*[34] and NCMs in the intron of *FOXP1*[88], would be interesting and relevant candidates for future studies.

We identified significant CREs harbouring NCMs at the *MIR100HG* introns using both computational approaches, highlighting its importance for further functional validation. Previously, the transcription of *MIR100HG* has been linked to TGF-β expression/induction through SMAD2/3 binding sites in PDAC cell lines and *in vivo* studies leading to the release of its hosted miRs, including the oncogenic miR-100 and miR-125b-1[25,54]. The CRISPRi-based perturbation of *cis*-regions harbouring the most significant NCMs in the third intron of MIR100HG (identified using luciferase or high-throughput STARR-seq experiments) led to a down-regulation of *MIR100HG* expression and, in turn, cell mobility (Fig. 4c). This was correlated with a significant downregulation in critical PDAC related pathways included KRAS, P53, TGF-β and TNFα signalling[72,73]. Although not tested here, the direct targeting of these *cis*-regions leading to a down-regulation of *MIR100HG* transcription may inhibit the release of its hosted oncogenic miRs, as previously reported[25,54].

Applying 4D genome interaction data[62], we observed looping of our targeted CRE-two with the promoter of proximal *EGFR-related* gene, *UBASH3B*[66] and CRE-three with the promoter of the RhoA regulating GEF protein *ARHGEF12*[67]. Using RT-qPCR, we demonstrated CRE-two had the ability to downregulate *UBASH3B* expression, and CRE-three inhibition led to the significant reduction of *ARHGEF12*. These putative interactions may contribute to the downregulation of core pathways revealed by the GSEA analysis, as seen by the downregulation of EGFR signatures upon CRE-two inhibition[66] and MYC-target downregulation with CRE-three inhibition[89]. This is the first report to our knowledge of NCMs in the introns of the lncRNA *MIR100HG* and the suggestion of *cis* genes other than MIR100HG being altered in expression[25,54]. Considering a large number of transcripts MIR100HG has, further assessment of these CREs and NCMs on splicing would be important.

Genetic changes are critical for PDAC initiation, and up until recently, with the clinically available $KRAS^{G12C}$ inhibitor (AMG 510)[90] and the preclinical development of the $KRAS^{G12D}$ inhibitor MRTX1133[91], core mutated genes are largely undruggable. The reversibility of epigenetic changes allows the opportunity for therapeutic targeting. Previously in prostate cancer cells, the silencing of MIR100HG has led to the sensitisation to cytotoxic drugs[54]. We have shown here that perturbation of MIR100HG-associated CREs has collectively led to the downregulation of multiple core signalling pathways, including those previously not implicated in MIR100HG disruption, such as KRAS and TNFa signalling[25,54]. In

addition to the above considerations of this study, further investigation into the therapeutic potential of targeting this enhancer cluster rich in CREs and NCMs would be the next step.

We have limited this study to active enhancers widely reported to be marked by H3K27ac and H3K4me1[19]. However, we observed NCMs located outside of PDAC-associated CREs to have high functional predictive scores, suggesting they may lead to a gain/loss in functional activity at the gene level (Extended Data Fig. 2b). Moreover, use of H3K27ac alone to predict active enhancers may be too simplistic as many enhancers are marked with H4K16ac and H3K122ac but lack H3K27ac[92,93], suggesting many more CRE associated NCMs may be missed here. We have demonstrated the enhancer function for the *MIR100HG* locus harbouring PDAC-specific NCMs. However, further work is needed to demonstrate the pathogenic role of other NCMs identified in PDAC. Overall, our work identified and validated functional CREs and associated NCMs that may contribute to PDAC tumourigenesis and we have demonstrated a systematic framework to study *cis*-regulatory mutations in other human diseases.

Methods

Data acquisition

Data from the International Cancer Genome Consortium were downloaded from the ICGC portal (https://dcc.icgc.org/) release 27[7]. This data included simple somatic mutation (SSM) data for pancreatic ductal adenocarcinoma samples from the PACA-AU and PACA-CA cohorts. Clinical data, array-based expression (EXP-A from the PACA-AU cohort) and sequencing-based gene expression data (EXP-S from the PACA-CA cohort) were also downloaded. Gene Expression Omnibus (GEO) acquired datasets GSE64560[21] and GSE99311[22] were used to obtain ChIP-seq data to identify active enhancer-associated regions of the genome (H3K27ac and H3K4me1) based on seven PDAC cell lines and two patient-derived organoid samples. Additional marks were used to annotate further putative promoters (H3K4me3) and repressive domains (H3K9me3, H3K27me3).

ICGC data processing

Downloaded SSMs were annotated and filtered using Annovar tools, retaining only those residing in non-coding elements (i.e., intergenic, intronic, synonymous and UTR)[94]. Annovar

'filter-based' annotation method with packages: hg19_avsnp147, hg19_snp138, hg19_cytoBand, hg19_dbnsfp30a, hg19_ensGeneMrna was used. Available raw array-based expression (EXP-A) data was retrieved for 269 out of 391 patients from the AU cohort and normalised. Raw RNA-seq data for 234 out of 268 patients from the CA cohort were also downloaded. Quality-checked sequencing reads were aligned to build hg38 of the human genome using Hisat2 (version 2-2.1.0)[95] and annotated using Gencode release 27 hg38[96]. Read counts were estimated for each gene in all samples using HTSeq[97]. Counts were normalised and transformed to $\log_2$-counts per million ($\log_2$CPM) using Voom (Limma package by BioConductor)[98]. $\log_2$CPM counts were then used as a measurement of gene expression.

ChIP-seq data processing and manipulation

Raw sequencing reads in fastq files were extracted from GEO, and checked for quality using FastQC (version 0.11.5)[99]. Where adaptors were present, sequences were trimmed using Trimmomatic tools [100]. Subsequent reads were aligned to the human reference genome (hg38) using Bowtie2 (verison 2/2.3.0) with default parameters [101], and duplicate reads were marked with Picard (MarkDuplicates)[102] and removed using SAMtools 'rmdup'[103]. Uniquely aligned reads were downsampled between ChIP-seq samples and input control pairs to avoid read yield bias. Genome-wide narrow peaks were called for H3K27ac and transcription (TF) samples, and broad peaks for H3K4me1, H3K4me3 and H3K9me3 samples against the input control using MACS2 (version 2.1.0) default settings where data was available[104]. Peaks were further filtered for quality, preserving peaks with a Q-value of E-10. Subsequent BedGraph file outputs from MACS2 were converted to BigWig files using the UCSC binary tool, BedGraphToBigWig. H3K27ac peaks located with an inter-peak distance of 2,000bp to other PDAC cell line H3K27ac regions, were merged using the 'merge' function from Bedtools (version 2.26.0) to produce one consensus H3K27ac region across all samples. H3K27ac peak co-ordinates were 'lifted' over to hg19 using the UCSC command line tool 'liftOver' to overlap with SSMs. H3K27ac regions harbouring non-coding mutations affecting >2% of the patient cohort were retained for further analysis (≥8 NCMs in ≥8 patients).

The identification of putative functional mutations using approach one (non-coding annotation/IW-scoring and LINSIGHT algorithms).

SSMs from filtered and merged H3K27ac peaks were subjected to functional testing and filtering using the IW-scoring algorithm[23]. The workflow for the identification of novel variants was utilised, excluding the use of GWAVA scores (for known variants). The median IW-functional score for all mutations within each H3K27ac consensus region was calculated. H3K27ac regions with a median IW-score of two or above were retained for further analysis. In addition, IW-scores of NCMs residing outside (H3K27ac negative) the top candidate H3K27ac consensus regions (~1kb) were obtained and compared to those of H3K27ac associated NCMs. The top candidate regions were also validated using the LINSIGHT algorithm. LINSIGHT scores were extracted as previously described[31]. The scores based on the likelihood of deleterious fitness consequences were extracted and used to compare NCMs located inside our consensus peak regions and NCMs located nearby outside peak regions (H3K27ac negative). An unpaired Wilcoxon signed rank test was used for all statistical significance testing.

The identification of putative functional mutations using approach two (LARVA algorithm).

To identify recurrently mutated regions (within H3K27ac consensus peaks) more than expected to nearby background regions, we used the algorithm LARVA[32]. This algorithm considers sample-specific mutation rates, heterogeneity and replication timing, as previously described[32]. NCMs that fell into blacklist regions were first removed, and the remaining NCMs overlapped with our H3K27ac consensus regions. Three models were used to calculate the mutation rate expected based on the stochastic background mutations. The first and second model calculates the number of local mutations within a given annotated region and estimates the probability of observing a mutation in each position. The *p*-value was drawn from a β- distribution, taking the average mutation rate and the over-dispersion, respectively into consideration. The third model considers the average replication timing within each H3K27ac element, a confounding genomic feature that would affect the background mutation rate[85]. For this, replication timing data from seven different cell lines were retrieved from ENCODE and the average timing per region calculated across all cell lines (HepG2, MCF- 7, GM12878, K562, BJ, IMR-90 and SK-N-SH GSE34399)[105]. *P*-values were

adjusted with the Benjamini-Hochberg method across all three models. We prioritised those significant H3K27ac regions with a *q* value of <0.01.

Luciferase reporter assays

Sequences surrounding NCMs of interest (~2kb total) were amplified using specific primers (Extended Data Table S4). Mutations were introduced with site-directed mutagenesis (QuikChange II Site-Directed Mutagenesis Agilent) as per the manufacturer's instructions and checked using Sanger Sequencing and correct regions cloned into the pGL2 vector upstream of the SV40 promoter. Thirty-five thousand cells (HEK293T and PANC-1) were plated 24-hours before transfection in a 24-well plate with either 100ng WT or MUT pGL2 plasmids (Promega Cat E1631) and 5ng of Renilla luciferase control (Promega Cat E2231). Luciferase activity was measured 48-hours post-transfection with the Dual-Luciferase Reporter Assay System (Promega Cat E1910). Overall activity was calculated by taking a ratio of the Firefly over the Renilla expression control vector. The background signal was quantified using un-transfected cells and subtracted from readings. An unpaired *t-test* was used to obtain statistical significance between wild-type (WT) and mutant (MUT) luciferase activity.

STARR-seq library design and cloning of candidate *cis*-regions into the STARR-seq plasmid

The STARR-seq library consisted of 6,080 constructs representing 587 candidate mutations, corresponding WT sequences and 210 controls. Constructs were represented in a 194bp sequence context, flanked by a 15bp linker region for adaptor ligation and amplification (Extended Data S4). One hundred and ten positive controls were selected from super-enhancers previously reported in PDAC [106] and additional regions from the super-enhancer database (SEdb)[107]. Putative negative controls were selected from gene deserts lacking H3K27ac and H3K4me1 marks in PDAC cell lines. A unique 6bp barcode was placed between the 5' 15bp linker and the candidate sequence to allow the differentiation between WT, MUT and control (CTRL) sequences, resulting in a final construct of 230bp. To understand the activity of mutations in different genomic contexts and maximise the chance of capturing regulatory activity, each mutation was represented in the library five times, shifting the genomic context of the sequences 10bp and 20bp left and right from the middle of the construct, thereby representing the mutation in left_20bp, left_10bp, centre,

right_10bp and right_20bp positions. The synthetic oligonucleotide library was amplified and cloned as previously described[44,108]. Briefly, 5ng of the STARR-seq library was amplified, and vector homology arms were added to either side of the construct. The second generation hSTARR-seq ORI plasmid (Addgene: #99296) was digested with SalI-HF and AgeI-HF restriction enzymes, and the amplified library was cloned into the 3-UTR of the vector. Ligations (X5 reactions) were transformed by electroporation into MegaX DH10B™ T1R Electrocomp™ Cells (Invitrogen), and reactions pooled. The plasmid pool was extracted using the ZymoPURE Giga prep kit according to the manufacturer's instructions. To check the quality and overall representation of the library, sequence inserts were amplified from the STARR-seq plasmid using Illumina-compatible index primers (Extended Data S4). STARR-seq libraries were sequenced using 2 x 150bp chemistry on an Illumina Novaseq 6000 by Novogene ltd.

STARR-seq oligo-pool quality check

Paired end reads were merged into single amplicons using the USEARCH fastq_mergepairs command[109]. Merged reads were aligned back to the expected oligo library using BWA MEM with default parameters, penalising soft-clipping of alignment ends (-L80)[110]. GATK DepthofCoverage (version 3) was used to determine the sequencing depth per nucleotide and construct[111]. Of the 6,082 constructs sequenced, 98.63% had a minimum coverage of 30X, with both WT and MUT sequences represented. To identify sequencing errors, the Samtools 'mpileup' function was run on aligned reads and the oligo reference library to obtain read counts for each nucleotide position[103]. Subsequent mpile up files were run with the VarScan2 package and 'mpileup2cns' parameters to identify sequencing errors[103,112].

Transfection, RNA isolation and cDNA synthesis

Two million PANC-1 cells were plated per 10cm dish (5 dishes per biological replicate) for 24 hours. Plasmid libraries (14µg per plate) were transfected using lipofectamine 3000 as per manufacturer instructions. To monitor transfection efficiency, one 10cm dish was co-transfected with 2.8µg of pmaxGFP plasmid (Lonza). Immediately post-transfection, the interferon inhibitors C16 and BX-795 were added to each plate at a final concentration of 1µM (per inhibitor), as previously described[113,114]. Cells were incubated at 37°c for 16 hours before harvesting and counting. 1/10th of the cells were retained for plasmid DNA, and the

635 remaining cells were for RNA extraction. For RNA, cells were homogenised with the Qiagen
636 Qiashredder and total RNA was extracted using the Qiagen mini-RNA extraction kit as per
637 the manufacturer's instructions. Poly-$(A)^+$ RNA were isolated using Dynabeads^{TM} oligo(dT)25
638 followed by DNase treatment with TurboDNase (Invitrogen). Samples were purified with
639 RNA cleanupXP beads as previously described[108]. cDNA synthesis was carried out using
640 SuperScript III and a gene-specific primer (Extended Data S4). cDNA was purified with 1.4X
641 AMpureXP beads (and for subsequent steps described below). A second-strand synthesis
642 reaction was followed by purification. Using a P7-specific primer (Extended Data Table S4)
643 UMI's were added to cDNA (in 5 reactions) with Kapa 2x HiFi HotStart ReadyMix (Kapa
644 Biosystems). Reactions were pooled and purified. Junction PCR was used to amplify
645 reporter-specific transcripts for 16 cycles and thereafter purified. For the final library
646 preparation, Illumina sequencing primers were used in cDNA samples for 8-14 cycles
647 followed by purification with 1.2X of AMPure SPRI beads (Extended Data S4).
648 To obtain the DNA input library, STARR-seq plasmids were isolated from PANC-1 cells
649 using the Monarch plasmid miniprep kit, as per the manufacturer's instructions. One
650 hundred nanograms of DNA were amplified using Illumina-compatible index primers as
651 described above. The DNA plasmid and RNA-derived libraries were sequenced using the
652 150-cycle paired-end V3 chemistry reagents and run on a Miseq.
653
654 **Processing and analysis of STARR-seq screen**
655 Paired-end reads were processed with CutAdapt to remove residual sequencing adaptors
656 and STARR-seq vector linkers[115]. Reads were split based on the 6 bp barcodes WT, MUT and
657 CTRL into separate files. Barcodes were removed, and sequences aligned to the human
658 reference genome (hg19) using BWA MEM with default parameters[110]. Aligned BAM files
659 were converted to BAMPE format using the bedtools function 'bamtobed', and properly
660 paired reads were extracted for further analysis[116]. The Bedtools 'intersect' function was
661 used to overlap reads with the expected design oligo library and obtain raw read counts.
662 Samples were deduplicated based on UMI's with a custom-made Perl script. A minimum of
663 three unique UMI's were required for a construct to be counted. Deduplicated counts were
664 normalised to the total number of reads in the sample and then multiplied by 1M to obtain
665 the number of transcripts per million. The relative abundance of each construct transcribed
666 was calculated by dividing the observed RNA output by the DNA input, indicating the

667 relative activity of each WT, MUT and CTRL construct. To compare the transcriptional
668 activity of single oligos between WT vs MUT and negative vs positive CTRLs, an unpaired t-
669 test was used. To compare the transcriptional activity at the mutation level across the five
670 sliding windows (WT vs MUT), a Mann-Whitney U statistical test was used.
671
672 **CRISPRi guide RNA design and cloning**
673 For CRISPRi, guide RNAs were selected from the UCSC genome browser 'CRISPR Tracks',
674 selecting guides as close to mutations as possible with a minimum of two guides per *cis*-
675 region (Extended Data Table S4). Potential off-target effects were assessed using the MIT
676 specificity score, selecting guides with a score above 70%[117]. Homology arm sequences were
677 added to each guide to clone into the pU6-sgRNA EF1Alpha-puro-T2A-BFP expression
678 plasmid at the BstXI-BlpI3 digested site. gRNA oligos were phosphorylated, annealed and
679 cloned into pU6-sgRNA EF1Alpha-puro-T2A-BFP expression plasmid (Addgene #60955) as
680 previously described [118]. Inserts were verified with Sanger sequencing.
681
682 **Lentivirus transduction**
683 Lentivirus was generated as previously described[118]. Briefly, 4M cells were plated in a 10cm
684 dish for 24-hours before transfecting HEK293T cells with 9ug of dCas9-mCherry-KRAB
685 (Addgene #60954), 4ug of packing plasmids psPAX.2 and 2ug of the envelope vector
686 pMD2.G diluted in OptiMEM medium and Trans-Ltl transfection reagent (Mirus). For the
687 generation of gRNA lentivirus, 9ug of each cloned guide were transfected, and the virus was
688 collected as described above. Twenty-four hours post-transfection, media was refreshed,
689 and viral supernatant was collected at 48- and 72-hours post-transfection. Viral
690 supernatants were centrifuged and filtered (45um). PANC-1 cells were transduced in a one-
691 to-one dilution of the virus and growth medium supplemented with polybrene (5ug/ml).
692 Three days post-transduction, mCherry positive cells were sorted by FACs, selecting the top
693 50% of positive cells based on the overall mCherry signal. PANC-1 dCa9/KRAB expressing
694 cells were plated in 24-well dishes for 24-hours before transducing cells with lentiviral
695 supernatant from multiple guides (as indicated in Fig. 3b). At 24-hours post-infection, the
696 medium was replaced, and cells were selected with 2ug/ml of puromycin for 72-hours. Cells
697 were harvested, and the effect on the expression of MIR100HG, UBASH3B and ARHGEF12
698 was assessed using qPCR and subsequent RNA-sequencing (Extended Data Table S4).

699

700 **qPCR**

701 RNA was extracted and DNase I treated using the Qiagen mini-RNA extraction kit according
702 to manufacturer instructions. cDNA was synthesised from 1ug of DNase treated RNA using
703 the LunaScript® RT SuperMix (NEB), according to the manufacturer's protocols. We
704 performed qPCR on a StepOneTM Real-Time PCR System with the Luna® Universal qPCR
705 Master Mix (NEB). Gene specific primers are outlined in Extended Data Table S4.

706

707 **RNA-seq data generation and analysis**

708 500 ng of total RNA was used to enrich mRNA using an oligo dT-based mRNA isolation
709 module (NEB Cat number E7490L). RNA sequencing libraries were prepared by
710 NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB catalogue number
711 E7760S). Libraries were sequenced as 150 bp paired-end reads using a Novaseq 6000. After
712 the quality check and trimming, reads were aligned to the reference genome hg38 using
713 STAR v2.7.9a[119], followed by the gene count quantification using RSEM[120] based on the
714 Ensembl gene annotation GRCh38.p13 Release 105. Genes with low mapped read across all
715 samples were removed. The normalised RPKM (Reads per kilobase of transcript per Million
716 reads mapped) expression values for all filtered genes across samples were subsequently
717 derived and used for the differential expression (DE) analysis. The DE analysis was
718 performed using Limma[121], comparing each CRISPRi perturbation group (G 5-6 and G 7-8) to
719 the dCa9/KRAB control group respectively. The significant DE genes were identified using a
720 threshold of FDR<0.05 and absolute $\log_2$FC>1. GSEA[68] was then performed based on the
721 Limma output against gene sets curated in MSigDB hallmark[69] and oncogenic signature gene
722 sets, to identify dysregulated gene activities in the CRISPRi group relative to the control.

723

724 **Cell migration assays**

725 Approximately four thousand dCas9/KRAB expressing PANC-1 cells transduced with
726 lentiviral gRNA combinations were seeded into 96-well plates. Cells were scratch wounded
727 using a 20ul pipette tip. Cells were washed with PBS to remove cell debris, and phase-
728 contrast images were taken at 0-, 24- and 48-hours at three specific wound sites per well
729 using a Leica microscope with an X4 objective. The ability of the cells to migrate and close
730 the wound area was evaluated by comparing the pixels of the open wound region at each

time point using image J (MRI wound healing plugin)[122]. An unpaired t-test was used to compare each treated time point to the negative control.

Data availability

The RNA-seq data for the CRISPRi perturbation of MIR100HG enhancer regions has been deposited to the Gene Expression Omnibus under the accession number of GSE229499. ChIP-seq data were available under GSE64560 and GSE99311. Mutation and expression data of PDAC patients were downloaded from the ICGC data portal. The STARR-seq data and all scripts to analyse the data can be requested and obtained by contacting the corresponding authors.

Acknowledgement

The work was funded by the Medical Research Council Doctoral Training Partnership (DTP) PhD programme to QMUL and University of Southampton, and by Academy of Medical Sciences Springboard Award (SBF003\1025 to J.W.). The authors acknowledge support from the Cancer Research UK City of London Major Centre core funding support to Barts Cancer Institute (C16420/A18066), and the Accelerator Award Program funded by Cancer Research UK (C355/A26819) and FC AECC and AIRC. Medical Research Council UKRI/MRC grant (MR/T000783/1) (MMP, DP) Barts charity small grant (MGU0475) (MMP).

Author contributions

J.W. conceived the study. J.W., M.M.P., and M.B.P. designed the study and developed the methodology. M.B.P., and J.W. performed the computational discovery analysis. M.B.P., A. R.-M., J.H., J.F., and M.M.P. performed the Luciferase reporter assay experiments. M.B.P., D.P., and M.M.P. performed the STARR-seq experiments. M.B.P., S.S.A., and J.W. performed the STARR-seq data analysis. M.B.P., E.M., and J.W. performed the RNA-seq analysis. M.B.P., H.K., and M.M.P. performed the in vitro functional assays. J.W., and M.M.P. supervised the study and acquired the funding. M.B.P., J.W., and M.M.P. interpreted the data and wrote the manuscript. All authors reviewed the manuscript and approved the final version of the manuscript.

Competing interests: Authors declare that they have no competing interests.

Figure legend

Fig. 1. Identification of functionally significant PDAC-CRE-associated NCMs and putative CRE drivers in PDAC. a. Overview of our investigative strategy to detect significant CRE-associated NCMs and CRE drivers **b.** The variant filtering of somatic mutations using the ICGC PDAC Australia (AU) cohort as an example. The number of H3K27ac peaks and mutations (in red) is listed at each filtering step. **c.** Putative CRE drivers and the most proximal labelled genes identified by the two independent in-silico approaches: one implementing the IW-Scoring algorithm and LINSIGHT validation, the other using the LARVA model to identify CRE-regions with recurrent NCMs. **d.** Two gene set enrichment pathway analyses of CRE-associated nearby genes identified by the two in-silico approaches. **e.** Genome browser tracks (hg19) showing the histone modifications, CRE-associated NCMs in the third intron of *MIR100HG* and *FOXP1* selected for the Luciferase reporter assay validation (grey shade). **f.** Boxplots depicting the luciferase reporter activity of selected NCMs in the introns of *MIR100HG* and *FOXP1* tested in HEK293 and PANC-1 cell lines. The top panels are for NCMs in the *MIR100HG* CRE, and the bottom panels are for NCMs in the *FOXP1* CRE. Data is representative of 3 technical replicates from 3-4 independent experiments. The statistics was performed using the unpaired t-test, with the significance p-value shown as, *<0.05, **<0.01, ***<0.001, ****<0.0001.

Fig. 2. STARR-seq to validate the regulatory activity of candidate NCMs. a. STARR-seq NCM candidate selection strategy (left) and oligo design (right). **b.** A scatter plot showing the correlation of the STARR-seq regulatory activity between the two biological replicates. STARR-seq activity was measured as the $\log_2$ transformed transcript per million (TPM) of RNA output over the DNA input. The correlation coefficient (*R*) and *p*-values are shown. **c.** Violin plots depicting the mean $\log_2$ transformed STARR-seq activity (TPM) of the two replicates comparing the negative and positive controls, p***≤0.001 (t-test). **d.** Volcano plot showing the mean $\log_2$ fold change *vs.* the log10 p-value (t-test) between MUT and WT oligos for all constructs. Pink dots demonstrate candidates with a *p*-value <0.05. Selected candidate CREs with a *p*-value <0.01 (t-test) are labelled with the closest proximal gene. **e.** Oligos with the most significant changes compared to its WT counterpart (p<0.01). MUT oligos with a higher activity than their WT sequence (gain of function) are in red bars, while MUT oligos with a lower activity than the WT control are in green. Predicted motifs identified by MotifbreakR are shown beside bars for mutations where relevant. Oligo names M1-48 are listed in Extended Data S5. **f.** Heatmap showing the gene expression profile (GEP) from the ICGC PDAC cohort (n=269) of predicted TFs putatively perturbed or gained in the top significant NCMs (p <0.01). Normalised microarray expression values are shown in the heatmap. **g.** Motif gain and loss (break) from two mutations in the *MIR100HR* enhancer cluster. The TF binding motifs for TFs NR6A1 and SOX10 are shown, and the affected nucleotide is marked in a dotted line.

Fig. 3. CRISPRi for selected CREs with NCMs within the MIR100HG enhancer cluster. a. Genome browser tracks (hg19) showing the overview of the *cis*-regulatory landscape at the MIR100HG enhancer cluster (11q24.1) and the selected CRE's for CRISPRi perturbation (grey vertical bars). The first region (left) is within a significant CRE identified by the first *in-silico* approach based on IW-Scoring, and two regions (centre right) within the significant CRE by the second approach, based on LARVA. H3K27ac and H3K4me1/3, DNase I hypersensitive sites (DHS), NCMs, guide RNA sites, microRNAs and the *BLID* gene are shown. **b.** Zoom-in of the three targeted *MIR100HG* CRE regions. RTqPCR data showing fold change in MIR100HG, *UBASH3B* (for region two G5-6) and *ARHGEF12* (for region two G7-8) levels normalised to Actin-β upon CRISPRi compared to no guide RNA control (dCAS9/KRAB). **c.** Principal component analysis (PCA) of the RNA-seq samples among the three groups, dCas9/KRAB

control, CRISPRi for region two (G 5-6) and region three (G 7-8). PC1 and PC2 were used for the separation of samples. **d.** Scatter plot of the $\log_2$ fold changes between G 5-6 and G 7-8 groups in comparison to the dCas9/KRAB control group for all profiled genes in the RNA-seq data. The correlation coefficient and *p*-value are shown. **e.** The overlap of significantly differentially expressed (DE) genes between G 5-6 and G 7-8 groups in comparison to the control. The significance cut-off is shown, and numbers of shared and unique DE genes are listed. **f.** The level of gene expression of *MIR100HG* among the dCas9/KRAB, G 5-6 and G 7-8 groups were derived from the RNA-seq data (n=3 in each group). Log_2 RPKM values were used to measure the RNA expression. A t-test was performed between the groups, with the significance p-value shown as *<0.05, **<0.01.

Fig. 4. CRISPRi for *MIR100HG* CREs results in a downregulation of KRAS and TGF-β pathways a. Significantly dysregulated pathways (false discovery rate, FDR<0.05) in the CRISPRi perturbation groups G 5-6 and G 7-8 compared to the dCas9/KRAB control group. Gene set enrichment analysis (GSEA) against the hallmark and oncogenic signature gene sets was performed [68,69]. The normalised enrichment scores (NES) were used to create the heatmap, with the positive NESs (in red) indicating the upregulation and negative NESs (in blue) indicating the downregulation of activities in the CRISPRi perturbation groups compared to the dCas9/KRAB control. **b.** GSEA plots for the TGF-β and KRAS signalling gene sets for the CRISPRi perturbation G 5-6 and G 7-8 groups compared to the dCas9/KRAB control group. The NES and FDR values for each analysis are shown. **c.** Wound healing assay with G 5-6 and G 7-8 CRISPRi clones compared to the dCas9/KRAB control samples. 0, 24 and 48-hour time points are shown. **d.** Measurement of the relative wound closure in the three groups, dCas9/KRAB control, region two (G 5-6) and region three (G 7-8) (images are a representation of n=3 biological replicates in each group). An unpaired t-test was used to compare perturbation clones vs. control groups. p-values *≤0.05. ns, not significant.

Extended Data Fig. 1. Prioritised CREs using the first *in silico* approach. (**a**) Australia (AU) and (**b**) Canada (CA) cohorts NCMs were submitted to the IW-scoring algorithm. Each dot denotes a CRE, it's combined median IW-score, across all chromosomes. The horizontal dotted line indicates the median IW-score threshold (p=0.1). The prioritised CREs with an IW-median score ≥2 are labelled by the nearest proximal gene for each cohort.

Extended Data Fig. 2. An independent validation of the top prioritised CREs using LINSIGHT.

a. Boxplots of LINSIGHT scores in H3K27ac-associated NCMs (inside peaks) in comparison to NCMs in nearby H3K27ac negative regions (outside peak) for the top 5 CREs identified by the first approach. **b**. LINSIGHT scores and location of NCMs inside (green) vs outside (black) H3K27ac peaks.

Extended Data Fig. 3. Identification of two putative CREs within the *MIR100HG* enhancer cluster. The MIR100HG enhancer cluster was the only shared element between the two in-silico approaches. One CRE was identified by the first approach based on IW-scoring and LINSIGHT validation, and the other was identified by the second approach based on the LARVA algorithm. Genome browser tracks (hg19) of the H3K27ac peaks across the 7 PDAC cell lines and 2 patient-derived organoids are shown. The MIR100HG-hosted microRNAs and associated gene *BLID* are indicated.

Extended Data Fig. 4. First *in silico* approach prioritises a significant CRE in the third intron of the TF *FOXP1*. Genome browser tracks (hg19) presenting the H3K27ac peaks across the 7 PDAC cell lines and 2 patient-derived organoids and the annotation of the merged consensus peaks. The location of NCMs inside and outside H3K27ac consensus is indicated for each cohort.

Extended Data Fig. 5. Quality analysis of the cloned STARR-seq plasmid library. a. Length distribution of cloned oligo constructs. The percentage of each length is shown. The oligo construct library had 49% of oligos with the expected correct length, followed by 1- (28%) and 2-bp (10%) deletions. **b.** Depiction of the number of construct synthesis errors across the sequenced oligos. The error occurrence is shown along the base pair positions.

Extended Data Fig. 6. Comparison of regulatory activities derived between Luciferase reporter assay and STARR-seq. Boxplots comparing the Luciferase reporter assay activity and STARR-seq, demonstrating concurrent directional changes in enhancer activity at NCMs profiled by both techniques.

Extended Data Fig. 7. BNC1 associated enhancer cluster and significant NCMs from the STARR-seq screen. a. Genome browser (hg19) of the H3K27ac peaks across the 7 PDAC cell lines and 2 patient-derived organoids. The STARR-seq significant NCMs are indicated, with the vast majority residing upstream of the *BNC1* gene promoter. **b.** Boxplots showing the expression of genes within 1Mb of the *BNC1* CRE that have significant alterations (*BTBD1*, *FAM103A1* and *TM6SF1*) and the nearest proximal gene (*BNC1*), comparing mutant (MUT) and wildtype (WT) patient gene expression profiles (GEPs). The p-values were derived from the Wilcoxon rank sum test. **c**. Motif gain example for a gain-of-function NCM proximal to the *TBX3* gene (chr12:115067012:C>A). A binding motif for the TF JUN is created by this mutation. All TF binding predictions were carried out using MotifBreakR. **d.** Motif break example for one loss-of-function NCM located in the intron of *FOXP1* (chr3:71123616:G>T). The binding motif for ATF3 was disrupted by this mutation.

Extended Data Fig. 8. The interaction of MIR100HG CREs with distal genes revealed by the 4D genome. a. Genome browser (hg19) showing the H3K27ac signal for PANC-1 cells and the putative loop between CRE-2 and the promoter of *UBASH3B*. Putative loops were predicted using the interactions from the integrated method for predicting enhancer targets (IM-PET) and 4D Genome in PANC-1 cells. **b**. Genome browser (hg19) showing the H3K27ac signal for PANC-1 cells and the putative loop between CRE-3 and the promoter of *ARHGEF12*.

Extended Data Fig. 9. Boxplots comparing the gene expression profiles of TGF-β related genes *FGF1*, *KDM6B*, *LIF*, *PIK3CD*, *PXDC1* and *TAGLN* between G 5-6 / G 7-8 CRISPRi perturbations and the dCas9/KRAB control. The gene expression levels were measured by RNA-seq data, in the unit of log2 RPKM values.

Extended Data Fig. 10. The mutational burden in the top five significant CREs identified in the first approach in other common solid tumours. CREs overlaying genes *FOXA1*, *FOXP1*, *HOTAIR/HOXC* genes, *MIR100HG* and *NFEL2* were assessed for their mutational burden. **a.** Barplot showing the number of samples across the selected cancer cohorts. **b.** Barplot of

the total number of mutations within each cohort. **c.** Barplots showing the frequency of NCMs identified within each CRE across each cancer cohort.

Extended Data S1. List of NCMs in significant CRE's prioritised using the first *in-silico* approach in the AU and CA cohorts respectively.

Extended Data S2. Table showing the comparison of NCMs inside the top five significant CREs to NCMs located outside flanking H3K27ac negative regions. An unpaired Wilcoxon signed rank test was used to obtain p values.

Extended Data S3. List of all significant CREs prioritised by the LARVA algorithm and those significant CREs found to be in common between the AU and CA cohorts.

Extended Data S4. List of primers and CRISPRi guides.

Extended Data S5. List of top significant oligos as shown in Fig. 2e.

References:

1 Klein, A. P. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nature reviews. Gastroenterology & hepatology* **18** (2021). https://doi.org:10.1038/s41575-021-00457-x

2 Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518** (2015). https://doi.org:10.1038/nature14169

3 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531** (2016). https://doi.org:10.1038/nature16965

4 Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine* **17** https://doi.org:10.1038/nm.2344

5 Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature genetics* **47** (2015). https://doi.org:10.1038/ng.3398

6 Murphy, S. J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145** (2013). https://doi.org:10.1053/j.gastro.2013.07.049

7 Zhang, J. *et al.* The International Cancer Genome Consortium Data Portal. *Nature biotechnology* **37** (2019). https://doi.org:10.1038/s41587-019-0055-9

8 Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature genetics* **46** (2014). https://doi.org:10.1038/ng.3141

9 Rachakonda, P. S. *et al.* TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proceedings of the National Academy of Sciences of the United States of America* **110** (2013). https://doi.org:10.1073/pnas.1310522110

10 Hosen, I. *et al.* TERT promoter mutations in clear cell renal cell carcinoma. *International journal of cancer* **136** (2015). https://doi.org:10.1002/ijc.29279

11 Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)* **339** (2013). https://doi.org:10.1126/science.1230062

12 Rahman, S. *et al.* Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. *Blood* **129** (2017). https://doi.org:10.1182/blood-2016-09-742148

13 S., N.-Z. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534** (2016). https://doi.org:10.1038/nature17676

14 Rheinbay, E. *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578** (2020). https://doi.org:10.1038/s41586-020-1965-x

15 Dietlein, F. *et al.* Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science (New York, N.Y.)* **376** (2022). https://doi.org:10.1126/science.abg5601

16 Feigin, M. E. *et al.* Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nature genetics* **49** (2017). https://doi.org:10.1038/ng.3861

17 Mansour, M. R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science (New York, N.Y.)* **346** (2014). https://doi.org:10.1126/science.1259037

18 Abraham, B. J. *et al.* Small genomic insertions form enhancers that misregulate oncogenes. *Nature communications* **8** (2017). https://doi.org:10.1038/ncomms14385

19 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107** (2010). https://doi.org:10.1073/pnas.1016071107

20 Corona, R. I. *et al.* Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. *Nature communications* **11** (2020). https://doi.org:10.1038/s41467-020-15951-0

21 Diaferia, G. R. *et al.* Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. *The EMBO journal* **35** (2016). https://doi.org:10.15252/embj.201592404

22 Boj, S. F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160** (2015). https://doi.org:10.1016/j.cell.2014.12.021

23 Wang, J., Dayem Ullah, A. Z. & Chelala, C. IW-Scoring: an Integrative Weighted Scoring framework for annotating and prioritizing genetic variations in the noncoding genome. *Nucleic acids research* **46** (2018). https://doi.org:10.1093/nar/gky057

24 Milan, M. *et al.* FOXA2 controls the cis-regulatory networks of pancreatic cancer cells in a differentiation grade-specific manner. *The EMBO journal* **38** (2019). https://doi.org:10.15252/embj.2019102161

25 Ottaviani, S. *et al.* TGF-β induces miR-100 and miR-125b but blocks let-7a through LIN28B controlling PDAC progression. *Nature communications* **9** (2018). https://doi.org:10.1038/s41467-018-03962-x

26 Li, C. H. *et al.* EZH2 coupled with HOTAIR to silence MicroRNA-34a by the induction of heterochromatin formation in human pancreatic ductal adenocarcinoma. *International journal of cancer* **140** (2017). https://doi.org:10.1002/ijc.30414

27 Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nature reviews. Cancer* **10** (2010). https://doi.org:10.1038/nrc2826

28 Roe, J. S. *et al.* Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* **170** (2017). https://doi.org:10.1016/j.cell.2017.07.007

29 Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nature genetics* **47** (2015). https://doi.org:10.1038/ng.3164

30 Forcina, G. C. *et al.* Ferroptosis regulation by the NGLY1/NFE2L1 pathway. *Proceedings of the National Academy of Sciences of the United States of America* **119** (2022). https://doi.org:10.1073/pnas.2118646119

31 Huang, Y. F., Gulko, B. & Siepel, A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nature genetics* **49** (2017). https://doi.org:10.1038/ng.3810

32 Lochovsky, L., Zhang, J., Fu, Y., Khurana, E. & Gerstein. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. *Nucleic acids research* **43** (2015). https://doi.org:10.1093/nar/gkv803

33 Eissa Maryam, A. L. *et al.* Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood. *Clinical Epigenetics* **11**, 1-10 (2019). https://doi.org:doi:10.1186/s13148-019-0650-0

34 Perkhofer, L. *et al.* Tbx3 fosters pancreatic cancer growth by increased angiogenesis and activin/nodal-dependent induction of stemness. *Stem cell research* **17** (2016). https://doi.org:10.1016/j.scr.2016.08.007

35 Tauber, S. *et al.* Transcriptome analysis of human cancer reveals a functional role of heme oxygenase-1 in tumor cell adhesion. *Molecular cancer* **9** (2010). https://doi.org:10.1186/1476-4598-9-200

36 Peppino, G. *et al.* Teneurins: Role in Cancer and Potential Role as Diagnostic Biomarkers and Targets for Therapy. *International journal of molecular sciences* **22** (2021). https://doi.org:10.3390/ijms22052321

37 Qu, Q. H., Jiang, S. Z. & Li, X. Y. LncRNA TBX5-AS1 Regulates the Tumor Progression Through the PI3K/AKT Pathway in Non-Small Cell Lung Cancer. *OncoTargets and therapy* **13** (2020). https://doi.org:10.2147/OTT.S255195

38 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4** (2009). https://doi.org:10.1038/nprot.2008.211

39 Nikolskaya, T. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science (New York, N.Y.)* **321** (2008). https://doi.org:10.1126/science.1164368

40 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)* **339** (2013). https://doi.org:10.1126/science.1232542

41 Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics (Oxford, England)* **31** (2015). https://doi.org:10.1093/bioinformatics/btv470

42 Inoue, F. *et al.* A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome research* **27** (2017). https://doi.org:10.1101/gr.212092.116

43 Ulirsch, J. C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **165** (2016). https://doi.org:10.1016/j.cell.2016.04.048

44 Neumayr, C., Pagani, M., Stark, A. & Arnold, C. D. STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Curr Protoc Mol Biol* **128**, e105 (2019). https://doi.org:10.1002/cpmb.105

45 Eissa, M. A. L. *et al.* Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood. *Clinical epigenetics* **11** (2019). https://doi.org:10.1186/s13148-019-0650-0

46 Pisani, D. F. *et al.* Involvement of BTBD1 in mesenchymal differentiation. *Experimental cell research* **313** (2007). https://doi.org:10.1016/j.yexcr.2007.03.030

47 Gonatopoulos-Pournatzis, T., Dunn, S., Bounds, R. & Cowling, V. H. RAM/Fam103a1 is required for mRNA cap methylation. *Molecular cell* **44** (2011). https://doi.org:10.1016/j.molcel.2011.08.041

48 Xu, R. *et al.* Combined Analysis of the Aberrant Epigenetic Alteration of Pancreatic Ductal Adenocarcinoma. *BioMed research international* **2019** (2019). https://doi.org:10.1155/2019/9379864

49 de Groot, J. S. *et al.* Validation of DNA promoter hypermethylation biomarkers in breast cancer--a short report. *Cellular oncology (Dordrecht)* **37** (2014). https://doi.org:10.1007/s13402-014-0189-1

50 Perkhofer, L. *et al.* Tbx3 fosters pancreatic cancer growth by increased angiogenesis and activin/nodal-dependent induction of stemness. *Stem cell research* **17** (2016). https://doi.org:10.1016/j.scr.2016.08.007

51 Tessari, G. *et al.* The expression of proto-oncogene c-jun in human pancreatic cancer. *Anticancer research* **19** (1999).

52 Azizi, N. *et al.* Loss of activating transcription factor 3 prevents KRAS-mediated pancreatic cancer. *Oncogene* **40** (2021). https://doi.org:10.1038/s41388-021-01771-z

53 Tong, X. *et al.* SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/β-catenin pathway. *Oncotarget* **5** (2014). https://doi.org:10.18632/oncotarget.2512

54 Papoutsoglou, P. *et al.* The noncoding MIR100HG RNA enhances the autocrine function of transforming growth factor β signaling. *Oncogene* **40** (2021). https://doi.org:10.1038/s41388-021-01803-8

55 Su, X. *et al.* ELK1-induced upregulation of long non-coding RNA MIR100HG predicts poor prognosis and promotes the progression of osteosarcoma by epigenetically silencing LATS1 and LATS2. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **109** (2019). https://doi.org:10.1016/j.biopha.2018.10.029

56 Wang, S. *et al.* LncRNA MIR100HG promotes cell proliferation in triple-negative breast cancer through triplex formation with p27 loci. *Cell death & disease* **9** (2018). https://doi.org:10.1038/s41419-018-0869-2

1097 57 Emmrich, S. *et al.* LincRNAs MONC and MIR100HG act as oncogenes in acute
1098 megakaryoblastic leukemia. *Molecular cancer* **13** (2014).
1099 https://doi.org:10.1186/1476-4598-13-171
1100 58 Li, W. *et al.* Elevated MIR100HG promotes colorectal cancer metastasis and is
1101 associated with poor prognosis. *Oncology letters* **18** (2019).
1102 https://doi.org:10.3892/ol.2019.11060
1103 59 Li, J., Xu, Q., Wang, W. & Sun, S. MIR100HG: a credible prognostic biomarker and an
1104 oncogenic lncRNA in gastric cancer. *Bioscience reports* **39** (2019).
1105 https://doi.org:10.1042/BSR20190171
1106 60 Broustas, C. G. *et al.* The proapoptotic molecule BLID interacts with Bcl-XL and its
1107 downregulation in breast cancer correlates with poor disease-free and overall
1108 survival. *Clinical cancer research : an official journal of the American Association for*
1109 *Cancer Research* **16** (2010). https://doi.org:10.1158/1078-0432.CCR-09-2351
1110 61 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and
1111 Activation. *Cell* **159** (2014). https://doi.org:10.1016/j.cell.2014.09.029
1112 62 Teng, L., He, B., Wang, J. & Tan, K. 4DGenome: a comprehensive database of
1113 chromatin interactions. *Bioinformatics (Oxford, England)* **32** (2016).
1114 https://doi.org:10.1093/bioinformatics/btw375
1115 63 He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in
1116 human cells. *Proceedings of the National Academy of Sciences of the United States of*
1117 *America* **111** (2014). https://doi.org:10.1073/pnas.1320308111
1118 64 Meyers, N., Gérard, C., Lemaigre, F. P. & Jacquemin, P. Differential impact of the
1119 ERBB receptors EGFR and ERBB2 on the initiation of precursor lesions of pancreatic
1120 ductal adenocarcinoma. *Scientific reports* **10** (2020). https://doi.org:10.1038/s41598-
1121 020-62106-8
1122 65 Navas, C. *et al.* EGF receptor signaling is essential for k-ras oncogene-driven
1123 pancreatic ductal adenocarcinoma. *Cancer cell* **22** (2012).
1124 https://doi.org:10.1016/j.ccr.2012.08.001
1125 66 Kowanetz, K. *et al.* Suppressors of T-cell receptor signaling Sts-1 and Sts-2 bind to Cbl
1126 and inhibit endocytosis of receptor tyrosine kinases. *The Journal of biological*
1127 *chemistry* **279** (2004). https://doi.org:10.1074/jbc.M403759200
1128 67 Rath, N. *et al.* ROCK signaling promotes collagen remodeling to facilitate invasive
1129 pancreatic ductal adenocarcinoma tumor cell growth. *EMBO molecular medicine* **9**
1130 (2017). https://doi.org:10.15252/emmm.201606743
1131 68 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach
1132 for interpreting genome-wide expression profiles. *Proceedings of the National*
1133 *Academy of Sciences of the United States of America* **102** (2005).
1134 https://doi.org:10.1073/pnas.0506580102
1135 69 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set
1136 collection. *Cell systems* **1** (2015). https://doi.org:10.1016/j.cels.2015.12.004
1137 70 Morris, J. P., Wang, S. C. & Hebrok, M. KRAS, Hedgehog, Wnt and the twisted
1138 developmental biology of pancreatic ductal adenocarcinoma. *Nature reviews. Cancer*
1139 **10** (2010). https://doi.org:10.1038/nrc2899
1140 71 Sarkar Bhattacharya, S., Mandal, C., Albiez, R. S., Samanta, S. K. & Mandal, C.
1141 Mahanine drives pancreatic adenocarcinoma cells into endoplasmic reticular stress-
1142 mediated apoptosis through modulating sialylation process and Ca2+-signaling.
1143 *Scientific reports* **8** (2018). https://doi.org:10.1038/s41598-018-22143-w

72 Adjuto-Saccone, M. *et al.* TNF-α induces endothelial-mesenchymal transition promoting stromal development of pancreatic adenocarcinoma. *Cell death & disease* **12** (2021). https://doi.org:10.1038/s41419-021-03920-4

73 Benjamin, J. R. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer cell* **32** (2017). https://doi.org:10.1016/j.ccell.2017.07.007

74 Zimmerman, N. P. *et al.* Cyclic AMP regulates the migration and invasion potential of human pancreatic cancer cells. *Molecular carcinogenesis* **54** (2015). https://doi.org:10.1002/mc.22091

75 Santiago, L., Daniels, G., Wang, D., Deng, F. M. & Lee, P. Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment. *American journal of cancer research* **7** (2017).

76 Carter, E. P. *et al.* Dissecting FGF Signalling to Target Cellular Crosstalk in Pancreatic Cancer. *Cells* **10** (2021). https://doi.org:10.3390/cells10040847

77 Lee, S. H., Kim, O., Kim, H. J., Hwangbo, C. & Lee, J. H. Epigenetic regulation of TGF-β-induced EMT by JMJD3/KDM6B histone H3K27 demethylase. *Oncogenesis* **10** (2021). https://doi.org:10.1038/s41389-021-00307-0

78 Shi, Y. *et al.* Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* **569** (2019). https://doi.org:10.1038/s41586-019-1130-6

79 Agnetti, J. *et al.* PI3Kδ activity controls plasticity and discriminates between EMT and stemness based on distinct TGFβ signaling. *Communications Biology* **5**, 1-15 (2022). https://doi.org:doi:10.1038/s42003-022-03637-w

80 Chen, Z. *et al.* TGF-β-induced transgelin promotes bladder cancer metastasis by regulating epithelial-mesenchymal transition and invadopodia formation. *EBioMedicine* **47** (2019). https://doi.org:10.1016/j.ebiom.2019.08.012

81 Zhou, S. *et al.* Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nature communications* **11** (2020). https://doi.org:10.1038/s41467-020-14318-9

82 Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The ensembl regulatory build. *Genome biology* **16** (2015). https://doi.org:10.1186/s13059-015-0621-5

83 Elliott, K. & Larsson, E. Non-coding driver mutations in human cancer. *Nature reviews. Cancer* **21** (2021). https://doi.org:10.1038/s41568-021-00371-z

84 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499** (2013). https://doi.org:10.1038/nature12213

85 Stamatoyannopoulos, J. A. *et al.* Human mutation rate associated with DNA replication timing. *Nature genetics* **41** (2009). https://doi.org:10.1038/ng.363

86 Zhang, P. *et al.* High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. *Nature communications* **9** (2018). https://doi.org:10.1038/s41467-018-04451-x

87 Schöne, S. *et al.* Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise. *PLoS genetics* **14** (2018). https://doi.org:10.1371/journal.pgen.1007793

88 Luo, X. *et al.* The clinicopathological significance of forkhead box P1 and forkhead box O3a in pancreatic ductal adenocarcinomas. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **39** (2017). https://doi.org:10.1177/1010428317699129

89 Sauzeau, V., Berenjeno, I. M., Citterio, C. & Bustelo, X. R. A transcriptional cross-talk between RhoA and c-Myc inhibits the RhoA/Rock-dependent cytoskeleton. *Oncogene* **29** (2010). https://doi.org:10.1038/onc.2010.134

90 Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575** (2019). https://doi.org:10.1038/s41586-019-1694-1

91 Kemp, S. B. *et al.* Efficacy of a Small-Molecule Inhibitor of KrasG12D in Immunocompetent Models of Pancreatic Cancer. *Cancer discovery* **13** (2023). https://doi.org:10.1158/2159-8290.CD-22-1066

92 Pradeepa, M. M. *et al.* Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature genetics* **48** (2016). https://doi.org:10.1038/ng.3550

93 Pal, D. *et al.* H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function. *Nature structural & molecular biology* (2023). https://doi.org:10.1038/s41594-023-01016-5

94 Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols* **10** (2015). https://doi.org:10.1038/nprot.2015.105

95 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37** (2019). https://doi.org:10.1038/s41587-019-0201-4

96 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research* **47** (2019). https://doi.org:10.1093/nar/gky955

97 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* **31** (2015). https://doi.org:10.1093/bioinformatics/btu638

98 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15** (2014). https://doi.org:10.1186/gb-2014-15-2-r29

99 *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2020).

100 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014). https://doi.org:10.1093/bioinformatics/btu170

101 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 https://doi.org:10.1038/nmeth.1923

102 Ebbert, M. T. *et al.* Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC bioinformatics* **17 Suppl 7** (2016). https://doi.org:10.1186/s12859-016-1097-3

103 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25** (2009). https://doi.org:10.1093/bioinformatics/btp352

104 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137

105 Consortium., E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489** (2012). https://doi.org:10.1038/nature11247

106 Ghosh, C. *et al.* Super-enhancers: novel target for pancreatic ductal adenocarcinoma. *Oncotarget* **10**, 1554-1571 (2019). https://doi.org:10.18632/oncotarget.26704

107 Jiang, Y. *et al.* SEdb: a comprehensive human super-enhancer database. *Nucleic acids research* **47** (2019). https://doi.org:10.1093/nar/gky1025

108 Muerdter, F. *et al.* Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* **15**, 141-149 (2018). https://doi.org:10.1038/nmeth.4534

109 Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics (Oxford, England)* **31** (2015). https://doi.org:10.1093/bioinformatics/btv401

110 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* (2013).

111 Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud*. (@OReillyMedia, 2023).

112 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* **22** (2012). https://doi.org:10.1101/gr.129684.111

113 Muerdter, F. *et al.* Resolving systematic errors in widely used enhancer activity assays in human cells. *Nature methods* **15** (2018). https://doi.org:10.1038/nmeth.4534

114 Neumayr, C., Pagani, M., Stark, A. & Arnold, C. D. STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Current protocols in molecular biology* **128** (2019). https://doi.org:10.1002/cpmb.105

115 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **17(1):3** (2011). https://doi.org:https://journal.embnet.org/index.php/embnetjournal/article/view/200

116 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26** (2010). https://doi.org:10.1093/bioinformatics/btq033

117 Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome biology* **17** (2016). https://doi.org:10.1186/s13059-016-1012-2

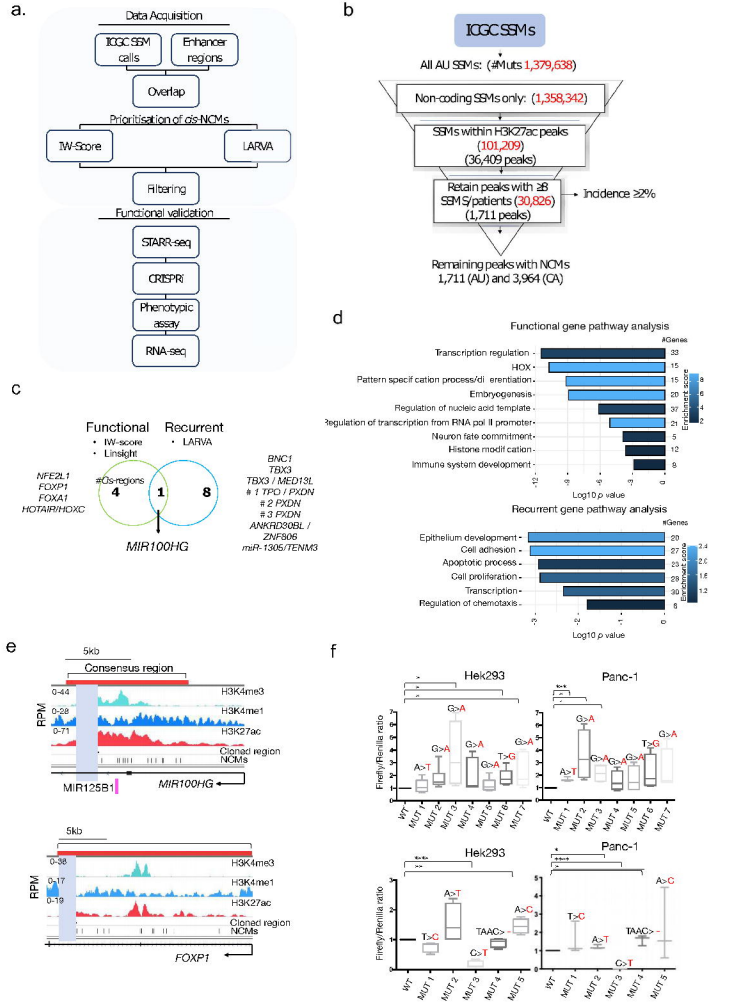
118 Stojic, L. *et al.* Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. *Nucleic acids research* **46** (2018). https://doi.org:10.1093/nar/gky437

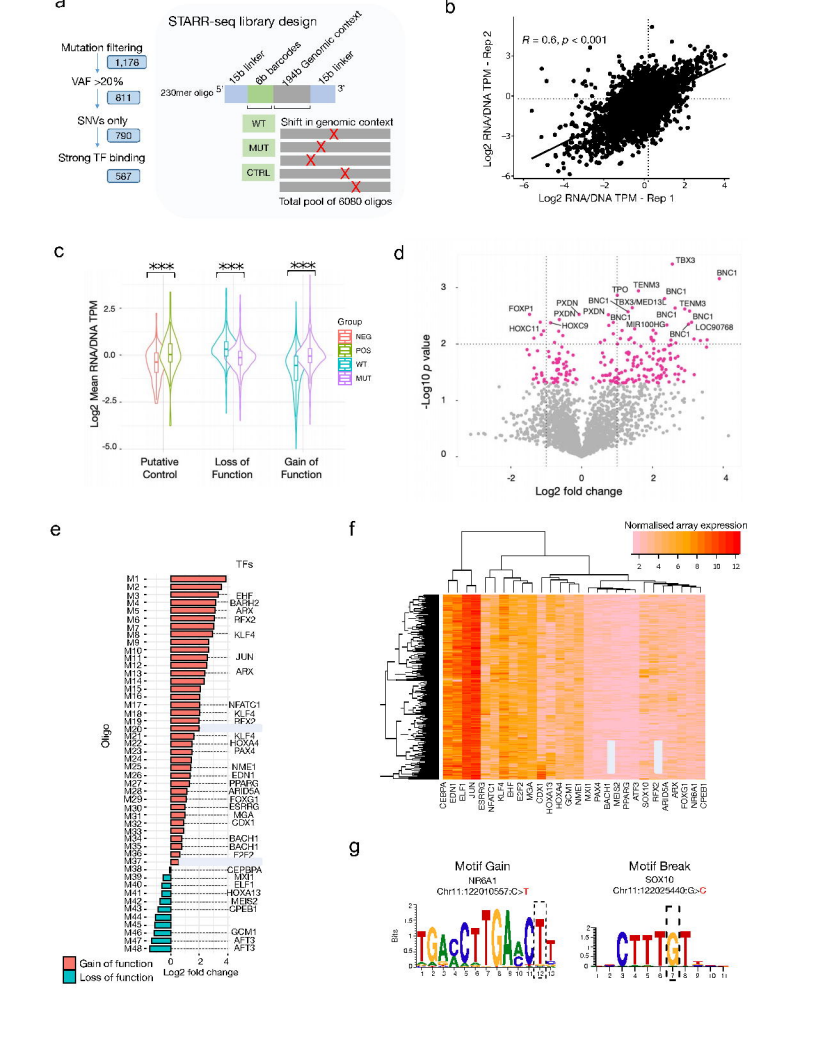
119 Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics* **51** (2015). https://doi.org:10.1002/0471250953.bi1114s51

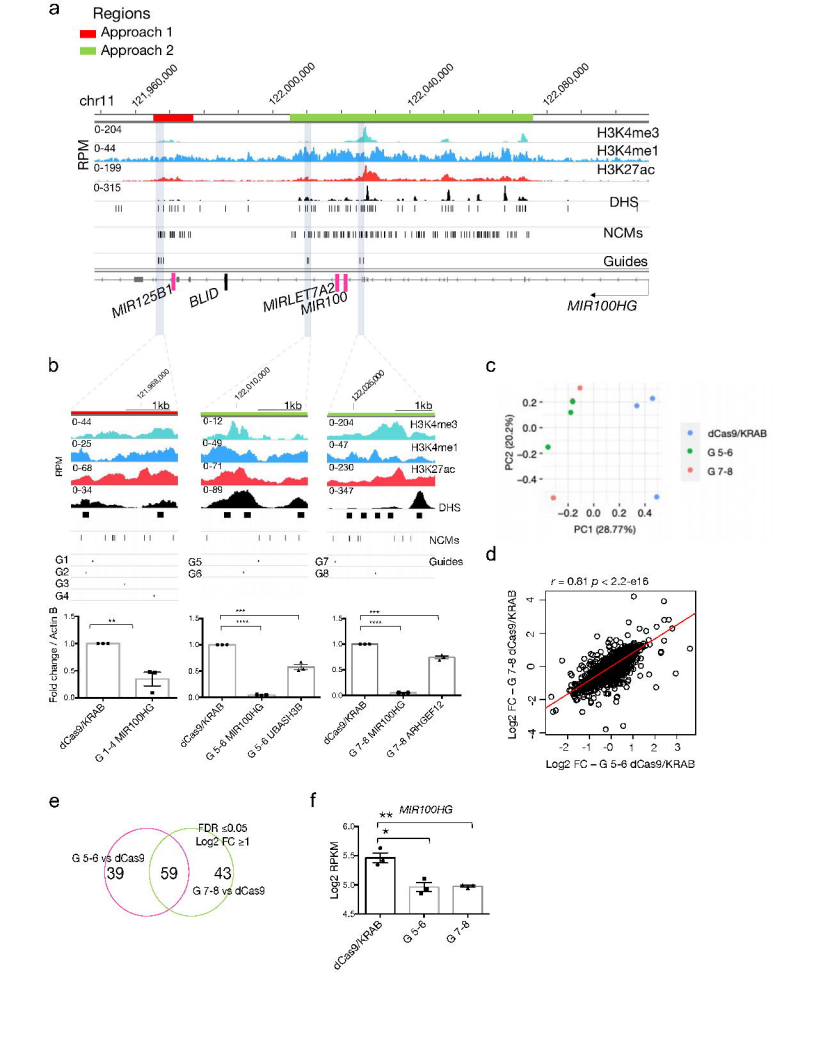
120 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12** (2011). https://doi.org:10.1186/1471-2105-12-323

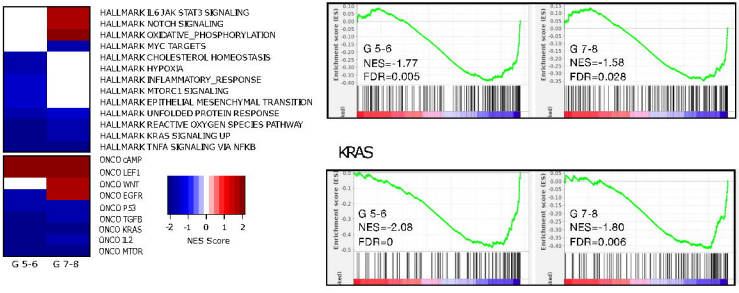
121 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43** (2015). https://doi.org:10.1093/nar/gkv007

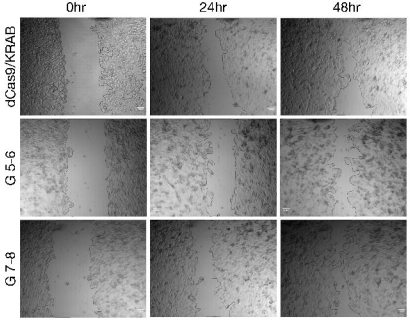
122 Suarez-Arnedo, A. *et al.* An image J plugin for the high throughput image analysis of in vitro scratch wound healing assays. *PloS one* **15** (2020). https://doi.org:10.1371/journal.pone.0232565

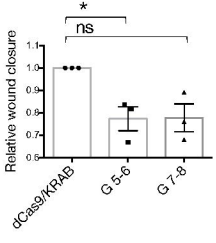

a.
Data Acquisition
ICGC SSM calls
Enhancer regions
Overlap
Prioritisation of cis-NCMs
IW-Score
LARVA
Filtering
Functional validation
STARR-seq
CRISPRi
Phenotypic assay
RNA-seq
b
ICGC SSMs
All AU SSMs: (#Muts 1,379,638)
Non-coding SSMs only: (1,358,342)
SSMs within H3K27ac peaks (101,209) (36,409 peaks)
Retain peaks with ≥8 SSMS/patients (30,826) (1,711 peaks)
Incidence ≥2%
Remaining peaks with NCMs 1,711 (AU) and 3,964 (CA)
c
Functional
IW-score
Linsight
Recurrent
LARVA
#cis-regions
4
1
8
NFE2L1
FOXP1
FOXA1
HOTAIR/HOXC
MIR100HG
BNC1
TBX3
TBX3 / MED13L
1 TPO / PXDN
2 PXDN
3 PXDN
ANKRD30BL / ZNF806
miR-1305/TENM3
d
Functional gene pathway analysis
#Genes
Transcription regulation
HOX
Pattern specification process/dorsal ventral axis
Embryogenesis
Regulation of nucleic acid template
Regulation of transcription from RNA pol II promoter
Neuron fate commitment
Histone modification
Immune system development
Enrichment score
Log10 p value
Recurrent gene pathway analysis
#Genes
Epithelium development
Cell adhesion
Apoptotic process
Cell proliferation
Transcription
Regulation of chemotaxis
Enrichment score
Log10 p value
e
5kb
Consensus region
RPM
H3K4me3
H3K4me1
H3K27ac
Clonal region NCMs
MIR125B1
MIR100HG
FOXP1
f
Hek293
Panc-1
Firefly/Renilla ratio
WT
MUT 1
MUT 2
MUT 3
MUT 4
MUT 5
MUT 6
MUT 7


a
STARR-seq library design
Mutation filtering
1,176
VAF >20%
611
SNVs only
790
Strong TF binding
567
230mer oligo 5'
15b linker
8b barcodes
194b Genomic context
15b linker
3'
WT
MUT
CTRL
Shift in genomic context
Total pool of 6080 oligos
b
R = 0.6, p < 0.001
Log2 RNA/DNA TPM - Rep 2
Log2 RNA/DNA TPM - Rep 1
c
***
***
***
Log2 Mean RNA/DNA TPM
Group
NEG
POS
WT
MUT
Putative Control
Loss of Function
Gain of Function
d
TBX3
BNC1
TENM3
TPO
PXDN
BNC1
TBX3
MED13L
TENM3
FOXP1
PXDN
PXDN
BNC1
BNC1
BNC1
HOXC11
HOXC9
MIR100HG
LOC90768
BNC1
-Log10 p value
Log2 fold change
e
TFs
Oligo
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
M23
M24
M25
M26
M27
M28
M29
M30
M31
M32
M33
M34
M35
M36
M37
M38
M39
M40
M41
M42
M43
M44
M45
M46
M47
M48
EHF
BARHL2
ARX
RFX2
KLF4
JUN
ARX
NFATC1
KLF4
RFX3
KLF4
HOXA4
PAX4
NME1
EDN1
PPARG
ARID5A
FOXG1
ESRRG
MGA
CDX1
BACH1
BACH1
FOF2
CEBPA
MAF
ELF1
HOXA13
MEIS2
CPEB1
GCM1
ATF3
ATF3
Gain of function
Loss of function
Log2 fold change
f
Normalised array expression
CEBPA
EDN1
ELF1
JUN
ESRRG
NFATC1
KLF4
EHF
MGA
CDX1
HOXA13
GCM1
NME1
MAF
PAX4
BACH1
MEIS2
PPARG
ATF3
SPIB
RFX3
ARID5A
ARX
FOXG1
NR6A1
CPEB1
g
Motif Gain
NR6A1
Chr11:122015057:C>T
Motif Break
SOX10
Chr11:122025440:G>C
Bits


a
Regions
Approach 1
Approach 2
chr11
121,960,000
122,000,000
122,040,000
122,080,000
RPM
0-204
H3K4me3
0-44
H3K4me1
0-199
H3K27ac
0-315
DHS
NCMs
Guides
MIR125B1
BLID
MIRLET7A2
MIR100
MIR100HG
b
121,968,000
122,010,000
122,056,000
1kb
0-44
0-25
0-68
0-34
0-12
0-49
0-71
0-89
0-204
0-47
0-230
0-347
G1
G2
G3
G4
G5
G6
G7
G8
Fold change / Actin B
dCas9/KRAB
G 1-4 MIR100HG
G 5-6 MIR100HG
G 5-6 UBASH3B
G 7-8 MIR100HG
G 7-8 APH3EF12
c
PC2 (20.2%)
PC1 (28.77%)
dCas9/KRAB
G 5-6
G 7-8
d
r = 0.81 p < 2.2-e16
Log2 FC – G 7-8 dCas9/KRAB
Log2 FC – G 5-6 dCas9/KRAB
e
G 5-6 vs dCas9
39
59
43
G 7-8 vs dCas9
FDR ≤0.05
Log2 FC ≥1
f
MIR100HG
Log2 RPKM
dCas9/KRAB
G 5-6
G 7-8


a
HALLMARK IL6 JAK STAT3 SIGNALING
HALLMARK NOTCH SIGNALING
HALLMARK OXIDATIVE_PHOSPHORYLATION
HALLMARK MYC TARGETS
HALLMARK CHOLESTEROL HOMEOSTASIS
HALLMARK HYPOXIA
HALLMARK INFLAMMATORY_RESPONSE
HALLMARK MTORC1 SIGNALING
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
HALLMARK UNFOLDED PROTEIN RESPONSE
HALLMARK REACTIVE OXYGEN SPECIES PATHWAY
HALLMARK KRAS SIGNALING UP
HALLMARK TNFA SIGNALING VIA NFKB
ONCO cAMP
ONCO LEF1
ONCO WNT
ONCO EGFR
ONCO P53
ONCO TGFB
ONCO KRAS
ONCO IL2
ONCO MTOR
G 5-6
G 7-8
-2 -1 0 1 2
NES Score
b
TGF-β
G 5-6
NES=-1.77
FDR=0.005
G 7-8
NES=-1.58
FDR=0.028
kRAS
G 5-6
NES=-2.08
FDR=0
G 7-8
NES=-1.80
FDR=0.006


c
0hr
24hr
48hr
dCas9/KRAB
G 5-6
G 7-8


d
*
ns
Relative wound closure
1.1
1.0
0.9
0.8
0.7
0.6
dCas9/KRAB
G 5-6
G 7-8