Cell Reports, Volume *40*

Supplemental information

Arid1a mutation suppresses TGF-β signaling and induces cholangiocarcinoma

Bing Guo, Scott C. Friedland, William Alexander, Jacquelyn A. Myers, Wenjia Wang, Michael R. O'Dell, Michael Getman, Christa L. Whitney-Miller, Diana Agostini-Vulaj, Aaron R. Huber, Stephano S. Mello, Paula M. Vertino, Hartmut K. Land, Laurie A. Steiner, and Aram F. Hezel

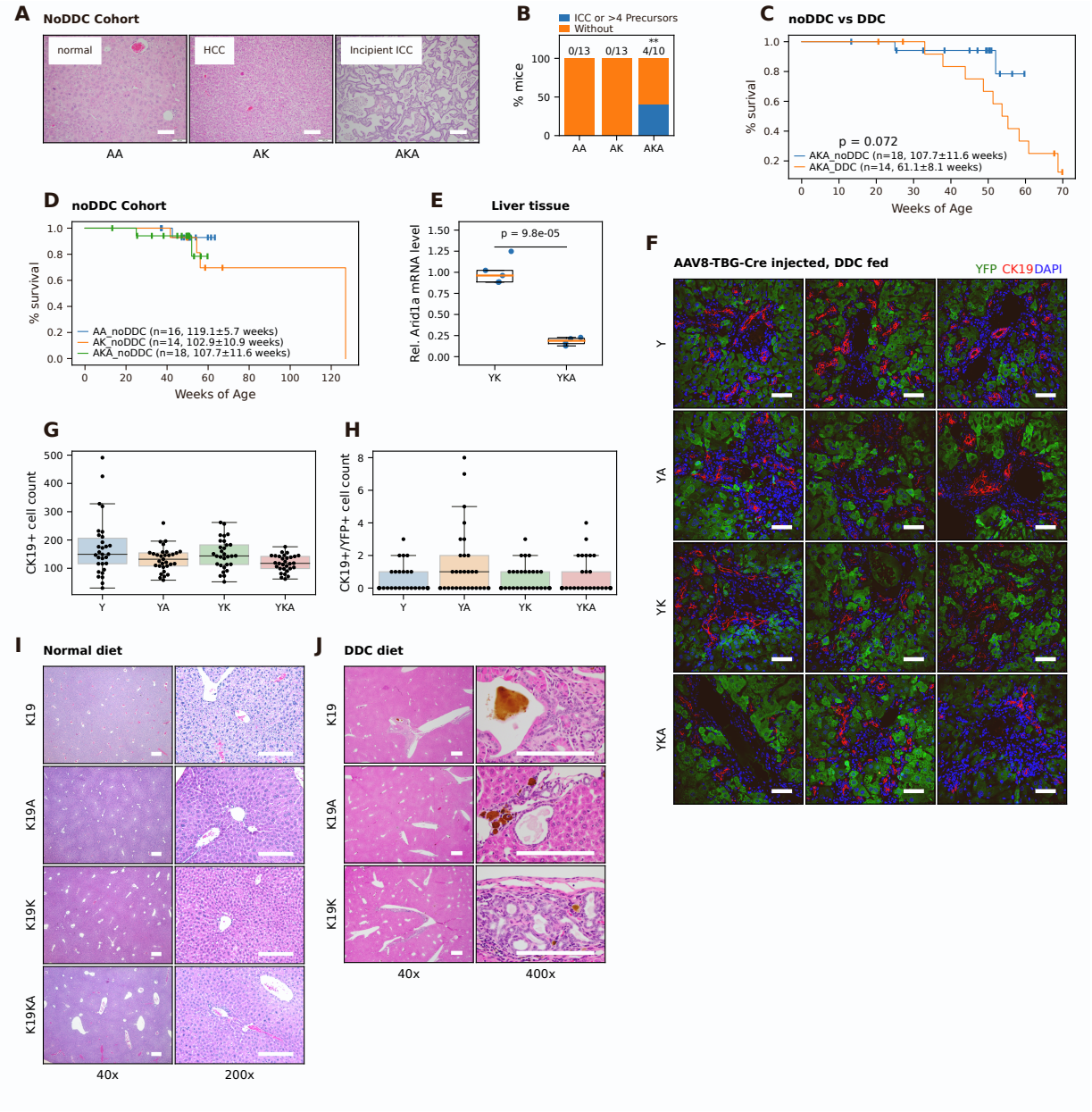


Figure S1. Extended data on the cooperative role of *Kras/Arid1a* mutations in liver homeostasis and tumorigenesis (Related to Figure 1)

A. Representative H&E images of noDDC *AKA*, *AK*, and *AA* livers harvested at around 50 weeks old. Scale bar 100 µm.
B. Frequencies of CC and CC precursors in *AKA* (n=10), *AK* (n=13), and *AA* (n=13) mice of the noDDC cohorts. The frequency in the *AKA* cohort is different *AA* and *AK* (Fisher's exact test, $p < 0.05$). More details are available in **Table S1.**
C. Survival curve of *AKA_DDC* versus *AKA_noDDC* cohort. Log-rank test p value is 0.072.
D. Survival curve of *AKA*, *AK*, *AA* mice without DDC treatment.
E. RT-qPCR quantification of *Arid1a* mRNA in *YK* (n=5) and *YKA* (n=4) livers that were infected with AAV8-TBG-Cre viruses. *Arid1a* mRNA is significantly lower in *YKA* liver (0.183+/- 0.046) than YK liver (1.000 +/- 0.153, Student's t-test p = 9.8e-5)
F. YFP/CK19 co-staining images of livers from *Y*, *YK*, *YA* and *YKA* mice that were injected with AAV-TBG-Cre viruses and fed with DDC for 6 weeks. Each image represents a randomly selected field of a different mouse of the specified genotype. The widespread signal of YFP confirmed a high AAV8-TBG-Cre recombination efficiency consistent with RT-qPCR data in **E**. Scale bar 50 µm.
G and H. Cell counts of CK19+ cells (G) and CK19+/YFP+ cells (H) in AAV8-DDC mice (extending **Figure 1H)**.
I. H&E staining showing pathology of *K19KA*, *K19K*, *K19A*, *K19* livers that were not injured when feeding a normal diet (no DDC). Scale bar 200 µm.
J. Rare lesions observed in *K19K*, *K19A*, *K19* livers that were injured when feeding with DDC diet. The top right and bottom right panels show rare expansion of large ducts of *K19* and *K19K* liver (DDC diet); the middle right panel shows rare singular dilated duct in *K19A* liver (DDC diet). Scale bar 200 µm.

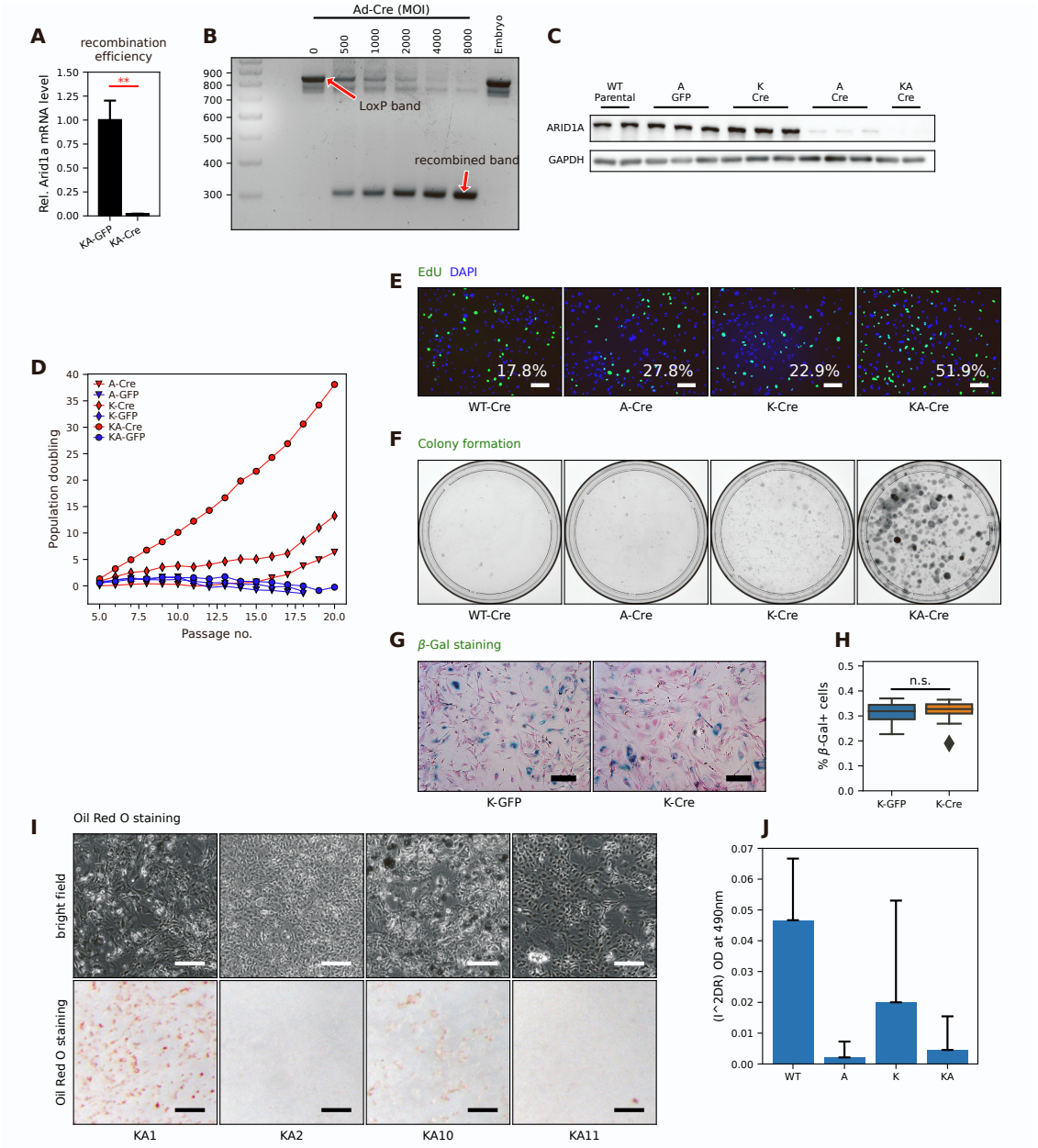


Figure S2. Validation of Cre-recombination efficiency and cancer-associated phenotypes in *Kras/Arid1a* co-mutant MEFs (Related to Figure 2)

A. RT-qPCR quantification of *Arid1a* mRNA in *KA* MEFs after *Ad5-CMV-GFP* and *Ad5-CMV-Cre* infection.
B. PCR products of genomic DNA showing decreased intensity of the LoxP (unrecombined, 845bp) band and increased intensity of the recombined band (298 bp) as increasing multiplicity of infection (MOI) of *Ad5-CMV-Cre* viruses were used to infect the *Arid1a*$^{L/L}$ MEFs.
C. Western blot showing the ARID1A protein levels in MEFs with different genotypes after virus infection (*Ad5-CMV-Cre* vs *Ad5-CMV-GFP*).
D. A separate set of 3T3 experiments accompanying **Figure 2B** showing a similar pattern of population doubling of *A-Cre*, *K-Cre*, and *KA-Cre* MEFs and their controls.
E. Representative images of EdU incorporation and DAPI staining of *WT-Cre*, *A-Cre*, *K-Cre*, and *KA-Cre* MEFs. Scale bar 50 μm.
F. Colony formation assay of recombined MEFs (different cell lines from those in Figure 2D) showing a similar and obvious increase in colonies formed in *KA-Cre* compared to the other three genotypes.
G-H. Representative images (G) and quantification (H) of senescence-associated beta-galactosidase staining in *K-Cre (K* MEFs infected with *Ad5-CMV-Cre* viruses) and *K-GFP* (*K* MEFs infect with *Ad5-CMV-GFP* control viruses, n=3). Scale bar 50 μm.
I. Representative images of bright field (upper panel) and Oil Red O (a dye staining lipids) staining (lower panel) of MEFs of different genotypes after the treatment of I2DR, a cocktail known to drive adipocyte differentiation. Scale bar 50 μm.
J. Quantification of Oil Red O staining for adipocyte differentiation in MEFs of different genotypes.

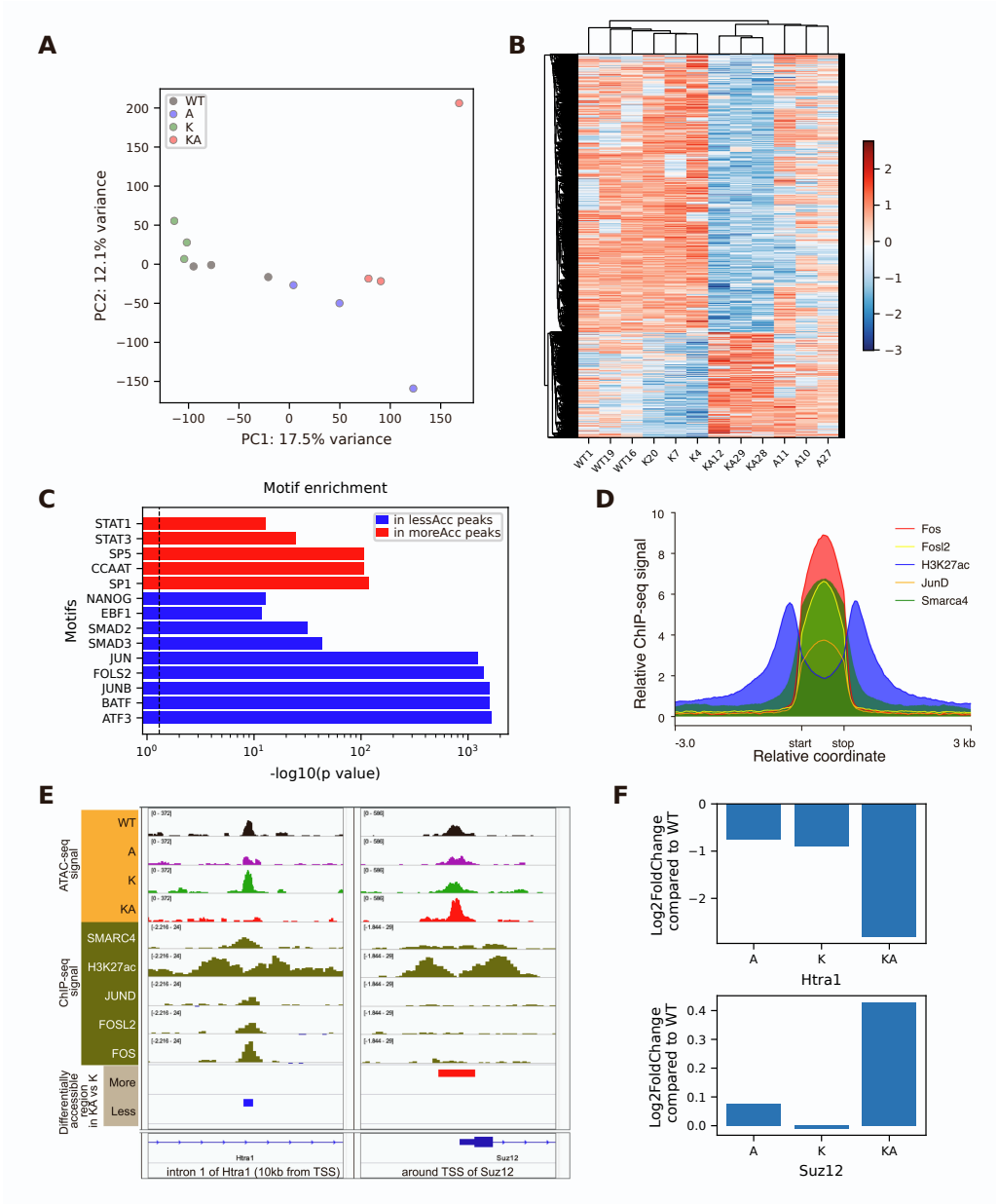


Figure S3. Extended data supporting chromatin accessibility changes induced by *Kras/Arid1a* mutations in MEFs (Related to Figure 2)

A. Principal component analysis of ATAC-seq counts (quantile normalized, log2 transformed) across all ATAC regions in *KA* versus *K* MEFs.

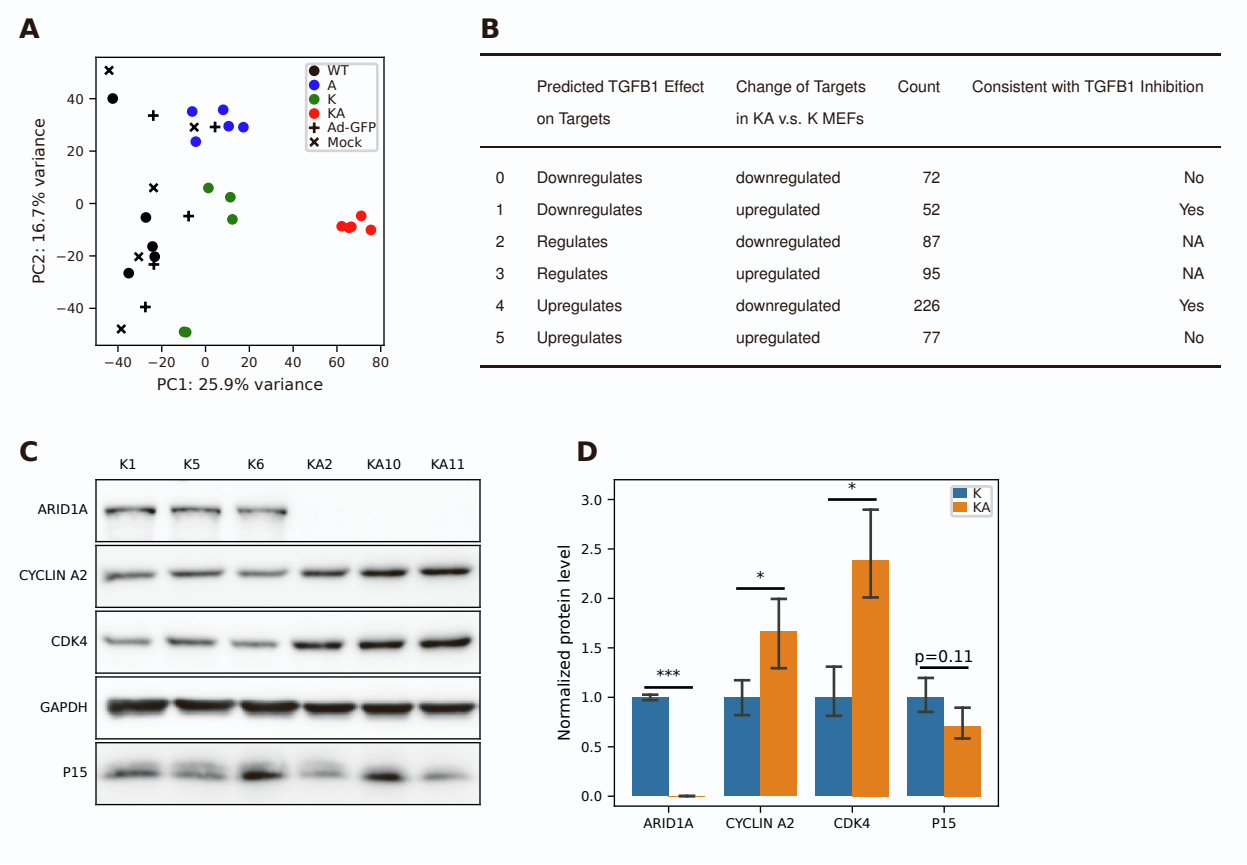
B. Heatmap of row-wise z score of log2 transformed ATAC-seq count matrix for differentially accessible regions (padj<0.05 and log2FoldChange >0.5 or <-0.5). Hierarchical clustering was performed on columns and rows using the Canberra distance metric and average linkage method.

C. Motif enrichment analyses in more accessible peaks and less accessible peaks (in *KA* vs *K* MEFs) via Homer findMotifsGenome.pl. The most significant (with least p values) enrichment items of motifs of interest (indicated as y axis tick labels) were plotted. The vertical dotted line represents a p value equal to 0.05.

D. Publicly available ChIP-Seq data (GSE83295) demonstrating co-location of AP1 factor binding regions with less accessible regions in *KA* versus *K* MEFs.

E. Representative differentially accessible regions aligned with ATAC-seq data and ChIP-seq data as well as nearest genes. *Htra1* represents genes associated with less accessible regions (left), and *Suz12* represent genes with more accessible regions (right).

F. RNA-seq log2FoldChange of *Htra1* and *Suz12* in *A*, *K*, and *KA MEFs* compared to *WT* MEFs.



	Predicted TGFB1 Effect on Targets	Change of Targets in KA v.s. K MEFs	Count	Consistent with TGFB1 Inhibition
0	Downregulates	downregulated	72	No
1	Downregulates	upregulated	52	Yes
2	Regulates	downregulated	87	NA
3	Regulates	upregulated	95	NA
4	Upregulates	downregulated	226	Yes
5	Upregulates	upregulated	77	No

Figure S4. Extended data supporting the dysregulation of cycle cell control and TGFβ pathways in *Arid1a* null *Kras^{G12D}* MEFs (Related to Figure 3)

A. Principal component analysis of RNA-seq data showing the distribution of both recombined MEFs (*WT-Cre*, *A-Cre*, *K-Cre*, and *KA-Cre*) and non-recombined (*Ad5-CMV-eGFP* or Mock infected) MEFs.

B. A summary table showing the majority of TGFB1 targets (provided by Ingenuity Pathway Analysis program) are regulated in a pattern consistent with TGFB inhibition according to the predicted TGFB1 effects on its targets.

C-D. Western blot (C) and its quantification (D) for proliferation-related proteins in *KA* (n=3 cell lines) vs *K* (n=3 cell lines) MEFs. Relative protein band intensity was calculated against GAPDH per lane, which was then normalized to the mean of K MEFs per protein. The expression of each protein in *K* vs *KA* MEFs were compared using *Student's t* test. * $p < 0.5$ and *** $p < 0.001$.

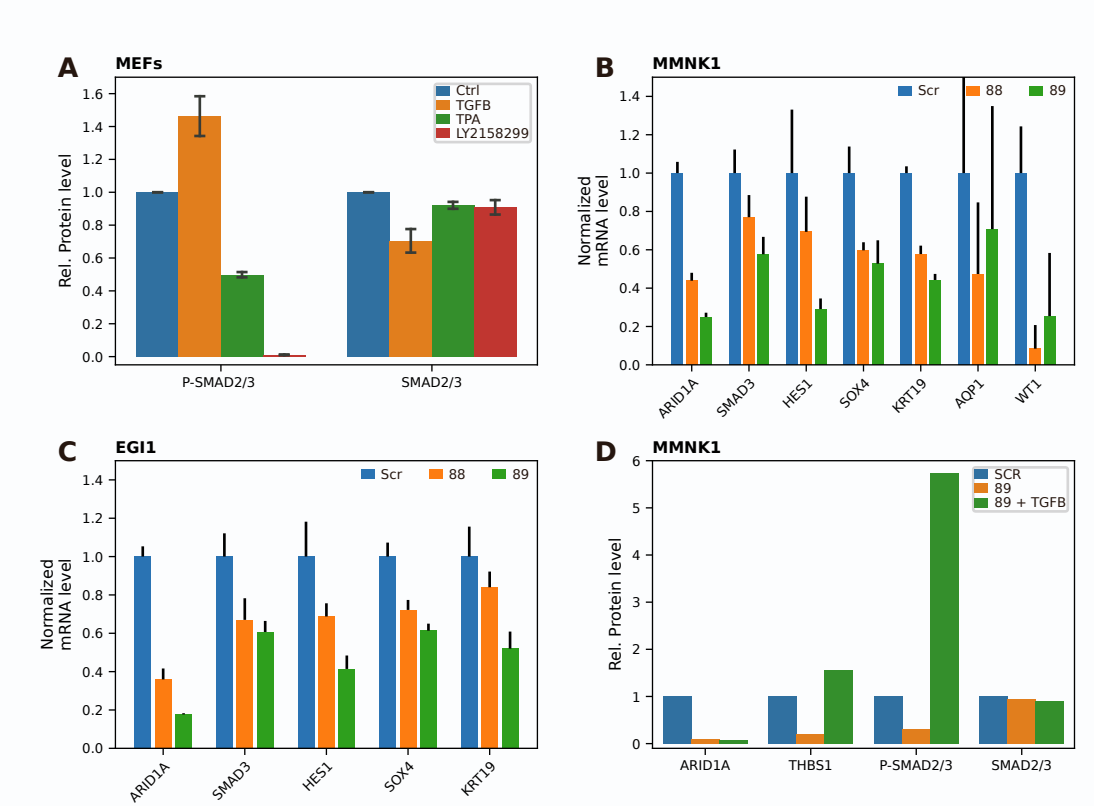


Figure S5. Extended data supporting the effect of ARID1A loss in cholangiocyte differentiation and TGFβ-Smad4 signaling (Related to Figure 5)

A. Quantification of Western blot bands of *KA* MEFs (for **Figure 5B**). The protein levels were normalized to the corresponding controls (GAPDH).
B-C. RT-qPCR of cholangiocyte differentiation markers and key genes of Notch pathway in MMNK1 cells (B) and in human cholangiocarcinoma cell EGI1 (C) with *ARID1A* KD (#88 and #89 knockdown constructs) or without (Scr).
D. Quantification of Western blot bands of MMNK1 *ARID1A* knockdown cells (**Figure 5G**). The protein levels were normalized to the corresponding controls (GAPDH).

Table S2. Top 20 inhibited and activated upstream regulators of differentially expressed genes in KA versus K MEFs from IPA analysis (Related to Figure 3)

Upstream Regulator	Expression Log Ratio	Molecule Type	Predicted Activation State	Activation z-score	p-value of overlap
TGFB1	-0.772	growth factor	Inhibited	-6.568	3.40E-30
CDKN2A	-0.792	transcription regulator	Inhibited	-6.562	3.53E-11
TNF		cytokine	Inhibited	-5.651	2.74E-12
XBP1	-0.33	transcription regulator	Inhibited	-5.517	1.32E-06
IL1A		cytokine	Inhibited	-4.815	4.44E-03
SREBF1		transcription regulator	Inhibited	-4.744	3.29E-02
IL1B		cytokine	Inhibited	-4.604	1.60E-04
NUPR1	-1.028	transcription regulator	Inhibited	-4.431	7.74E-11
TCF7L2	0.316	transcription regulator	Inhibited	-4.314	2.52E-06
TET2		enzyme	Inhibited	-4.251	2.83E-02
TP53	0.205	transcription regulator	Inhibited	-4.209	9.52E-34
IFNG		cytokine	Inhibited	-4.146	9.92E-03
AGT	-2.392	growth factor	Inhibited	-4.137	9.85E-10
TGM2		enzyme	Inhibited	-4.119	8.49E-03
SMARCA4	0.208	transcription regulator	Inhibited	-4.112	5.03E-06
RB1		transcription regulator	Inhibited	-4.064	5.41E-09
IL5		cytokine	Inhibited	-4.019	4.13E-05
TFEB		transcription regulator	Inhibited	-3.977	9.32E-03
ATF4		transcription regulator	Inhibited	-3.92	6.65E-06
CD38		enzyme	Inhibited	-3.916	9.46E-05
NTRK2		kinase	Activated	3.106	1.88E-06
PRKAA1		kinase	Activated	3.114	1.71E-07
PRKAA2		kinase	Activated	3.116	2.62E-06
ZNF106	-0.292	other	Activated	3.207	1.35E-01
BMI1		transcription regulator	Activated	3.214	9.60E-03
NEUROG1		transcription regulator	Activated	3.273	3.10E-02
MTM1	-0.657	phosphatase	Activated	3.274	1.11E-03
AREG		growth factor	Activated	3.372	2.15E-02
miR-122-5p		mature microrna	Activated	3.423	3.67E-02
TBX2	1.737	transcription regulator	Activated	3.5	2.93E-10
AHR		ligand-dependent nuclear receptor	Activated	3.554	1.33E-08
MYC		transcription regulator	Activated	3.573	4.14E-18
E2F1		transcription regulator	Activated	3.587	5.05E-10
MYCN	1.444	transcription regulator	Activated	3.615	6.93E-02
E2f		group	Activated	4.156	1.23E-07
EP400		other	Activated	4.205	3.74E-10
Alpha catenin		group	Activated	4.274	2.45E-07
RABL6		other	Activated	4.342	7.19E-08
26s Proteasome		complex	Activated	4.479	6.08E-04
INSIG1		other	Activated	4.503	2.37E-01

Table S3. Genotyping and RT-qPCR primer sequences (Related to Figure 1, 2, 4 and 5)

Category	Gene or allele	Species	Primer sequence (5'-3')	Comment
Genotyping primer	*Alb-Cre*	Mouse	ATGAAATGCGAGGTAAGTATGG	
			CGCCGCATAACCAGTGAAAC	
	CK19-Cre ^{ERT}	Mouse	AATCGCCAGGAATTGACCAATGGGG	
			CGGCAAACGGACAGAAGCATTTTCC	
			CGCCCGTACCCCCAAAGGAAGACAT	
	Kras $^{LSL-G12D}$	Mouse	TCCGAATTCAGTGACTACAGATG	
			CTAGCCACCATGGCTTGAGT	
			ATGTCTTTCCCCAGCACAGT	
	Arid1a $^{L/L}$	Mouse	GTAATGGGAAAGCGACTACTGGAG	
			TGTTCATTTTTGTGGCGGGAG	
	Rosa26 $^{LSL-eYFP}$	Mouse	GCGAAGAGTTTGTCCTCAACC	
			GGAGCGGGAGAAATGGATATG	
			AAAGTCGCTCTGAGTTGTTAT	
RT-qPCR primer	*ARID1A*	Human	CAGTACCTGCCTCGCACATA	Forward
			GCCAGGAGACCAGACTTGAG	Reverse
	RHOA	Human	TGGAAAGACATGCTTGCTCAT	Forward
			GCCTCAGGCGATCATAATCTTC	Reverse
	SMAD3	Human	TCAACACCAAGTGCATCACC	Forward
			CGGCAGTAGATGACATGAGG	Reverse
	HES1	Human	GAAAGATAGCTCGCGGCATT	Forward
			TACTTCCCCAGCACACTTGG	Reverse
	SOX4	Human	AGCGACAAGATCCCTTTCATTC	Forward
			CGTTGCCGGACTTCACCTT	Reverse
	KRT19	Human	CACCAGCCGGACTGAAGAAT	Forward
			GCAGGTCAGTAACCTCGGAC	Reverse
	AQP1	Human	TAACCCTGCTCGGTCCTTTG	Forward
			AGTCGTAGATGAGTACAGCCAG	Reverse
	WT1	Human	GTGACTTCAAGGACTGTGAACG	Forward
			CGGGAGAACTTTCGCTGACAA	Reverse
	Arid1a	Mouse	ATGGCCAATATGCCACCTCA	Forward
			CCATAGGGAGGTCCAGTTCC	Reverse
	Rhoa	Mouse	AGCTTGTGGTAAGACATGCTTG	Forward
			GTGTCCCATAAAGCCAACTCTAC	Reverse
	Hes1	Mouse	CCAGCCAGTGTCAACACGA	Forward
			AATGCCGGGAGCTATCTTTCT	Reverse
	Tgfb1	Mouse	AGCTGCGCTTGCAGAGATTA	Forward
			AGCCCTGTATTCCGTCTCCT	Reverse