Published in final edited form as:

Nature. 2023 July ; 619(7970): 632–639. doi:10.1038/s41586-023-06254-7.

Histone demethylase KDM5D upregulation drives sex differences in colon cancer

Jiexi Li[1], **Zhengdao Lan**[1], **Wenting Liao**[1,2], **James W. Horner**[3], **Xueping Xu**[3], **Jielin Liu**[4], **Yohei Yoshihama**[1], **Shan Jiang**[3], **Hong Seok Shim**[1], **Max Slotnik**[1], **Kyle A. LaBella**[1], **Chang-Jiun Wu**[5], **Kenneth Dunner Jr.**[1], **Wen-Hao Hsu**[1], **Rumi Lee**[1], **Isha Khanduri**[6], **Christopher Terranova**[5], **Kadir Akdemir**[5,7], **Deepavali Chakravarti**[1], **Xiaoying Shang**[1], **Denise J. Spring**[1], **Y. Alan Wang**[1,8], **Ronald A. DePinho**[1,*]

[1]Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[2]Department of Experimental Research, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, 510060, China.

[3]TRACTION Platform, Division of Therapeutics Discovery, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[4]Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[5]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[6]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[7]Department of Neurosurgery, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[8]Current address: Indiana University, Indianapolis, Indiana, USA

[*]Corresponding author: Ronald A. DePinho, RDePinho@mdanderson.org.

Author Contributions

J.L., Y.A.W. and R.A.D. designed the project and analyzed the data; Z.L. established the iKAP primary and metastatic cell lines, conducted the *in vivo* study in nude mice and helped with the survival analysis in CRC patients; W.L. helped with tumour staging; J.W.H. and X.X. generated the iAP-KDM5D mouse model; J. Liu helped with MHC-I detection, antigen presentation assay and T cell killing assay; Y.Y. helped with immunofluorescence, confocal imaging, paracellular permeability assay and the quantitation of the tight junction integrity; S.J. maintained, monitored and recorded the health condition of the genetically engineered mice; H.S.S. helped with qPCR of FFPE tumour tissue, IP-MS and co-IP experiment; M.S. helped with immunohistochemistry; K.A.L. helped with organoid culture and engineering; C.W. analyzed the RNA-seq data and performed GSEA; K.D. performed the transmission electron microscopy and helped with the quantitation of the tight junction integrity; W.H. established the iKAP tumour organoid lines and helped with animal surgeries; R.L. helped with the luciferase reporter assay; I.K. helped with the tumour staging; C.T. and K.A. helped with the HiChIP experiment and analysis; D.C. helped with organoid culture; X.S. helped with mouse genotyping; J.L. and R.A.D. wrote the manuscript; D.J.S. edited the manuscript and reviewed data.

Competing interests

R.A.D. holds equity as a former advisor and/or director of Tvardi Therapeutics, Asylia Therapeutics, Stellanova Therapeutics and Sporos Bioventures. The remaining authors declare no competing interests.

Supplementary Information is available for this paper.

Abstract

Sex exerts a profound impact on cancer incidence, spectrum and outcomes, yet the molecular and genetic bases of such sex differences are ill-defined and presumptively ascribed to X-chromosome genes and sex hormones[1]. Such sex differences are particularly prominent in colorectal cancer (CRC) where men experience higher metastases and mortality. A murine CRC model, engineered with an inducible transgene encoding oncogenic mutant $KRAS^{G12D}$ and conditional null alleles of *Apc* and *Trp53* tumour suppressors (designated iKAP)[2], revealed higher metastases and worse outcomes specifically in males with oncogenic mutant KRAS (KRAS*) CRC. Integrated cross-species molecular and transcriptomic analyses identified Y-chromosome gene histone demethylase *KDM5D* as a transcriptionally upregulated gene driven by KRAS*-mediated activation of the STAT4 transcription factor. KDM5D-dependent chromatin mark and transcriptome changes showed repression of regulators of the epithelial cell tight junction and MHC class I complex components. Deletion of *Kdm5d* in iKAP cancer cells increased tight junction integrity, decreased cell invasiveness, and enhanced cancer cell killing by $CD8^+$ T cells. Conversely, iAP mice engineered with a *Kdm5d* transgene to provide constitutive *Kdm5d* expression specifically in iAP cancer cells exhibited an increased propensity for more invasive tumours *in vivo*. Thus, KRAS*-STAT4-mediated upregulation of Y chromosome *KDM5D* contributes substantially to the sex differences in KRAS* CRC via its disruption of cancer cell adhesion properties and tumour immunity, providing an actionable therapeutic strategy for metastasis risk reduction for men afflicted with KRAS* CRC.

Introduction

Sex differences in cancer clinical outcomes and tumour biology are evident, with males typically being more adversely affected than females[3]. In addition, metastasis rates[4] and treatment responses vary between males and females[1]. Current perspectives posit that sex differences relate to healthier lifestyle habits in women, such as lower consumption of red and processed meats and alcohol, and sex hormones[5]. However, how sex-specific mechanisms may intersect with genotype is not well understood. In this regard, the Y chromosome contains a small number of protein-coding genes that, except for sex-determination genes, have received limited attention as potential determinants of cancer hallmarks and clinical outcomes.

One of the most prominent sex-specific differences in cancer outcomes is observed in CRC, the second most common cause of cancer mortality which often relates to metastases[6]. The most common genetic alterations in sporadic human CRC include activating mutations of the *KRAS* oncogene and inactivating mutations or deletion of the *APC* and *TP53* tumour suppressors[7]. We have generated a mouse model of CRC engineered with these signature alterations: $VillinCre^{ERT2}$ allele, Lox-stop-Lox (LSL)-inducible and doxycycline-inducible transgene encoding oncogenic $KRAS^{G12D}$ and conditional null alleles of *Apc* and *Trp53* (modelling the most common mutations in patients). This model, designated iKAP, recapitulates the evolution of human CRC from adenoma to adenocarcinoma to invasive and metastatic disease. Moreover, this model established that KRAS* (oncogenic mutant KRAS) drives CRC invasiveness and metastasis[2,8], mirroring another model engineered with $Apc^{\Delta 716}$, $Trp53^{R270H}$ and $Kras^{G12D}$ mutations[9]. Mechanistically, in both male and

female mice, KRAS* drives an epithelial-to-mesenchymal transition (EMT) and progression to metastasis in part through regulation of the TGF-β pathway[2] as well as IRF2 which upregulates CXCL3 to recruit CXCR2-positive myeloid-derived suppressor cells (MDSCs) to suppress anti-tumour immunty[8].

In this study, we observed higher metastasis frequency and shortened survival specifically in male iKAP mice with KRAS*-expressing CRC relative to female mice; and such sex differences mirrored those in humans. These circumstantial observations led to the identification of Y chromosome gene *KDM5D* as a potential basis for sex-specific differences in KRAS* CRC progression. KDM5D is a member of the KDM5 family of H3K4me2/3 demethylases that repress target gene expression[10]. Integrated analyses reveal that KRAS*-STAT4-KDM5D drives invasion and metastases via repression of genes governing cell-cell junction complex integrity and $CD8^+$ T cell-mediated anti-tumour immunity.

Results

KRAS* increases metastasis via KDM5D.

KRAS oncogenic mutations correlate with more invasive tumour stages in both iKAP mice[2] and CRC patients[11–13]. In the iKAP model, KRAS* expression correlates with poorer survival in tumour-bearing males relative to females: median survival of 19 weeks in males versus 25 weeks in females (Fig. 1a). This sex difference was not observed in the iAP model (Fig. 1a). Accordingly, endpoint tumour stage analyses of iAP tumours showed no sex-related difference, whereas tumours from male iKAP mice exhibited more advanced disease than the female iKAP mice (Extended Data Fig. 1a). Correspondingly, Kaplan-Meier survival analysis of integrated CRC data sets (n = 831 patients) by Colorectal Cancer Subtyping Consortium (CRCSC) (https://www.synapse.org, Synapse ID: syn2623706)[14] revealed that male KRAS* CRC patients exhibited poorer survival relative to female KRAS* CRC patients (median survival 86 months in males and not reached in females) (Fig. 1b). Moreover, these sex differences in survival were not present in wild-type (WT) *KRAS* CRC (Fig. 1b). In male CRC patients, KRAS* was associated with a median survival of 86 months versus 112 months in WT *KRAS* controls, whereas *KRAS* status had no impact on outcome in female CRC patients (median survival not reached in KRAS* versus 183 months in WT *KRAS* controls) (Extended Data Fig. 1b). Additional analyses suggested that the sex differences were not related to mutation status of *TP53* or *APC* (Extended Data Fig. 1c). In line with overall survival analysis, disease-free survival (DFS) showed KRAS*-specific worse survival in males than in females (Extended Data Fig. 1d). As second primaries or local recurrences are very rare in CRC, we further analyzed the DFS of patients with initial staging of stage 2 or below, in whom primary tumours are treated with surgery alone and disease recurrence manifests as metastasis. As expected, the DFS also showed a KRAS*-related sex-bias pattern (Extended Data Fig. 1e). Together, these data indicated a key role of KRAS* in driving sex differences in both murine and human CRC.

Women show a tendency to develop proximal CRC, whereas men more often develop distal CRC[15]. As proximal and distal CRC show differences in the timing of *KRAS* mutations and the frequency of associated signature mutations, we further analyzed the generality of

the KRAS*-driven sex differences in proximal and distal CRC across various genotypic contexts. The worse survival in males with KRAS* CRC was readily evident only in proximal CRC (Extended Data Fig. 2a) likely relating to the higher frequency (Extended Data Fig. 2b, c) and more clonal nature of *KRAS* mutations that arise earlier in the evolution in proximal CRC and are more readily detected[16]; in contrast, in distal CRC, the clonal evolution is more complex and divergent, and *KRAS* mutations, in particular, are more subclonal[16] as they occur later and are restricted to the tumour invasive front[2]. Indeed, in KRAS* Stage 1 and 2 distal CRC where *KRAS* mutations have arisen earlier and are more clonal, KRAS* correlates positively with worse DFS only in male patients (Extended Data Fig. 2d). With respect to signature mutations, proximal CRC most frequently presents with 'KAP' (20.16%) and 'KA' (21.71%) versus distal CRC prominent profile of 'AP' (39.94%) (Extended Data Fig. 2e). The sex-related survival difference was specifically observed in 'KAP' patients (Extended Data Fig. 2f), aligning with findings in the iKAP model. Due to the limited sample size, the patients were not further subdivided into proximal and distal CRC groups. Lastly, consistent with iKAP distal CRC (Extended Data Fig. 1a), proximal tumours induced by 4-Hydroxytamoxifen (4-OHT) injection in the proximal colon also showed sex disparities in tumour stages (Extended Data Fig. 2g). Together, these data suggest that KRAS* may drive sex differences in both proximal and distal CRC in mice and humans.

Differential gene expression patterns were audited in primary versus metastatic male iKAP tumours and further intersected with KRAS*-dependent gene expression changes in male iKAP tumours with KRAS* on versus off (Database: SRP097890). *Kdm5d* emerged as the sole Y chromosome gene with differential expression (Fig. 1c). The KRAS*-KDM5D relationship was validated in (i) normal male colonoids from mouse proximal or distal colon (Extended Data Fig. 3a, b), (ii) independently derived male iKAP primary and metastatic cell lines with doxycycline-controlled KRAS* on and off (Fig. 1d), as well as (iii) male CRC patients (Fig. 1e, Extended Data Fig. 3c). Such relationship is independent of WT or $Apc^{null}/Trp53^{null}$ genetic background, as shown in male colonoids (Extended Data Fig. 3d, e). Moreover, $KRAS^{G12C}$ could equally upregulate *Kdm5d*, supporting the centrality of oncogenic KRAS*-KDM5D signalling (Extended Data Fig. 3f). The KDM5D-metastasis relationship was further validated across (i) independently derived male iKAP primary and metastatic cell lines (Fig. 1d, Extended Data Fig. 3g), (ii) organoids established from male iKAP primary and matched spontaneous liver metastatic tumours (Extended Data Fig. 3h–j), (iii) primary versus metastatic CRC tumours in male patients (Fig. 1f), and (iv) male CRC cell lines in the CCLE data set (Extended Data Fig. 3k). Lastly, higher *KDM5D* expression in male patients correlated with worse overall survival (Extended Data Fig. 3l). Of note, the X chromosome paralog KDM5C is not regulated by KRAS* nor differentially expressed in primary versus metastatic disease (Extended Data Fig. 3m–o). Together, these data support the view that KRAS* increases KDM5D levels driving worse outcomes in human and mouse males with KRAS* CRC.

As KRAS* drives invasion and metastasis in CRC[2], we explored the potential role of KDM5D in metastasis and related processes. To functionally validate KDM5D's pro-metastasis activity, CRISPR-Cas9 gene editing was used to delete *Kdm5d* in the male iKAP liver (the most common metastatic site of CRC[17]) metastatic cell line L2–3 (Extended

Data Fig. 4a), which is a matching pair of the male iKAP cell line 3346 isolated from a distal primary CRC tumour. In this model, *Kdm5d* deletion did not affect cell proliferation *in vitro* (Extended Data Fig. 4b) and *in vivo* (Extended Data Fig. 4e), but inhibited cell migration and invasion *in vitro* (Extended Data Fig. 4c, d), and decreased liver metastases and increased survival following intrasplenic injection in nude mice *in vivo* (Extended Data Fig. 4f–l). Conversely, enforced expression of WT *KDM5D*, but not an enzymatic-dead (ED) mutant, in female iKAP distal primary tumour cell line 2806 enhanced migration and invasion *in vitro* and increased liver metastasis *in vivo* (Extended Data Fig. 4m–r). Finally, to drive *Kdm5d* expression in the absence of KRAS*, a *Kdm5d* transgene was introduced into the iAP model (designated iAP-KDM5D). In the iAP-KDM5D model, low dose 4-OHT enema delivery activates *Villin*-directed Cre^{ERT2} resulting in the deletion of *Apc* and *Trp53* and removal of a Lox-stop-Lox (LSL) cassette to enable constitutive *Kdm5d* expression in both male and female mice (Fig. 2a, b). Compared to iAP mice which rarely generate tumours that progress beyond pT2 stage and show no metastasis, enforced *Kdm5d* expression enhanced tumour invasiveness and promoted metastasis to lymph nodes, though not as robustly as KRAS* as expected from its additional mechanisms operative in both male and females[8] (Fig. 2c, d). Together, these results indicate that KDM5D promotes cancer cell migration, invasion and metastasis in CRC.

To determine how KRAS* regulates KDM5D, we first used the JASPAR algorithm[18] to identify putative transcription factor binding elements in the mouse *Kdm5d* and human *KDM5D* promoters, and intersected this list with genes upregulated by KRAS* (Database: SRP097890) (Fig. 3a, b). Among three overlapped transcription factors, we focused on STAT4 as its motif is conserved in mice and humans (Extended Data Fig. 5a) and it is a transcriptional activator[19], whereas mouse MSX3 (human ortholog VENTX)[20,21] and SNAI2[22] are transcriptional repressors. Phospho-STAT4, which is indicative of activated STAT4 signalling[19], colocalized with phospho-ERK, which is indicative of activated KRAS signalling, in both male and female proximal or distal iKAP tumours along the invasive front (Extended Data Fig. 5b). Correspondingly, a tumour microarray of primary tumours from male patients with proximal or distal CRC showed a positive correlation of H-scores of phospho-STAT4 and phospho-ERK in cancer epithelial cells (Extended Data Fig. 5c, d). Like other STAT family proteins[23], STAT4 binding to the genome and the *Kdm5d* promoter decreased by extinction of KRAS* expression (Fig. 3c, d). Luciferase reporter assays demonstrated that mutation of the STAT4 binding element in the *Kdm5d* promoter abolished STAT4-induced upregulation of reporter activity (Fig. 3e) and mutant (phospho-dead) STAT4 decreased reporter activity driven by the *Kdm5d* promoter (Fig. 3f), whereas phosphomimetic STAT4 further upregulated *Kdm5d* expression (Fig. 3g, h) as well as its known downstream genes *Socs1, Socs2* and *Socs3*[24] (Extended Data Fig. 5e, f). The positive correlations between *STAT4*, *SOCS1*, *SOCS2*, *SOCS3* expression and *KDM5D* expression were validated in both proximal and distal human CRC (Extended Data Fig. 5g–i). Lastly, the synchronous increase of phospho-ERK1/2, phospho-STAT4 and total STAT4 signals in male metastatic cancer cells relative to primary tumour-derived cancer cells further supports the role of KRAS*-STAT4-KDM5D in CRC metastasis (Fig. 3i).

KDM5D disrupts cell-cell junctions.

To identify the downstream targets of KDM5D involved in metastasis, we profiled KDM5D-dependent changes in H3K4me2/3 marks and gene expression. ChIP-seq analyses of male iKAP primary cells, metastatic cells and *Kdm5d−/−* metastatic cells showed decreased H3K4me2/3 marks in iKAP *Kdm5d+/+* metastatic cells and restoration of these marks in isogenic *Kdm5d−/−* controls (Fig. 4a, b). In addition to its classical activity in the demethylation of H3K4me2/3, KDM5D expression was associated with increased H3K4me1 marks (Extended Data Fig. 6a). Next, these chromatin profiles and RNA-seq expression profiles were intersected to identify genes showing (i) repressed expression in iKAP metastatic cancer cells relative to primary cancer cells, (ii) de-repressed expression in isogenic iKAP *Kdm5d−/−* metastatic cells and (iii) altered H3K4me2/3 marks with pattern similar to expression changes in (i) and (ii) (Fig. 4c). This analysis yielded several genes essential in the maintenance of cell-cell junctions including *Amot* and *Pkp1*, regulating tight junctions[25] and desmosomes[26], respectively. Here we focused on the topmost differentially expressed gene, *Amot*, a gene encoding angiomotin, which is essential in maintaining tight junctions[25].

Loss of cell-cell junctions can result in cancer cell dissemination, a key step in early metastasis. *Amot* was downregulated in iKAP metastatic cancer cells and de-repressed by *Kdm5d* deletion, upregulated in human CRC metastatic cell LoVo by *KDM5D* deletion, downregulated in organoids derived from metastatic iKAP tumours and in metastatic tumours from CRC patients (Fig. 4d and Extended Data Fig. 6b–e). Additionally, in iKAP metastatic cells, *Amot* showed decreased H3K4me2/3 marks in its promoter (Extended Data Fig. 6f). Negative correlation of *AMOT* and *KDM5D* expression was present in tumours from male CRC patients (Extended Data Fig. 6g); and, enforced *Kdm5d* expression in colonoids repressed *Amot* expression across iKAP, iAP, $Kras^{G12D}$ single mutation (‘K’) and wild-type (‘WT’) genotypes (Extended Data Fig. 6h). Enforced *Amot* expression in metastatic iKAP cancer cells did not affect cell proliferation *in vivo* (Extended Data Fig. 6i) but reduced migration *in vitro* (Fig. 4e, f) and metastasis *in vivo* and extended animal survival (Extended Data Fig. 6j–n), findings consistent with its metastasis suppressor function.

AMOT recruits intracellular ZO-1 (zonula occludens 1) to the cytoplasmic membrane surface[27,28] and connects to multiple transmembrane junction proteins, such as claudins[29]. Accordingly, cytoplasmic membrane levels of claudin-1 and ZO-1 were both lower in metastatic iKAP cells than in primary cells and were higher in metastatic cells either null for *Kdm5d* or with enforced *Amot* expression (Extended Data Fig. 6o and Fig. 4g). Furthermore, the integrity of tight junctions, assessed via dextran permeability assay measuring dextran-FITC transit across cultured cell monolayers, showed increased leakiness in metastatic relative to primary iKAP cancer cell monolayers; and *Kdm5d* deletion partially restored tight junction integrity in metastatic cancer cell monolayers (Extended Data Fig. 6p, q). Gene Set Enrichment Analysis (GSEA) affirmed that the apical junction signalling was weaker in metastatic cells and rescued by *Kdm5d* deletion (Extended Data Fig. 6r). Finally, transmission electron microscopy of iAP, iAP-KDM5D and iKAP tumours revealed that the tight junctions between epithelial cells were (i) more intact in female than male

iKAP tumours but similar in iAP and (ii) were impaired by *Kdm5d* transgene expression or KRAS* expression, as indicated by the number of 'kissing points' (fusion of outer leaflets of adjoining cell membranes)[30] at the tight junction (Fig. 4h, i). Notably, *AMOT* expression was lower in the human male KRAS* CRC cell lines compared with females (Extended Data Fig. 6s). Together, these data support the view that AMOT serves as a key downstream effector of KDM5D-driven CRC metastasis in males.

KDM5D inhibits MHC-I antigen presentation.

In addition to KDM5D's canonical function of demethylating H3K4me2/3, ChIP-seq data showed that *Kdm5d* deletion increased H3K27ac, an active enhancer mark (Extended Data Fig. 7a, b). Application of the ROSE algorithm for super-enhancer calling[31,32] revealed a decrease in super-enhancers in metastatic iKAP cells (268 putative super-enhancers) compared with primary cells (571 putative super-enhancers) and restoration of super-enhancers by *Kdm5d* deletion in metastatic cells (674 putative super-enhancers) (Extended Data Fig. 7c). To identify the genes regulated by these super-enhancers, Hi-C chromatin immunoprecipitation (HiChIP) for high-hierarchy chromatin interaction profiling was performed to capture the long-range enhancer-gene interactions associated with H3K27ac[33]. HiChIP revealed increased number and activity of cis long-range chromatin-interactions mediated by H3K27ac in metastatic cells null for *Kdm5d* compared with WT controls (16% in null versus 8% in WT) (Extended Data Fig. 7d, e). The intersection of the interaction pairs and the putative super-enhancer regions that were decreased in metastatic cells and re-appeared after *Kdm5d* deletion identified the super-enhancer region of the MHC gene cluster as one of the most dynamic regions (Extended Data Fig. 7c and Supplementary Table 1). Affirmation of this super-enhancer was reinforced by the co-existence of clustered H3K27ac and H3K4me1 histone mark signals at the predicted super-enhancer region (Extended Data Fig. 7f), the sequence of which has 71.08% high cross-species conservation in DNA sequence between mouse and human (at corresponding HLA gene cluster on chromosome 6).

Activation of an EMT and immune signature correlates strongly with metastasis in KRAS* CRC[2,8]. Impaired tumour antigen processing and presentation (e.g., loss of MHC class I) is a major escape mechanism of tumour immune recognition and $CD8^+$ cytotoxic T cell-mediated tumour elimination[34]. The intersection of murine MHC cluster genes downregulated in metastatic cells, restored by *Kdm5d* deletion, and possessing a human HLA class I gene ortholog identified five genes, *Tap1*, *Tap2*, *Pfdn6* (*H2-Ke2*), *H2-T23*, and *H2-Ke6* (Fig. 5a). Among these genes, *Tap1*/*TAP1* and *Tap2*/*TAP2* were validated to be uniformly altered in multiple cell lines and organoid systems (Fig. 5b and Extended Data Fig. 8a–d) and in various genetic contexts (Extended Data Fig. 8e). *TAP1* and *TAP2* are antigen processing genes and encode essential transporters of cytosolic endogenous peptides to the endoplasmic reticulum where the peptides are attached to the MHC class I complex and later transported to the cell surface for antigen recognition by $CD8^+$ cytotoxic T cells[35,36]. TAP1 downregulation or loss has been observed in all human tumour types, including 11% of CRC carcinomas, whereas TAP1 levels are maintained in adenomas and normal mucosa[37]. In human CRC, slightly though not significantly lower *TAP1* and *TAP2* expression was observed in liver metastasis compared to primary tumours (Extended Data

Fig. 8f, g); and lower *TAP1* and *TAP2* expression correlated with worse survival (Extended Data Fig. 8h, i).

Tap1 and *Tap2* upregulation associated with *Kdm5d* deletion in iKAP cells would be expected to increase antigen presentation and cancer cell death induced by $CD8^+$ T cells. To test this possibility, we performed several immune-relevant functional studies. First, analysis of KDM5D-dependent MHC-I changes documented that *Kdm5d* deletion increased cell surface MHC-I levels which were further enhanced by interferon-gamma (IFNγ) treatment (Fig. 5c). Second, antigen presentation assays using ovalbumin (OVA) protein-treated iKAP cells showed that *Kdm5d*-null metastatic cells were able to process and present more OVA 257–264 (SIINFEKL) peptides through cell surface MHC-I relative to *Kdm5d*-intact controls, indicating enhanced antigen processing and presentation after elimination of *Kdm5d* (Fig. 5d). Third, in T cell killing assays using purified splenic OT-1 $CD8^+$ T cells previously stimulated with OVA 257–264 peptide, OVA protein-treated *Kdm5d*-null metastatic cancer cells were found to be more susceptible to $CD8^+$ T cell-mediated cell death than *Kdm5d*-intact metastatic cancer cell controls, which was further enhanced by IFNγ treatment (Fig. 5e). Fourth, we observed a reduced number of $CD8^+$ T cells at the tumour invasive front in iAP-KDM5D and iKAP mice, compared to iAP controls (Fig. 5f, g), a profile consistent with the established correlation between tumour metastasis and a less intact $CD8^+$ T cell barrier at the invasive margin[38]. GSEA of TCGA male CRC patients revealed enrichment of the stimulated $CD8^+$ T cell signature in *KDM5D*-low patients (Fig. 5h). Together, these findings reveal a novel non-canonical function of KDM5D in promoting immune evasion by impairing antigen processing and presentation. Along with similar data in human CRC metastasis[39], these findings suggest that this non-canonical function of KDM5D may serve to promote metastasis by impairing antigen processing and presentation in cancer cells.

On the basis of the report that KDM5D can bind the ZMYND8 co-repressor which reads the dual histone mark H3K4me1-H3K14ac[40], we speculated that, in CRC, KDM5D may have binding partners that could erase H3K27ac. To identify such binding partners, we performed KDM5D immunoprecipitation (IP) and mass spectrometry (MS), revealing that SAP18 (Sin3A Associated Protein 18) interacted with KDM5D in the human CRC cell line LoVo (Supplementary Table 2). This interaction was validated in HEK293T cells (Extended Data Fig. 9a). SAP18 is a subunit of Sin3-HDAC complex which includes histone deacetylases HDAC1 and HDAC2 and represses gene transcription[41–43]. Pull-down assays confirmed that HDAC1 and HDAC2 complexes with KDM5D (Extended Data Fig. 9a). The KDM5D and SAP18 interaction was dependent on the intact enzymatic domain of KDM5D (Extended Data Fig. 9b). Accordingly, the downstream genes of KDM5D – *Amot*, *Tap1*, *Tap2* – were regulated by WT but not enzymatic-dead KDM5D (Extended Data Fig. 9c). Together, these findings are consistent with the model that KDM5D downregulates H3K27ac via its interaction with SAP18 in the Sin3-HDAC complex.

Discussion

This study identified molecular and cellular mechanisms by which Y chromosome gene *KDM5D* drives male-specific metastasis and worse outcome in CRC harboring KRAS*.

KRAS* activates STAT4 to increase histone demethylase KDM5D which represses the expression of genes governing cell adhesion and immune recognition (Extended Data Fig. 10). Enforced *Kdm5d* transgene in the iAP background can promote invasion in the absence of KRAS*. The murine molecular and phenotypic findings mirror those in human proximal and distal CRC, providing a molecular biological basis for the sex disparities in human KRAS* CRC.

Cancer cell adhesion properties and tumour immunity both govern tumour metastasis. Our study showed that KDM5D impairs cancer cell tight junction structure by epigenetic repression of the key tight junction gene *AMOT*. The tight junction sustains intercellular adhesion to maintain epithelial tissue integrity and cell polarity and contributes to the regulation of intercellular signalling. Breakage of the tight junction, which can occur within a cluster of cancer cells, results in cancer cell dispersion and metastasis[44]. Besides *AMOT*, KDM5D may also repress *PKP1*, encoding a component of the desmosome junction, the regulation of which warrants additional study. With respect to anti-tumour immunity, our analysis of KDM5D-mediated repression of MHC-I and TAP1/2 uncovered a non-canonical function of KDM5D in the regulation of histone acetylation via its interactions with Sin3-HDAC complex and in the regulation of super-enhancer activities. KDM5D-Sin3-HDAC removes H3K27ac at MHC gene cluster super-enhancer to promote immune evasion. Cancer metastasis shows lower HLA class I expression compared with primary tumour tissue[45], and relates to T cell-mediated immune selection of the heterogeneous primary tumour cancer cells[46,47]. These correlations align with KDM5D's functions in suppressing MHC class I antigen presentation and $CD8^+$ T cell-mediated killing, and with promoting tumour invasion and metastasis. Notwithstanding the findings, further study is needed to directly link KDM5D-mediated immune modulation to the rate of metastasis. In accordance with our findings, it is notable that HDAC inhibitor treatment has been shown to increase TAP1 expression and increase $CD8^+$ T cell-mediated killing *in vitro* and *in vivo*[48], suggesting that increased H3K27ac enhancer marks either by KDM5D deletion or HDAC inhibitor treatment may enhance antigen processing/presentation and tumour immunogenicity. Together, these mechanistic insights provide a framework for the development of sex-specific precision medicine strategies for the treatment of men afflicted with metastatic CRC.

Methods

Mouse models.

The iAP and iKAP CRC mouse models were established and described previously[2]. For the iAP-KDM5D mouse model, the CAGGS-LSL-Luc,TdT,*Kdm5d*-pA-Frt-Hyg-Frt targeting construct was introduced into iAP mouse embryonic stem cells by electroporation. Hygromycin-resistant clones were picked and genotyped, and two independent lines were injected into C57BL/6 blastocysts and implanted into surrogate mothers, yielding four high-percentage chimeras. Germline transmission was confirmed by crossing the chimeras to C57 albino females. The C57BL/6J (B6), C57BL/6-Tg (TcraTcrb)1100Mjb/J (OT-1) mice and the NU/J (nude) mice were purchased from the Jackson Laboratory. The iAP, iKAP,

iAP-KDM5D mice of both sexes, B6 mice of both sexes, male OT-1 mice, and nude mice of both sexes matching the sexes of the cell lines injected were used.

Animal experiments.

All mouse manipulations were reviewed and approved by MD Anderson Cancer Center's Institutional Animal Care and Use Committee (IACUC). All animals were maintained in pathogen-free conditions and cared for in accordance with policies and certification of the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC International). Housing rooms were maintained on a 12:12-h light:dark cycle (lights on at 6 am, lights off at 6 pm) with humidity ranging from 30% to 70%. Temperature ranged from 20 to 24 °C (68 to 76 °F).

For intrasplenic injection, 8-week-old nude mice were anesthetized with isoflurane. The spleen was exteriorized through a left lateral flank incision. Cells resuspended in growth factor-reduced Matrigel (Corning) were injected using a 27G needle. The injection site was pressed with a cotton stick for one minute after withdrawing the needle to minimize cell leakage. The peritoneum was closed with suturing and skin was closed with wound clipping.

For subcutaneous tumour cell inoculation, 8-week-old nude mice were anesthetized with isoflurane. Cells were resuspended in growth factor-reduced Matrigel (Corning) and injected subcutaneously to the right flank using a 27G needle. Tumour size was estimated with the formula: volume = width × width × length / 2.

For tumour detection using luminescence, each mouse was injected with 1.5 mg D-Luciferin intraperitoneally (100 μl) and imaged using the IVIS Spectrum Imaging System (Perkin Elmer) after 2 minutes. Images were acquired and analyzed using Living Image 4.3 software (Perkin Elmer) and quantified using Aura (Spectral Instruments Imaging) (version 4.0.8).

Spontaneous tumourigenesis was induced by administering 4-Hydroxytamoxifen (4-OHT, 1 mg/ml) through enema (50 μl 4-OHT) or proximal colon injection after laparotomy (20 μl 4-OHT) and/or doxycycline water (2 mg/ml, for iKAP mice) as previously described[2] to genetically engineered mice at the age of 7–15 weeks old. Mouse health was observed daily.

Regarding experimental endpoints, the max tumour size for subcutaneous tumours was a linear measurement of 2 cm at the largest diameter. For mice undergoing intrasplenic or spontaneous tumourigenesis induction, a loss of 20% of body weight, distress, hypothermia, bowel obstruction, severe hunched posture or inactivity were considered potential endpoints. Mice were humanely sacrificed according to IACUC Euthanasia guidelines. Endpoint criteria were strictly followed in all animal experiments.

The number of mice per group (minimum 3) was decided based on previous experience[2], common practice in the field, animal welfare guidelines and availability of animals, while minimizing the use of animals in accordance with animal care guidelines from MD Anderson Cancer Center's Institutional Animal Care and Use Committee (IACUC) and the National Institutes of Health.

Age-matched nude mice were randomized into each group. For genetically engineered mouse models, mice with matched ages were used based on the genotypes. Tumour growth curves and tumour staging statistics were not plotted until all the data points were collected and analyses were performed by three independent investigators blinded to group information.

Histopathology analysis and immunohistochemistry.

Mouse tumours were sectioned crossing the multi-layers of the intestine. Tissues were fixed in 10% neutral buffered formalin, embedded in paraffin, and sent to Science Park Research Histology, Pathology & Imaging Core or DVMS Veterinary Pathology Services Core at The University of Texas MD Anderson Cancer Center for sectioning, hematoxylin and eosin (H&E) staining, and scanning using Aperio ScanScope (Leica) at 20X. The scanned images were viewed using Aperio ImageScope (Leica) software (version 12.4.6.5003). For CRC tumour staging, blinded histology assessments were performed by three independent pathologists following staging criteria previously described[2]. The tumour area in liver in the intrasplenic injection experiment was measured with Fiji ImageJ (version 2.1.0). IHC was performed following an established protocol[8]. Briefly, the slides were baked at 60 °C, deparaffinized and rehydrated by incubating in a gradient of xylene, ethanol and water. Antigen unmasking was done using Citrate-Based Antigen Unmasking Solution (Vector Laboratories) in a pressure cooker at 95 °C for 30 minutes and 115 °C for 10 seconds. Slides were incubated in 3% hydrogen peroxide to quench endogenous peroxidases. Non-specific signal was blocked by Rodent Block M (Biocare Medical) and stained with primary antibody overnight at 4 °C. Primary antibodies include anti-CD8 (Cell Signaling Technology, Cat#98941, 1:500), anti-phospho-STAT4 (Ser721) (Invitrogen, Cat#PA5–64562, 1:200), and anti-phospho-ERK1/2 (Cell Signaling Technology, Cat#4370, 1:500) antibodies. On the next day, antibody binding was detected by DAB Quanto (Epredia) and slides were dehydrated and mounted. Images were taken on a Leica DM 1000 microscope. The staining intensity and quantity were measured in Fiji ImageJ software (version 2.1.0). The phospho-STAT4 and phospho-ERK H-scores were calculated using formula: H-score = 1 × (% cells of score 1+) + 2 × (% cells of score 2+) + 3 × (% cells of score 3+).

RNA isolation and real-time quantitative PCR.

RNA from cell lines was isolated using the RNeasy Kit (Qiagen). RNA from FFPE slides was isolated with High Pure FFPE RNA Micro Kit (Roche). RNA was reverse transcribed using a 5x All-In-One RT MasterMix (Applied Biological Materials Inc.). Real-time qPCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) using the 7500 Fast Real-Time PCR System (Applied Biosystems). PCR product specificity was confirmed with a melting-curve analysis. Relative mRNA expression was calculated using the 2-ΔΔCt method. Relative expression was calculated by normalizing to *ACTB* for human and *Actb* for mouse unless specified in the figure legend. The primer sequences are: Mouse *Kdm5d* F: GCTGCTTGTGGTTGGAATCTC; R: AATGCTGAAAACACCATGCCC. Mouse *Actb* F: GGCTGTATTCCCCTCCATCG; R: CCAGTTGGTAACAATGCCATGT. Mouse *Gapdh* F: AGGTCGGTGTGAACGGATTTG; R: TGTAGACCATGTAGTTGAGGTCA. Mouse *Stat4* F: TGGCAACAATTCTGCTTCAAAAC; R: GAGGTCCCTGGATAGGCATGT. Mouse

Socs1 F: CTGCGGCTTCTATTGGGGAC; R: AAAAGGCAGTCGAAGGTCTCG. Mouse *Socs2* F: AGTTCGCATTCAGACTACCTACT; R: TGGTACTCAATCCGCAGGTTAG. Mouse *Socs3* F: ATGGTCACCCACAGCAAGTTT; R: TCCAGTAGAATCCGCTCTCCT. Mouse *Amot* F: CCGCCAGAATACCCTTTCAAG; R: CTCATCAGTTGCCCCTCTGT. Mouse *Tap1* F: CCAACGTGTGCGAGTCCATTA; R: AGAGTGAGGTACGGTGACCC. Mouse *Tap2* F: CTGGCGGACATGGCTTTACTT; R: CTCCCACTTTTAGCAGTCCCC. Mouse *H2-Ke2* F: GAAGCACAGCTAACGGAAAACA; R: TGATGTAGTCTAGCCTCTTCCCT. Mouse *H2-T23* F: ACAGTCCCGACCCAGAGTAG; R: CCACGTAGCCGACAATGATGA. Mouse *H2-Ke6* F: GCCTTCCAAGCGGATGTGT; R: AGATGGCGGGCGAGAAAAG. Mouse total *Kdm5d* (transgene + wild-type): F: TGGTGGGAATGGGACACCAAGTTCCT; R: TGTTGGCGGGATTTTTCCAGCTG. Human *KDM5D* F: TAACACACACCCGTTTGACAA; R: GCTGCTGAACTTTGAAGGCTG. Human *ACTB* F: CATGTACGTTGCTATCCAGGC; R: CTCCTTAATGTCACGCACGAT. Human *AMOT* F: ACAGGGATCTGAGAGAGCGT; R: ATTGGTAGAACGGGCAGTGG. Human *TAP1* F: CGCCTCACTGACTGGATTCTA; R: TCTGTTGGAAAAACTCCGTCTC. Human *TAP2* F: TGGACGCGGCTTTACTGTG; R: GCAGCCCTCTTAGCTTTAGCA.

RNA-seq.

RNA from cell lines was isolated as described above. RNA-seq was performed by the Advanced Technology Genomics Core at The University of Texas MD Anderson Cancer Center. Libraries were generated using Illumina's TruSeq Kit and were sequenced using the Illumina HiSeq2000 Sequencer. The NextSeq RTA (version 2.4.11.0) was used for base calling and bcl2fastq (version 2.20.0) was used to convert the .bcl files into .fastq files. Fastq files were aligned to the mouse reference genome (mm9) using HISAT2 (version 2.1.0) and assembled using StringTie (version 2.1.1). R (version 3.5.2) and R package DESeq2 (version 1.32.0) was used for data normalization and differential expression analysis. Each cell line was sequenced in triplicate.

Cell line generation and plasmids.

2806 and 3346 were derived from a female and male, respectively, iKAP primary distal CRC using Mouse Tumour Dissociation Kit (Miltenyi Biotec) followed by GFP+ sorting. Cells were cultured in DMEM with 10% tetracycline-free Fetal Bovine Serum (FBS), 1X penicillin/streptomycin (Invitrogen) and 1 μg/ml doxycycline. Next, cell line 3346 were injected into the spleen of male nude mice, the liver metastatic tumours were collected, and the liver metastatic GFP+ iKAP cells were sorted and cultured under the same conditions, named 'L2–1'. L2–1 cells were re-injected into the spleen of male nude mice and liver metastatic tumours were re-collected. The GFP+ metastatic iKAP cells were generated with the same procedure as above and named 'L2–3'. The human metastatic CRC cell line LoVo (Cat#CCL-229) was purchased from ATCC (authenticated by ATCC) and cultured with the recommended conditions.

The mouse *Kdm5d* and human *KDM5D* genes were knocked out in mouse iKAP tumour cell lines L2–1 and L2–3 and in the human CRC cell line LoVo, respectively. Knockout lentiviral plasmids for both mouse *Kdm5d* and human *KDM5D* were purchased from

Applied Biological Materials Inc. (Cat#2548411). Overexpression plasmids containing Flag-tagged wild-type or enzymatic-dead human *KDM5D* were generously gifted by Dr. Min G. Lee at The University of Texas MD Anderson Cancer Center. Mouse wild-type *Kdm5d* expression plasmid was purchased from OriGene (Cat#MR223231L4). Mouse enzymatic-dead *Kdm5d* expression plasmid was produced by GenScript by converting amino acid sequence 'WHIED' to 'WAIAD' in the enzymatic domain[49,50]. Mouse *Amot* expression plasmid was purchased from Applied Biological Materials Inc. (Cat#LV477274). Wild-type mouse *Stat4* expression plasmid was purchased from OriGene (Cat#MR227035L3). Phosphomimetic STAT4 mutagenesis on S721 and S723 was performed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturer's protocol.

Lentivirus containing the above plasmid was produced in HEK293T (ATCC, Cat#CRL-3216, authenticated by ATCC) by Lipofectamine 2000 (Thermo Fisher Scientific) transfection, and used to infect iKAP or LoVo cells. Single-cell clones were selected, and knockout or overexpression was confirmed by qPCR.

All iKAP tumour cell lines were derived from iKAP mice that were genotyped before performing experiments and were verified by regularly checking morphology. KRAS mutation in iKAP cells was validated with GFP+ flow sorting. LoVo and HEK293T cells were authenticated by ATCC and regularly checked by morphology. All cell lines were confirmed mycoplasma-free. The iKAP cell lines will be made available by the corresponding author under MTA.

Organoid generation and gene overexpression.

For tumour organoids, primary and metastatic tumours from a male iKAP mouse were collected and dissociated with Mouse Tumour Dissociation Kit (Miltenyi Biotec). GFP+ positive cells were sorted and cultured in growth factor-reduced, phenol-free Matrigel (Corning) in medium described below.

To isolate normal colonic crypts for organoid culture, a small piece of the proximal or distal large intestine was incubated in HBSS containing 5 mM EDTA at 4 °C for 45 minutes on a shaker. Incubated colon pieces were shaken vigorously to release crypts and filtered through a 100 uM filter. Crypts were washed and spun down at 300 x g, washed again, and spun again at 180 x g to enrich for intact crypts. Crypts were resuspended in Matrigel (Corning) and plated in 30 to 50 μL drops.

Colonic stem cell organoids were cultured in Advanced DMEM/F12 (Gibco) containing 50% Wnt3a conditioned media (ATCC, Cat#CRL-2647), 10% R-spondin conditioned media (cell line obtained from Dr. Calvin Kuo, Stanford University), 5% Noggin conditioned media (cell line obtained from Dr. Vanesa Muncan and Dr. Gijs van den Brink[51], Academic Medical Center, the Netherlands), 50 ng/mL EGF (Thermo Fisher Scientific, Cat#PMG8041), B27 (Gibco, Cat#17504044), N2 (Gibco, Cat#17502048), 1 mM N-acetylcysteine (Sigma, Cat#A7250), and 10 mM nicotinamide (Sigma, Cat#N3376). *Apc*-null colonic stem cells were cultured in the above media without Wnt3a or R-spondin. Apc^{null};*Kras*-mutant colonic stem cells were cultured in the above media without Wnt3a, R-spondin, and EGF[52].

Lentivirus containing mouse *Kras*^{G12D} overexpression plasmid[8] was produced in HEK293T by Lipofectamine 2000 (Thermo Fisher Scientific) transfection. Retrovirus containing mouse *Kras*^{G12C} (Addgene, Cat#64376) was produced in HEK293T by FuGENE (Promega) transfection.

To overexpress genes in organoids, organoids were dissociated with TrypLE (Gibco) for 5 minutes at 37 °C, washed with HBSS (Gibco) with FBS and spun down at 400 x g. The organoid pellets were resuspended in Advanced DMEM/F12 (Gibco) medium supplemented with 10 μM Y-27632 (EMD Millipore) and mixed with lentivirus or retrovirus in ultra-low attachment plate (Corning), spun at 600 x g for 60 minutes at 32 °C. The cells were incubated at 37 °C for 4 hours for recovery and seeded in Matrigel (Corning).

All iKAP tumour organoids were derived from iKAP mice that were genotyped before performing experiments and were verified by regularly checking morphology. KRAS mutation in iKAP organoids was validated with GFP+ flow sorting. Normal mouse colonoids were checked by western blot for mutations and related pathway activation/deactivation (that is, KRAS(G12D), pERK, p53, p21 and active β-catenin). All organoids were confirmed mycoplasma-free. The iKAP organoids will be made available by the corresponding author under MTA.

Migration and invasion assays.

For 2D assays, cells were plated on the upper chamber of the Transwell with pore size of 8 μm (Corning, coated with Matrigel for invasion assay) in culture media without FBS. For 3D organoid invasion assay, organoids were dissociated into single-cell clones and mixed with Matrigel before adding to upper chamber of the Transwell. Conditioned media with 20% FBS was added to the lower chamber of the Transwell. Cells were cultured for 24 hours and were fixed and stained with crystal violet with 20% methanol. Cells at the upper chamber were removed with a cotton swab. Images were taken and numbers were counted in Fiji ImageJ (version 2.1.0) for the cells attached to the bottom side of the Transwell membrane.

Proliferation assay.

Cells (5×10^4) were plated in 48-well plates. Cells were incubated in IncuCyte (Sartorius) and images were taken every 4 hours for 84 hours. Cell confluency was calculated in the IncuCyte software ZOOM 2018A.

Western blotting and membrane fractionation.

For Western blot analyses, cells were lysed on ice using RIPA buffer (Boston BioProducts) supplemented with protease and phosphatase inhibitors (Roche). Cell lysate was sonicated for 10 cycles with 30 seconds on and 30 seconds off using Biorupter (Diagenode). After spinning down, protein extracts were collected from the supernatant and concentrations were measured with the Pierce BCA Protein Assay kit (Thermo Fisher Scientific) on CLARIOstar plate reader (BMG LABTECH). Protein solutions were mixed with NuPAGE LDS Sample Buffer (4X) (Invitrogen) with 10% β-mercaptoethanol and boiled at 95 °C for 10 minutes before loading onto NuPAGE 4–12% Bis-Tris gels (Thermo Fisher Scientific). Precision Plus Protein Dual Color Standards (BIO-RAD) were used. The primary

antibodies used include anti-ERK (Cell Signaling Technology, Cat#9102, 1:1000), anti-phospho-ERK1/2 (Cell Signaling Technology, Cat#4370, 1:1000), anti-STAT4 (Invitrogen, Cat#714500, 1:500), anti-phospho-STAT4 (Ser721) (Invitrogen, Cat#PA5–64562, 1:1000), anti-β-actin (Sigma, Cat#A2228, 1:2000), anti-non-phospho (active) β-catenin (Cell Signaling Technology, Cat#8814, 1:1000), anti-total β-catenin (Cell Signaling Technology, Cat#9587, 1:1000), anti-Ras (G12D) (Cell Signaling Technology, Cat#14429, 1:1000), anti-vinculin (Sigma, Cat#V9131, 1:500), anti-p53 (Cell Signaling Technology, Cat#32532, 1:1000), anti-claudin-1 (Abcam, Cat#ab15098, 1:200), anti-p21 (Abcam, Cat#ab188224, 1:1000) and anti-NaK ATPase (Abcam, Cat#ab76020, 1:100000) antibodies. The secondary antibodies used include anti-rabbit IgG, HRP-linked (Cell Signaling Technology, Cat#7074, 1:5000) and anti-mouse IgG, HRP-linked (Cell Signaling Technology, Cat#7076, 1:5000). Membrane fractionation was performed using Subcellular Protein Fractionation Kit for Cultured Cells (Thermo Fisher Scientific) following the manufacturer's protocol. Western blot images were taken on ChemiDoc Touch Imaging System (BIO-RAD) and analyzed in Image Lab software (version 6.0.1). All western blot raw data are shown in Supplementary Figure 1.

ChIP-seq.

ChIP was performed following the previous publication[53]. Briefly, iKAP cells were harvested via cross-linking with formaldehyde followed by quenching with glycine. Then cells were washed with ice-cold PBS and lysed on ice for 30 minutes in ChIP harvest buffer. Lysed cells were sonicated with a Bioruptor (Diagenode) to obtain chromatin fragments (~200–500bp) and soluble chromatin fractions were obtained by centrifugation. Ten percent of supernatant was set aside as input. Soluble chromatin was diluted five times using ChIP dilution buffer with Mini Protease Inhibitor Cocktail (Roche). Anti-H3K4me1 (Cell Signaling Technology, Cat#5326, 1:50), anti-H3K4me2 (Abcam, Cat#ab7766, 5 μg for 3 × 10^6 cells), anti-H3K4me3 (Abcam, Cat#ab8580, 5 μg for 3 × 10^6 cells), anti-H3K27ac (Abcam, Cat#ab4729, 5 μg for 3 × 10^6 cells), anti-mouse STAT4 (Invitrogen, Cat#714500, 1:100) antibodies and Dynabeads Protein G (Invitrogen) were incubated at 4 °C followed by washing with ChIP dilution buffer. Antibodies and beads were added to samples and incubated with rotation at 4 °C overnight. On the next day, magnetic protein G beads-antibody-protein complexes were separated by placing the sample tube on a magnetic rack. The samples were then washed in RIPA wash buffer, RIPA-500 wash buffer and LiCl wash buffer. Together with input, samples were reverse crosslinked with direct elution buffer, RNase A (Invitrogen) and Proteinase K (Invitrogen) overnight. On the next day, the supernatant was transferred to a new tube and 2.3X AMPure XP beads (Beckman Coulter) were added to purify the DNA. DNA was washed with 70% ethanol and eluted with 10 mM Tris-Cl buffer. The ChIP-seq library was prepared using KAPA HyperPrep Kit (Roche) following manufacturer's protocol and sequenced on Illumina NextSeq 500 High Output with 76nt single-read by the Advanced Technology Genomics Core at The University of Texas MD Anderson Cancer Center. NextSeq RTA (version 2.4.11.0) was used for base calling and bcl2fastq (version 2.20.0) was used to convert the .bcl files into .fastq files. Fastqc (version 0.11.8) was used to QC fastq data. Trimmomatic (version 0.33) was used to trim adapter sequence. Fastq files were aligned to the mouse reference genome (mm9) using Bowtie (version 1.2.2). SAMtools (version 1.9) was used to convert .sam file to .bam

file. Deeptools (version 2.7.15) was used to generate .bigwig files that were loaded to IGV software (version 2.9.2) for visualization. Peaks were called using MACS (version 1.4.2) and annotated in R (version 3.5.2) using R package 'ChIPseeker' (version 1.28.3). Heatmaps were generated in R using package 'EnrichedHeatmap' (version 1.22.0). Metaplots were generated in R using package 'Tidyverse' (version 1.3.1). MACS2 (version 2.1.2) was used to identify the differential binding of each protein in different conditions.

Luciferase reporter assay.

Mouse *Kdm5d* promoters with or without mutations were cloned into pGL3 luciferase vector by GenScript Biotech. Mutagenesis for mouse phospho-dead STAT4 was also performed by GenScript. The sequences bearing mSTAT4-Y693F, S721A, or S723A mutations were inserted into pCMV6-Entry vector. Luciferase vector and STAT4 ORF vector were co-transected into HEK293T cells with Lipofectamine 2000 (Thermo Fisher Scientific). After 24 hours, the medium was removed, and luciferase reporter assay was performed using Dual-Luciferase Reporter Assay Kit (Promega) following the manufacturer's protocol. The firefly luciferase signal was normalized to *Renilla* luciferase (internal control).

Transmission electron microscopy and tight junction integrity measurement.

iAP, iAP-KDM5D, and iKAP tumours were collected from mice approximately 100 days after 4-OHT and +/− dox water induction. Samples were fixed with a solution containing 3% glutaraldehyde plus 2% paraformaldehyde in 0.1 M cacodylate buffer, pH 7.3, then washed in 0.1 M sodium cacodylate buffer and treated with 0.1% Millipore-filtered cacodylate buffered tannic acid, postfixed with 1% buffered osmium, and stained en bloc with 1% Millipore-filtered uranyl acetate. The samples were dehydrated in increasing concentrations of ethanol, infiltrated, and embedded in LX-112 medium. The samples were polymerized in a 60 °C oven for approximately 3 days. Ultrathin sections were cut in an Ultracut microtome (Leica), stained with uranyl acetate and lead citrate in a Leica EM Stainer, and examined in a JEM 1010 transmission electron microscope (JEOL) at an accelerating voltage of 80 kV. Digital images were obtained using AMT Imaging System (Advanced Microscopy Techniques Corp). After randomizing the images and removing sample information, the number of tight junction kissing points was counted by three independent researchers.

Dextran paracellular permeability assay.

Cells were plated on Transwell with pore size of 0.4 μm (Corning) for two days to form an even monolayer. Fluorescein isothiocyanate–dextran (Sigma, Cat#FD2000S) was added to the upper chamber of the Transwell and incubated at 37 °C for 3 hours. After incubation, cells in the upper chamber were detached and the cell numbers were counted. The FITC signal in the lower chamber was measured with a CLARIOstar plate reader (BMG LABTECH) (Exc: 485 nm and Em: 544 nm) and the signal was normalized to cell numbers.

Immunofluorescence and confocal imaging.

Cells were seeded on top of the membrane of a Transwell with 0.4 μm pore size (Corning) for two days to form a monolayer. After washing with PBS, cells were fixed with 2%

paraformaldehyde (PFA)/PBS, for 15 minutes at RT. After PBS washing, cells were incubated with 0.5% Triton X-100/PBS for 10 minutes at RT and washed with PBS. Cells were then incubated with 10% calf serum/PBS for 1 hour at RT to block unspecific binding to antibody. Anti-ZO-1 antibody (Invitrogen, Cat#617300, 1:50) was diluted in dilution solution (10 mM Tris-HCl [pH 7.5], 0.05% Tween-20, 150 mM NaCl, 0.1% [w/v] BSA) and added to samples for overnight incubation at 4 °C. On the next day, cells were washed with PBS with 0.05% Tween-20. Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488 (Invitrogen, Cat#A11008, 1:500) was added to cells and incubated for 1 hour at RT. After PBS-Tween washing, the membrane was cut and placed on the glass slide and flanked by two tape strips on each side to avoid squeezing the cells on the membrane. The membrane was then covered with mounting medium and coverslip. Confocal imaging was performed using Leica SP8 Laser Scanning Confocal Microscope (SP8) at the Advanced Microscopy Core at The University of Texas MD Anderson Cancer Center.

Immunoprecipitation-MS.

LoVo cells were infected by lentivirus containing Flag-hKDM5D construct. Cells were harvested and washed with ice-cold PBS then lysed in RIPA buffer (50 mM HEPES [pH 7.4], 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1X protease and phosphatase inhibitor cocktail [Thermo Fisher Scientific]). Cell debris was removed with centrifugation, and 5% of the supernatant was saved as input. Cell lysates were pre-cleared with Protein A/G Plus Agarose (Thermo Fisher Scientific) for 30 minutes at 4 °C. Samples were incubated with anti-Flag antibody (Sigma, Cat#F1804, 1:200) or anti-mouse IgG (Sigma, Cat#06–371, 1:200) overnight at 4 °C, and Protein A/G Plus Agarose was then added and incubated for an additional 2–3 hours at 4 °C. The immunoprecipitates were washed with RIPA buffer and PBS, recovered with s.d.S sample buffer and separated on s.d.S-PAGE. Gel bands were digested overnight with trypsin (Pierce) following destaining, reduction with DTT and alkylation with iodoacetamide (Sigma). The samples then underwent solid-phase extraction cleanup with an Oasis HLB plate (Waters) and the resulting samples were injected onto an Orbitrap Fusion Lumos mass spectrometer coupled to an Ultimate 3000 RSLC-Nano liquid chromatography system. Samples were injected onto a 75 μm i.d., 75-cm long EasySpray column (Thermo Fisher Scientific) and eluted with a gradient from 1–28% buffer B over 90 minutes. Buffer A contained 2% (v/v) ACN and 0.1% formic acid in water, and buffer B contained 80% (v/v) ACN, 10% (v/v) trifluoroethanol, and 0.1% formic acid in water. The mass spectrometer was operated in positive ion mode with a source voltage of 1.8 kV and an ion transfer tube temperature of 275 °C. MS scans were acquired at 120,000 resolution in the Orbitrap and up to 10 MS/MS spectra were obtained in the ion trap for each full spectrum acquired using higher-energy collisional dissociation (HCD) for ions with charges 2–7. Dynamic exclusion was set for 25 s after an ion was selected for fragmentation.

Raw MS data files were analyzed using Proteome Discoverer v2.2 (Thermo Fisher Scientific) with peptide identification performed using Sequest HT searching against the human protein database from UniProt. Fragment and precursor tolerances of 10 ppm and 0.6 Da were specified, and three missed cleavages were allowed. Carbamidomethylation of Cys

was set as a fixed modification, with oxidation of Met set as a variable modification. The false-discovery rate (FDR) cutoff was 1% for all peptides.

Co-IP.

HEK293T cells were transfected with plasmids of Flag-hKDM5D, hSAP18-GFP (OriGene, Cat#RC205607L4), or Flag (Addgene, Cat#20011) or GFP (OriGene, Cat#PS100072) control vectors. Flag-IP was performed as described above, followed by western blotting using antibodies for anti-Flag (Sigma, Cat#F1804, 1:1000), anti-GFP (Cell Signaling Technology, Cat#2956, 1:1000), anti-HDAC1 (Cell Signaling Technology, Cat#34589, 1:1000) and anti-HDAC2 (Abcam, Cat#ab16032, 1:1000).

Flow cytometry and antigen presentation assay.

The recommended protocol, BioLegend Cell Surface Flow Cytometry Staining Protocol, was followed. Briefly, the cells were washed in Cell Staining Buffer (BioLegend), centrifuged at 350 x g for 5 minutes at 4 °C, incubated for 10 minutes on ice in 1 μg of TruStain FcX anti-mouse CD16/32 antibody (BioLegend, Cat#101319, 1:50) diluted in Cell Staining Buffer, and stained for 20 minutes on ice in the dark with APC anti-mouse H-2Kb/H-2Db (BioLegend, Cat#114613, 1:40). Cells were analyzed by flow cytometer LSRFortessa X-20 Analyzer (Becton Dickinson) and FlowJo (version 10.8.0).

For antigen presentation assay, L2–3 and KO iKAP cells were treated with 10 ng/mL recombinant mouse Interferon-gamma (IFNγ) protein (Abcam, Cat#ab9922) and 2 mg of ovalbumin protein (InvivoGen, Cat#vac-pova) for 48 hours. For flow cytometry, the above cells were stained with Anti-H2Kb/SIINFEKL-APC antibody (Miltenyi Biotec, Cat#130–102-175, 1:10). Flow cytometry gating strategies are shown in Supplementary Figure 2.

OT-1 $CD8^+$ T cell stimulation, purification and T cell killing assay.

An 8-week-old male OT-1 mouse was euthanized following IACUC protocol and the spleen was removed. The spleen was disrupted in PBS containing 2% FBS. The aggregates and debris were removed by passing cell suspension through a 70-μm mesh nylon strainer. The cells were centrifuged at 300 x g for 10 minutes and resuspended in T cell culture medium (RPMI-1640 + 2.05 mM L-Glutamine [Cytiva], GlutaMAX [Thermo Fisher Scientific], 1x penicillin/streptomycin [Invitrogen]) with 5 μM ovalbumin (257–264) peptide (InvivoGen, Cat#vac-sin) and 30 U/mL mouse IL-2 (Abcam, Cat#ab9856) for 6 days for OT-1 $CD8^+$ T cell stimulation. After stimulation, the $CD8^+$ T cells were isolated with Mouse $CD8^+$ T Cell Isolation Kit (Stemcell) following the manufacturer's recommended protocol. The isolated $CD8^+$ T cells were cultured with T cell culture medium as mentioned above supplemented with 30 U/mL mouse IL-2 and β-mercaptoethanol (1000X).

The cancer cells were treated under the same conditions as in the antigen presentation assay. After 24 hours of treatment, ovalbumin and IFNγ were washed off, the cells were detached and counted, and 10 times the activated $CD8^+$ T cells were added. On the next day, all cells were collected and stained with SYTOX Blue Dead Cell Stain (Invitrogen). The single iKAP cell was separated with GFP+ signal and the percentage of dead cells (SYTOX Blue+) was calculated in FlowJo (version 10.8.0).

HiChIP.

HiChIP was performed following published protocol[33]. Briefly, the cells were fixed in 1% formaldehyde for 10 minutes at RT and quenched by glycine for 10 minutes. After washing with PBS, cells were resuspended in Hi-C lysis buffer and rotated at 4 °C for 30 minutes. Nuclei were pelleted, washed with Hi-C Lysis Buffer and resuspended in 0.5% s.d.S, which was diluted and quenched by Triton X-100. NEB buffer 2 and Mbol restriction enzyme (New England Biolabs, Cat#R0147) were used to fragment the genome and the DNA fragments were labeled with biotin-dATP (Thermo Fisher Scientific). Proximal DNA fragments were then ligated using NEB T4 DNA ligase (New England Biolabs, Cat#M0202). The nuclear pellet was lysed in nuclear lysis buffer and chromatins were then sonicated using a Covaris E220 and clarified by centrifugation. Clarified samples were diluted in ChIP Dilution Buffer. Samples were precleared with Dynabeads Protein G (Invitrogen). H3K27ac antibody (Abcam, Cat#ab4729) was added at a concentration of 3.75 μg per 5×10^6 cells and incubated overnight at 4 °C. Dynabeads Protein G beads were used to capture the antibody-protein-DNA complex. The captured samples were washed with low-salt wash buffer, high-salt wash buffer and LiCl wash buffer. The DNA was reverse crosslinked in DNA elution buffer and incubated with Proteinase K (Invitrogen). The DNA was purified using DNA Clean and Concentrator columns (Zymo Research) and eluted in water. After quantification using Qubit ds.d.NA HS Assay Kit (Thermo Fisher Scientific), Streptavidin C-1 beads (Thermo Fisher Scientific) were added to samples to capture the biotin-labeled DNA. After washing, DNA was mixed with Tn5 transposase (Illumina) for tagmentation for library preparation. DNA was amplified by Phusion HF (New England Biolabs), Nextera Ad1_noMx and Nextera Ad2.X (see ref.[33] for primer sequences). After removing the primers with Ampure XP beads (Beckman Coulter) cleanup, the Nextera_i7Short and Nextera_i5Short primers (see ref.[33] for primer sequences) were added to samples for additional PCR reactions. The DNA was size-selected using Ampure XP beads to capture fragments between 300bp and 700bp. Libraries passed the QC and were sequenced on NovaSeq6000 SP-xp flow cell. HiChIP data were aligned to mm9 genomes using the HiC-Pro pipeline[54] (version 2.11.1) with the default settings. Interaction loops were called with hichipper (version 0.7.7) (https://github.com/aryeelab/hichipper) and converted into .hic format files and loaded into juicer tools (version 1.13.02) (https://github.com/aidenlab/juicer/wiki/Download) for visualization.

Super-enhancer prediction.

Super-enhancers in iKAP cells were called using ROSE[31,32] (version 0.1) main program ‘ROSE_main.py’ on Python (version 2.7.3), with reference genome mm9. The data files used to run the program include .bed and .bam files from H3K27ac ChIP-seq and .bam file from the input in the ChIP-seq.

GSEA.

GSEA was done using GSEA (version 4.1.0). For GSEA analysis of *KDM5D*-high versus low patients, TCGA male CRC was divided into *KDM5D*-low and *KDM5D*-high based on *KDM5D* expression in RNA-seq with the optimal cutoff value of gene expression determined with the website-based tool, Cutoff Finder[55] (https://

molpathoheidelberg.shinyapps.io/CutoffFinder_v1/). Whole-genome expression data were loaded to GSEA (version 4.1.0) and gene set data 'c7.all.v7.4.symbols.gmt [Immunologic signatures]' was used. Results were filtered with FDR<0.25 as the cutoff.

Gene set variation analysis.

KRAS signalling score was calculated by R (version 3.5.2) and R package 'GSVA' (version 1.40.1) using TCGA male KRAS* CRC RNA-seq data and KRAS signalling activation signature derived from differentially expressed genes in RNA-seq of iKAP tumours with dox on versus off (SRP097890).

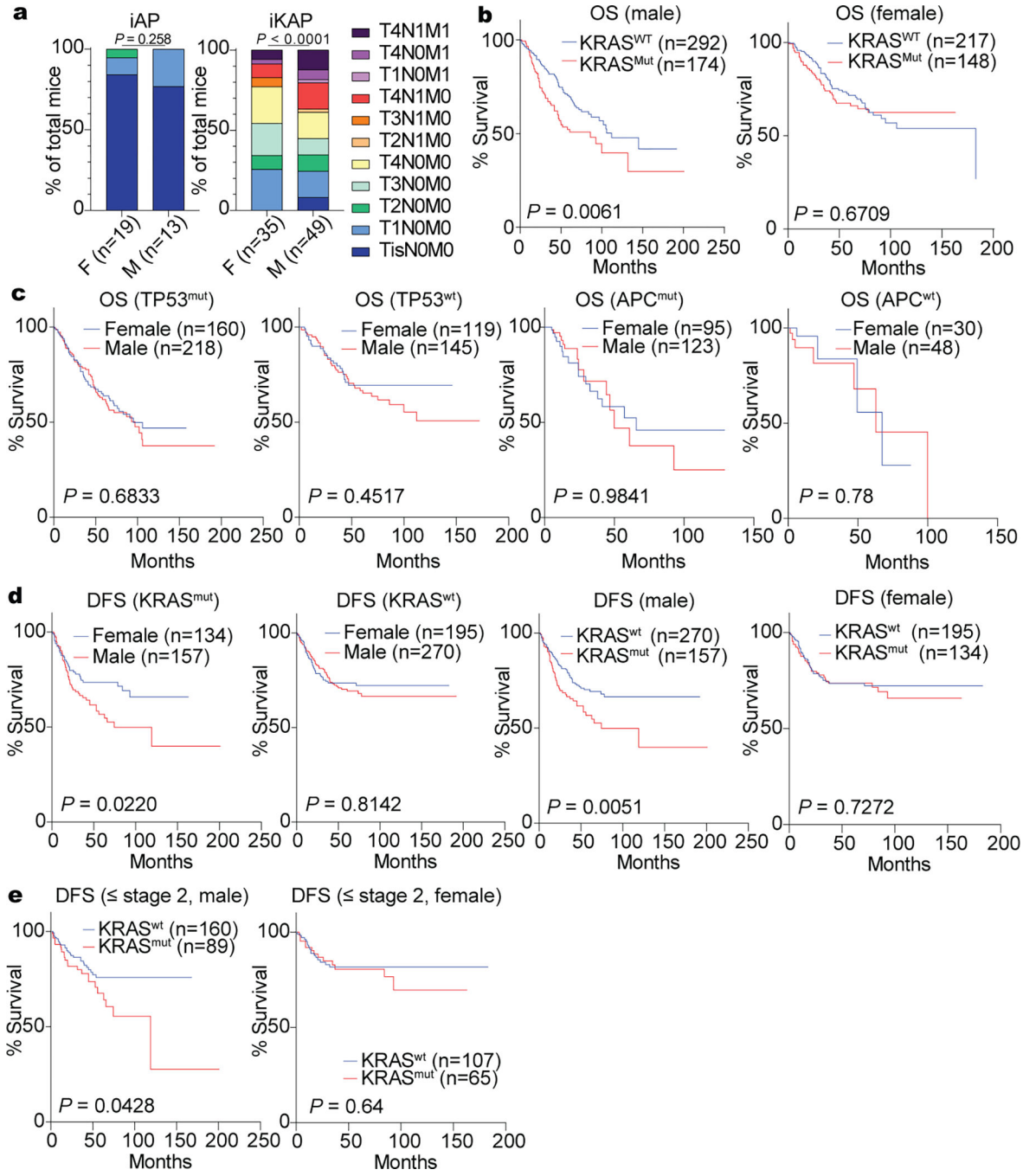
Transcription factor binding site prediction.

The transcription factor bound to mouse *Kdm5d* and human *KDM5D* promoter was predicted using JASPAR 2022 (http://jaspar.genereg.net/). Mouse and human DNA sequences of 1kb upstream and downstream of *Kdm5d*/*KDM5D* TSS were used as input for JASPAR 2022 with relative profile score threshold of 85%.

Statistics and reproducibility.

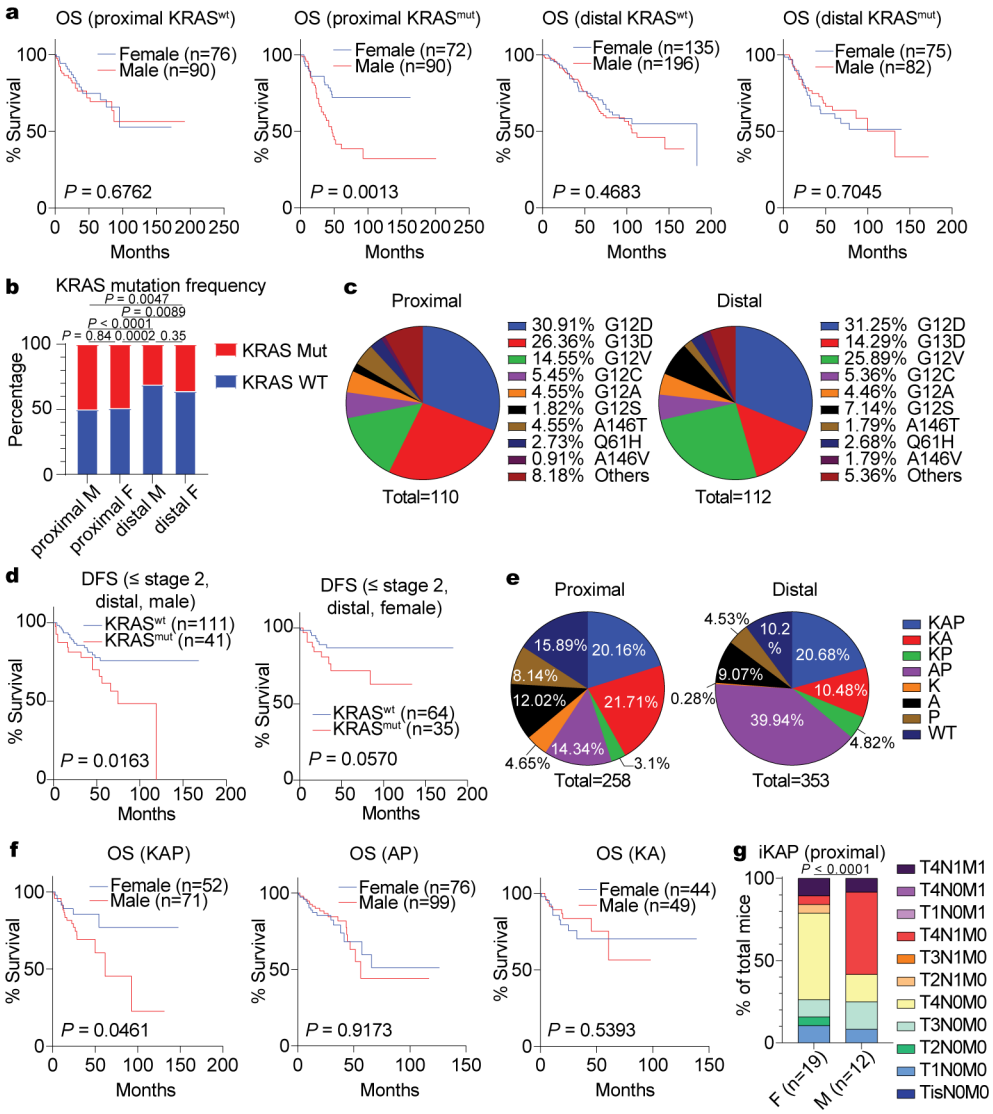
All statistical tests were performed in GraphPad Prism 8. All experiments were repeated independently at least three times with similar results unless specified in the figure legends. Statistical analyses are noted in figure legends.

Extended Data



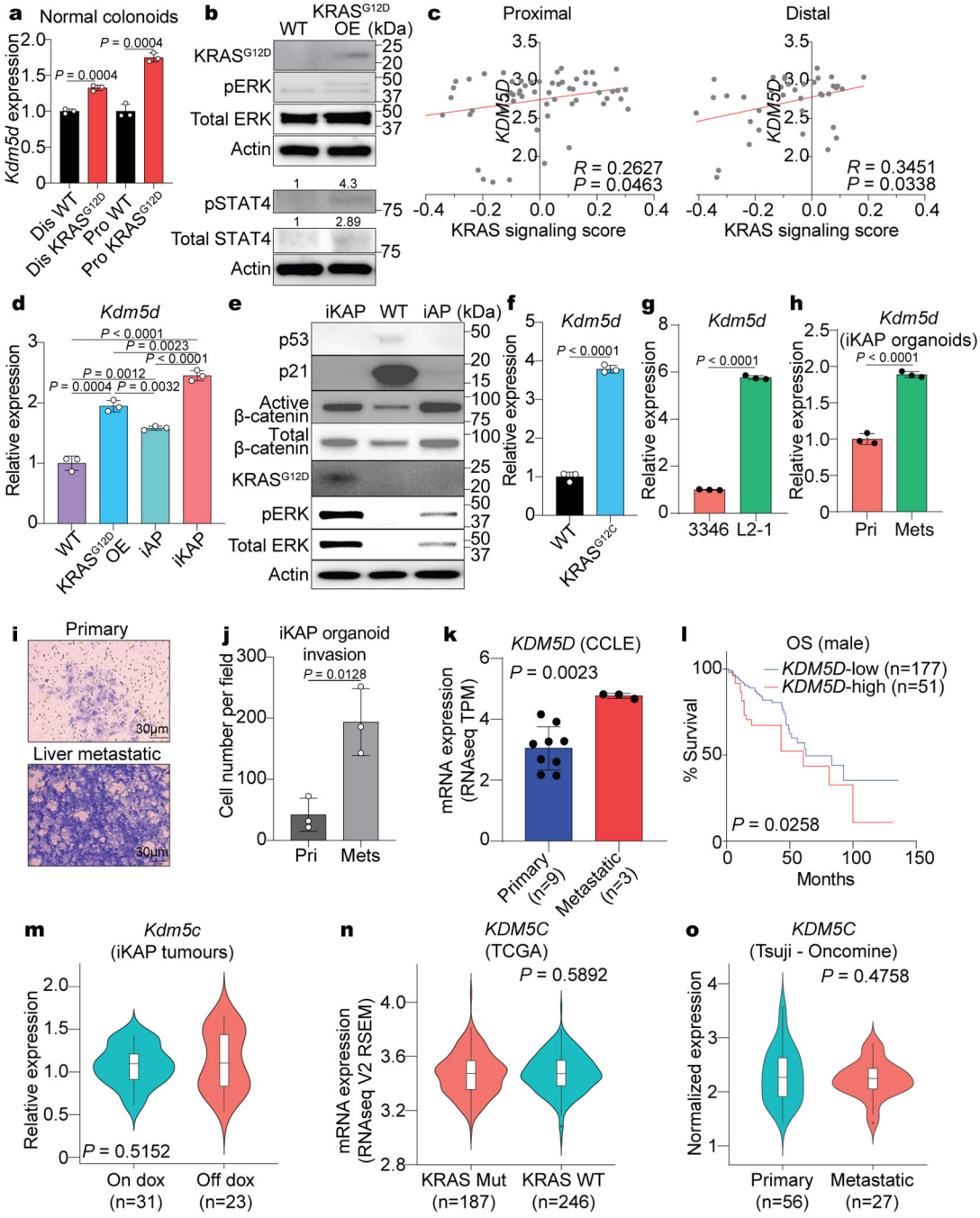
Extended Data Figure 1. CRC sex differences in survival and metastasis are KRAS*-specific.

a, Stages of female and male iAP and iKAP tumours (two-tailed Chi-square test). **b**, **c**, Kaplan-Meier OS analysis for patients in CRCSC data set. **d**, **e**, Kaplan-Meier disease-free survival (DFS) analysis for patients in CRCSC data set. For survival analysis (**b**-**e**), log-rank (Mantel-Cox) test was used.



Extended Data Figure 2. Comparisons of proximal and distal CRC in humans and mice.

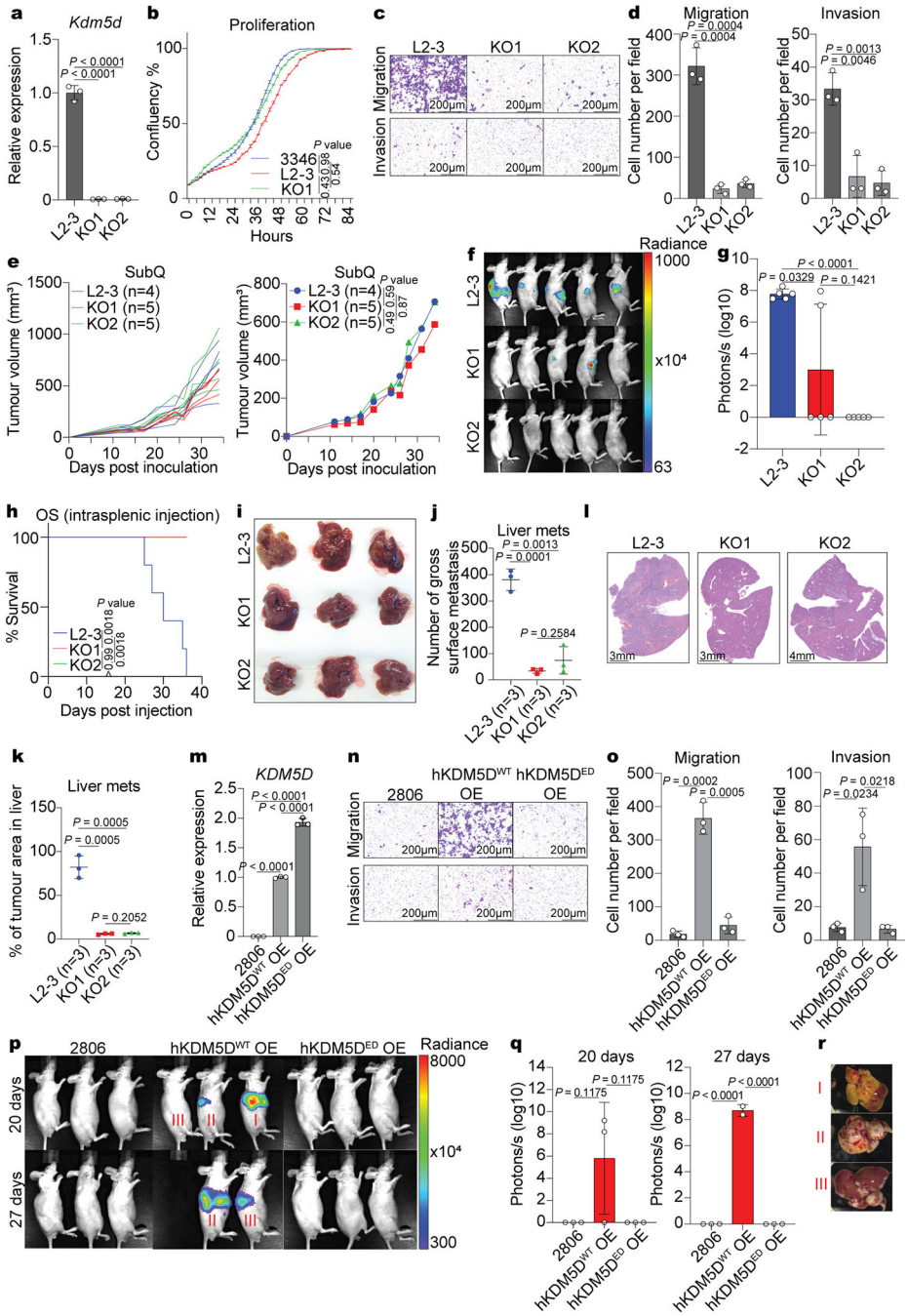
a, Kaplan-Meier OS analysis for patients in CRCSC data set. **b**, Frequency of *KRAS* mutation in patients in CRCSC data set (two-tailed Binomial test). **c**, Subtypes of *KRAS* mutations in patients in CRCSC data set. **d**, Kaplan-Meier DFS analysis for patients with stage 1 or stage 2 cancers in CRCSC data set. **e**, Percentage of mutation combinations of *KRAS* ('K'), *APC* ('A') and *TP53* ('P') in TCGA CRC patients. **f**, Kaplan-Meier OS analysis for patients with indicated mutations in TCGA CRC data set. **g**, Stages of iKAP proximal tumours induced by injection of 4-OHT at proximal colon (two-tailed Chi-square test). For survival analysis (**a**, **d**, **f**), log-rank (Mantel-Cox) test was used.



Extended Data Figure 3. KDM5D is upregulated in KRAS* driven metastasis.

a, qPCR of *Kdm5d* in male normal colonoids isolated from proximal ('pro') or distal ('dis') colon of healthy mice. Relative expression was calculated by normalizing to the 'WT' condition of each pair (n=3 wells per group). **b**, Western blot confirming successful overexpression of $KRAS^{G12D}$ and activation of KRAS signalling and STAT4 signalling in colonoids used in **a**. Phospho-STAT4 and total STAT4 levels were quantified by normalizing to Actin and 'WT' condition. The samples were derived from the same experiment and gel/blots were processed in parallel. **c**, Two-tailed Pearson correlation analysis of KRAS

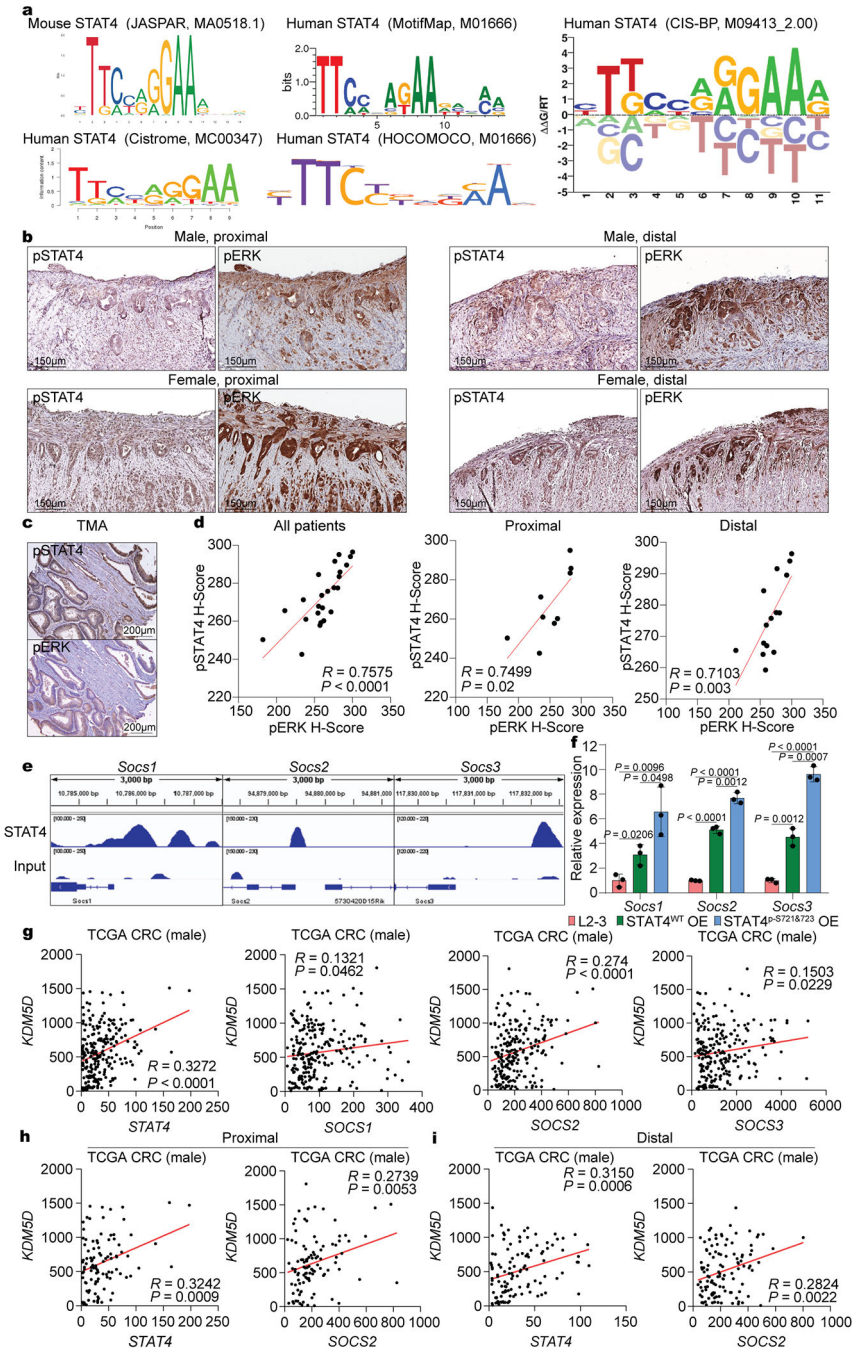
signalling score and *KDM5D* expression in TCGA male CRC patients. KRAS signalling score was calculated with GSVA using expression of all genes in TCGA RNA-seq data and KRAS signalling activation signature which was derived from differentially expressed genes in RNA-seq of iKAP tumours with dox on versus off (SRP097890) (*n*=58 proximal CRC patients; *n*=38 distal CRC patients). **d**, qPCR of *Kdm5d* in male normal colonoids with indicated genotypes (*n*=3 wells per group). **e**, Western blot of colonoids with indicated genotypes in **d**. The samples were derived from the same experiment and gel/blots were processed in parallel. **f**, qPCR of *Kdm5d* in male normal colonoids. Relative expression was calculated by normalizing to 'WT' (*n*=3 wells per group). **g**, qPCR of *Kdm5d* in 3346 and matched liver metastatic cell line L2–1 (independent of L2–3). Relative expression was calculated by normalizing to 3346 (*n*=3 wells per group). **h**, qPCR of *Kdm5d* in organoids derived from male iKAP primary ('Pri') and matched spontaneous liver metastatic ('Mets') tumours. Relative expression was calculated by normalizing to 'Pri' (*n*=3 wells per group). **i**, Representative images of in vitro invasion assays of organoids used in **h** to validate the increased mobility of metastatic cells compared to primary cells. **j**, Quantitation of **i** (*n*=3 wells per group). **k**, Expression of *KDM5D* in human male CRC cell lines with *KRAS* mutations in CCLE database. **l**, Kaplan-Meier OS analysis for TCGA male CRC patients with high or low *KDM5D* expression (log-rank (Mantel-Cox) test). The optimal cutoff value of *KDM5D* expression was determined with Cutoff Finder application. **m**, *Kdm5c* expression (*z* score) iKAP tumour RNA-seq data (SRP097890). **n**, *KDM5C* expression in TCGA CRC patient RNA-seq data. **o**, *KDM5C* expression in CRC patients in Tsuji Colorectal Cancer RNA microarray data set downloaded from Oncomine database. For box plots in **m**-**o**, center lines denote medians; box limits denote 25th–75th percentile (Q1-Q3); whiskers are drawn up to the smallest or largest observed value that was still within 1.5 times the interquartile range below the first quartile or above the third quartile, respectively; all other observed points are plotted as outliers). All bar plots show mean value ± s.d.; except for correlation analysis (**c**) and survival analysis (**l**), *P* was derived with two-tailed unpaired *t* test; no adjustment for multiple comparisons.



Extended Data Figure 4. *Kdm5d* deletion suppresses tumour cell metastasis.

a, qPCR of *Kdm5d* in L2–3 and two independent *Kdm5d* knockout cell lines ('KO1' and 'KO2') derived from L2–3. Relative expression was calculated by normalizing to 'L2–3' (n=3 wells per group). **b**, Proliferation assay of 3346, L2–3 and KO1 (n=3 wells per group; error bars indicate s.d.; two-tailed Mann-Whitney test; no adjustment for multiple comparisons). **c**, Representative images of migration and invasion assay. **d**, Quantitation of **c** (n=3 wells per group). **e**, Growth curves of subcutaneous tumours in male nude mice (two-tailed Mann-Whitney test; no adjustment for multiple comparisons). **f**, Liver metastases

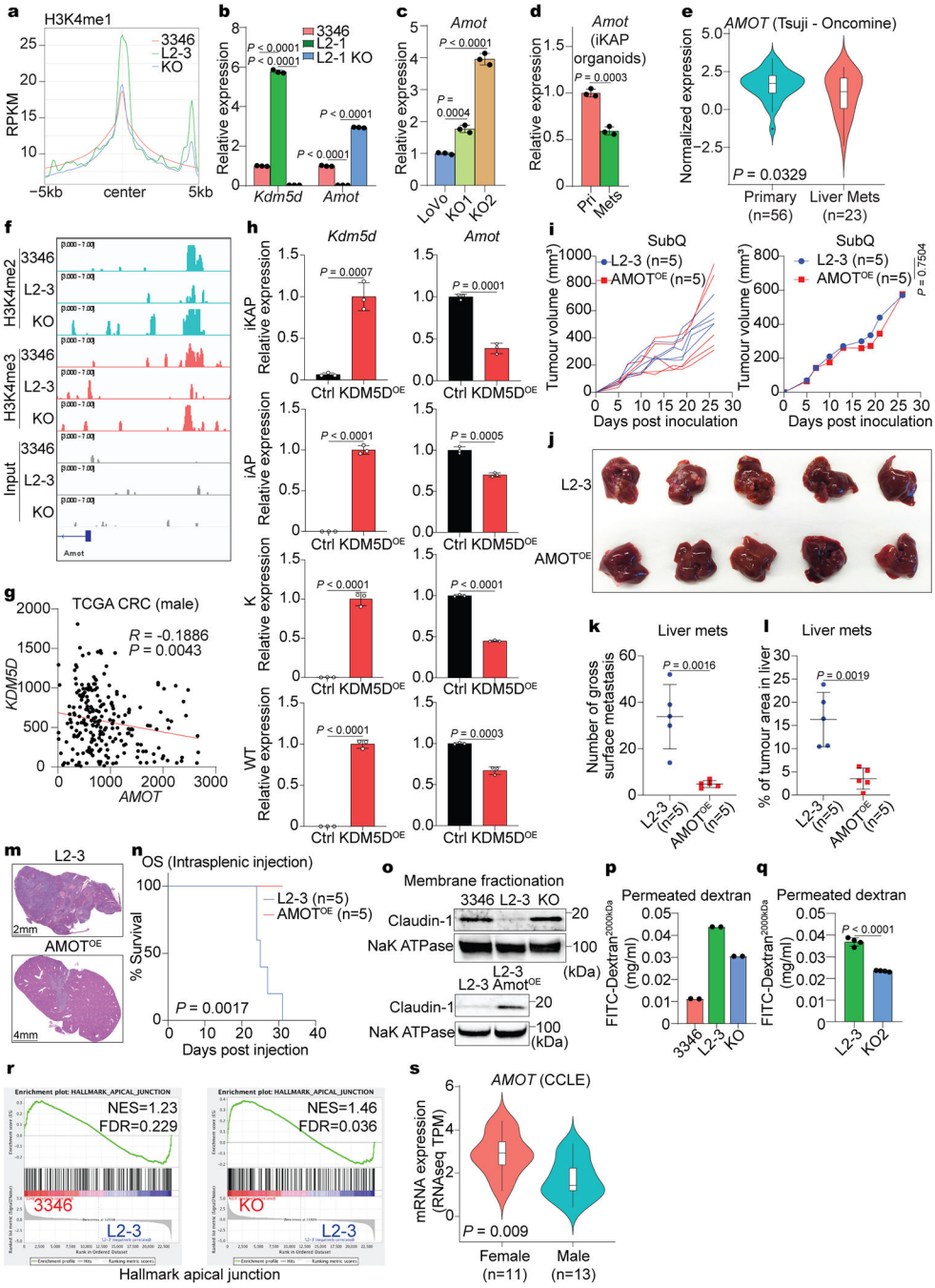
detected by luminescence in male nude mice with intrasplenically injected male iKAP cells with a luciferase reporter (n=5 mice per group). **g**, Quantitation of luminescence signals in **f** (n=5 mice per group). **h**, Kaplan-Meier OS analysis of mice in **f** (n=5 mice per group; Log-rank [Mantel-Cox] test). **i**, Representative liver tissues with metastatic tumours from an independent experimental repeat of **f-h** (n=3 livers per group). **j**, Number of gross surface metastasis in livers in **i**. Cells injected into spleen metastasize to liver through portal vein. Surrounding the blood vessel, the cells form initial liver metastasis, followed by whole liver spread. The formation of liver surface metastasis indicates higher metastatic potential of the cells and more aggressive disease. **k**, Percentage of histologically metastatic lesion area in all liver lobes in each mouse. **l**, Representative histology of liver metastasis (representing three tissues per group). **m**, qPCR of *KDM5D* in 2806, a cell line derived from female iKAP distal primary CRC, and 2806 with overexpression of WT or enzymatic-dead ('ED') human *KDM5D*. Relative expression was calculated by normalizing to 'hKDM5D^{WT} OE' (n=3 wells per group). **n**, Representative images of migration and invasion assays of cell lines in **m**. **o**, Quantitation of **n** (n=3 wells per group). **p**, Liver metastases detected by luminescence in female nude mice with intrasplenically injected female iKAP cells with a luciferase reporter (n=3 mice per group). **q**, Quantitation of luminescence signals in **p** (n=3 mice per group, except for n=2 mice for hKDM5D^{WT} OE group at 27 days). **r**, Liver metastatic tumours from mice in **p**. For all bar/dot plots (**a**, **d**, **g**, **j**, **k**, **m**, **o**, **q**), data are shown as mean value ± s.d.; P was derived with two-tailed unpaired t test; no adjustment for multiple comparisons.



Extended Data Figure 5. STAT4 activation correlates with KRAS signalling activation and *KDM5D* expression.

a, STAT4 motifs in mouse and human in indicated databases. **b**, Representative images of IHC of iKAP primary tumours (representing three individual tumours per group). **c**, Representative images of IHC of tissue microarrays (TMA) of male CRC patients' primary tumours (representing results from 24 patients, as detailed in **d**). **d**, Pearson correlation analysis of H-scores of phospho-STAT4 and phospho-ERK in nucleus of epithelial cells in TMA of male CRC patients' primary tumours (n=24 patients, 9 with proximal CRC,

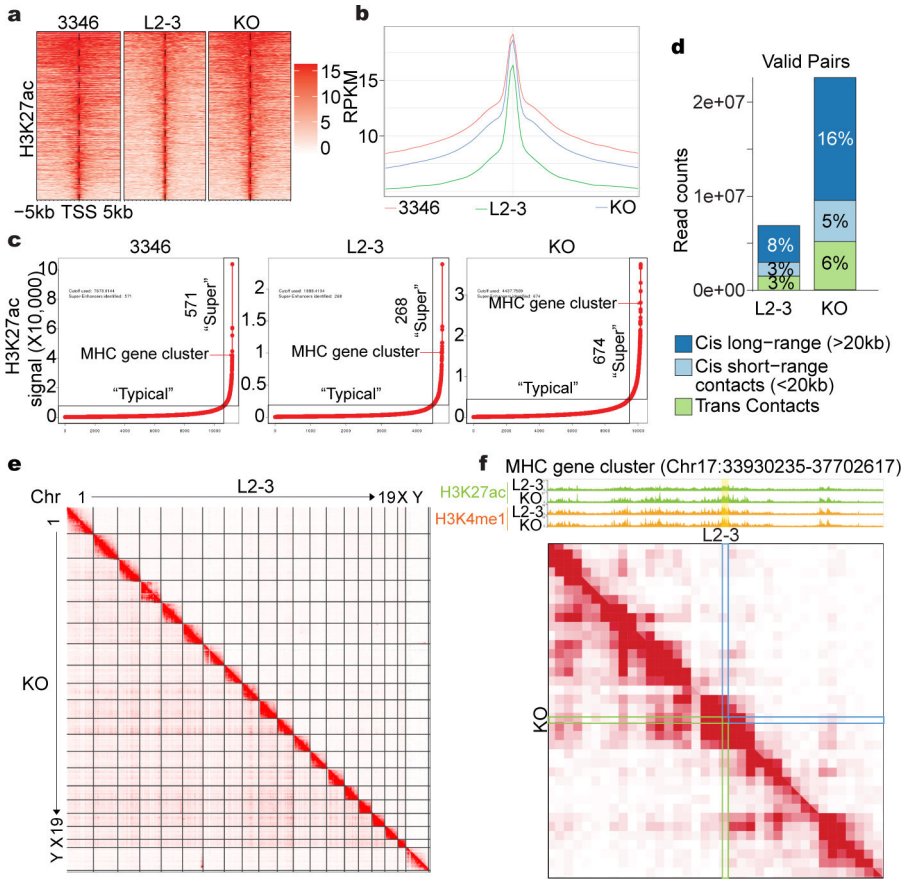
15 with distal CRC). **e**, ChIP-seq tracks of STAT4 IP and input around promoters of *Socs1*, *Socs2*, and *Socs3* in iKAP cells cultured with dox. **f**, qPCR of L2–3 with WT or phosphomimetic STAT4 overexpression. Relative expression was calculated by normalizing to 'L2–3' (n=3 wells per group; two-tailed unpaired t test; mean value ± s.d.; no adjustment for multiple comparisons). **g**, Pearson correlation analysis of *STAT4*, *SOCS1*, *SOCS2*, *SOCS3* and *KDM5D* expression in TCGA male CRC patients. X- and Y-axis units are mRNA expression (RNA-seq V2 RSEM) (n=229, 228, 230, 229 patients). **h**, **i**, Pearson correlation analysis of *STAT4*, *SOCS2* and *KDM5D* expression in TCGA male CRC patients with proximal or distal primary tumours. X- and Y-axis units are mRNA expression (RNA-seq V2 RSEM) (for **h**, n=101, 102 patients; for **i**, n=115, 115 patients).



Extended Data Figure 6. AMOT is downregulated in metastasis and negatively regulated by KDM5D.

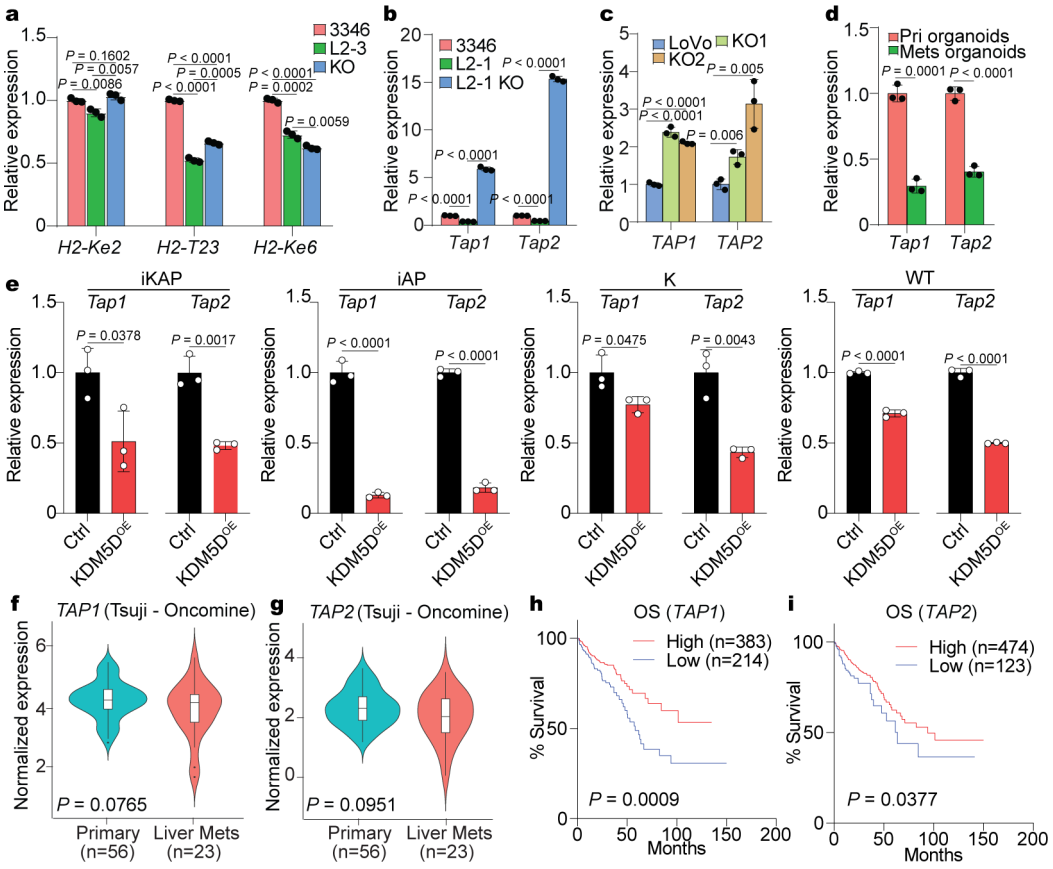
a, Metaplot of H3K4me1 in iKAP cells. Average signal in the window of 5 kb up-/down-stream of TSSs was used for two-tailed Mann-Whitney test. 3346 versus L2–3, $P = 0.3899$, L2–3 versus KO, $P = 0.0269$, 3346 versus KO, $P = 0.0013$. **b**, qPCR of 3346, L2–1 and L2–1 with *Kdm5d* knockout (L2–1 KO). Relative expression was calculated by normalizing to '3346' (n=3 wells per group). **c**, qPCR of human male metastatic CRC cells LoVo and two independent *KDM5D* knockout clones ('KO1' and 'KO2'). Relative

expression was calculated by normalizing to 'LoVo' (n=3 wells per group). LoVo cell line has *KRAS*^{G13D}, *APC* loss of function mutations (nonsense mutation and frameshift deletion) and missense mutation, and wild-type *TP53*. **d**, qPCR of organoids from iKAP primary ('Pri') and matched spontaneous metastatic ('Mets') tumours, which were validated in Extended Data Fig. 3h–j. Relative expression was calculated by normalizing to 'Pri' (n=3 wells per group). **e**, RNA microarray data of CRC primary and metastatic tumours from patients in Tsuji Colorectal Cancer data set in Oncomine (box plots, center lines denote medians; box limits denote 25th–75th percentile [Q1-Q3]; whiskers are drawn up to the smallest or largest observed value that was still within 1.5 times the interquartile range below the first quartile or above the third quartile, respectively; all other observed points are plotted as outliers). **f**, ChIP-seq tracks of H3K4me2/3 IP and input around *Amot* promoter in iKAP cells. **g**, Pearson correlation analysis of *AMOT* and *KDM5D* expression in TCGA CRC male patients. X- and Y-axis units are mRNA expression (RNA-seq V2 RSEM) (n=228 patients). **h**, qPCR of female colonoids with indicated genotypes ('WT', wild type; 'K', KRAS^{G12D} overexpression in WT colonoids) and corresponding colonoids with mouse *Kdm5d* overexpression. *Kdm5d* relative expression was calculated by normalizing to 'KDM5D^{OE}'; *Amot* relative expression was calculated by normalizing to 'Ctrl' (n=3 wells per group). **i**, Growth curves of subcutaneous tumours from L2–3 cell line and L2–3 with *Amot* overexpression in male nude mice (two-tailed Mann-Whitney test). **j**, Representative liver metastatic tumours in male nude mice with intrasplenically injected male iKAP cells. **k**, Number of gross surface metastasis in livers in **j**. **l**, Percentage of histologically metastatic lesion area in all liver lobes in each mouse in **j**. **m**, Representative histology of liver metastasis (representing five livers per group). **n**, Kaplan-Meier OS analysis of mice from an independent experimental repeat of **j**-**m** (log-rank (Mantel-Cox) test). **o**, Western blot of fractionated iKAP cell cytoplasmic membrane. NaK ATPase was used as the membrane-specific loading control. **p**, **q**, Paracellular permeability assay of iKAP cells. FITC-conjugated dextran with molecular weight of 2000 kDa was used (for **p**, n=2 wells per group; for **q**, n=4 wells per group). **r**, GSEA of iKAP cell RNA-seq. **s**, Expression of *AMOT* in human CRC cell lines with *KRAS* mutations in CCLE (box plots, center lines denote medians; box limits denote 25th–75th percentile [Q1-Q3]; whiskers are drawn up to the smallest or largest observed value that was still within 1.5 times the interquartile range below the first quartile or above the third quartile, respectively; all other observed points are plotted as outliers). All bar/dot plots (**b**-**d**, **h**, **k**, **l**, **p**, **q**) show mean value ± s.d.. For **b**-**e**, **h**, **k**, **l**, **p**, **q**, **s**, P was derived with two-tailed unpaired t test; no adjustment for multiple comparisons.



Extended Data Figure 7. KDM5D regulates H3K27ac and super-enhancer activity.

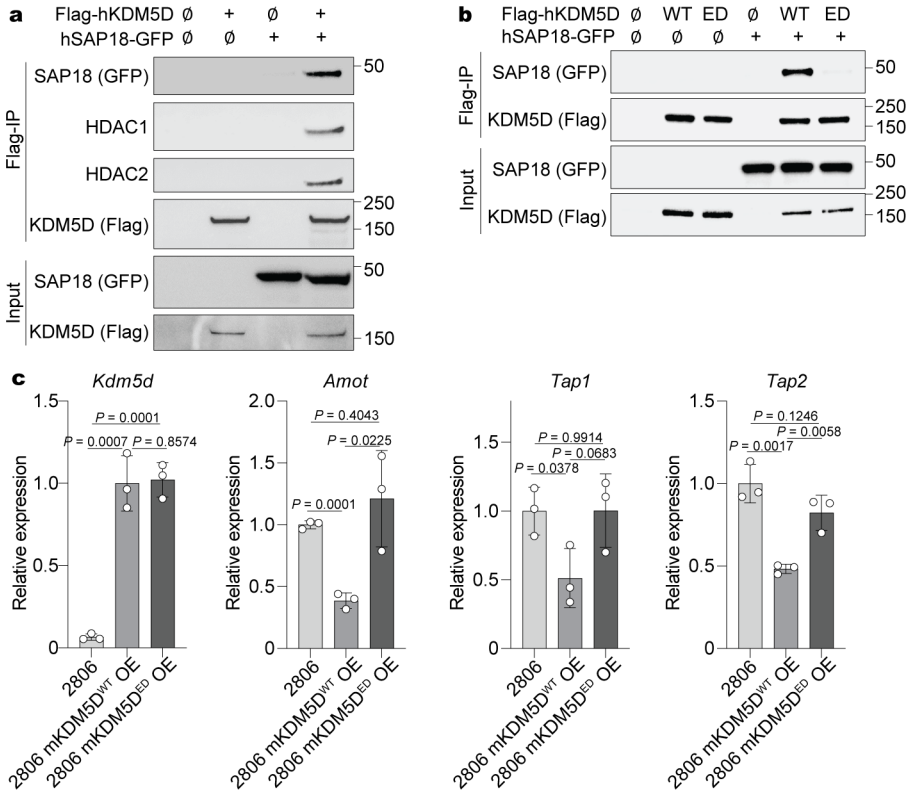
a, **b**, Heatmaps (**a**) and metaplots (**b**) of H3K27ac in iKAP cells. Average signal in the window of 5 kb up-/down-stream TSSs was used to focus on the gene enhancers, and two-tailed Mann-Whitney test was performed. 3346 versus L2–3, $P < 0.0001$, L2–3 versus KO, $P < 0.0001$, 3346 versus KO, $P < 0.0001$. **c**, Putative super-enhancers ranked by H3K27ac signal in iKAP cells. 'Super' indicates super-enhancer, 'Typical' indicates typical enhancer. The super-enhancer predicted to regulate the MHC gene cluster was pointed out. **d**, The percentage of long and short-range interactions in iKAP cells detected by HiChIP analysis. **e**, Interaction map between and within chromosomes in iKAP cells detected by HiChIP. **f**, Top: ChIP-seq tracks of enhancer markers H3K27ac and H3K4me1 in the MHC gene cluster, super-enhancer region was highlighted in yellow; bottom: the interaction map generated by HiChIP analysis in iKAP cells. For **e** and **f**, red to white color range indicates more to less interaction.



Extended Data Figure 8. TAP1/2 are regulated by KDM5D and are correlated with CRC metastasis and shorter survival.

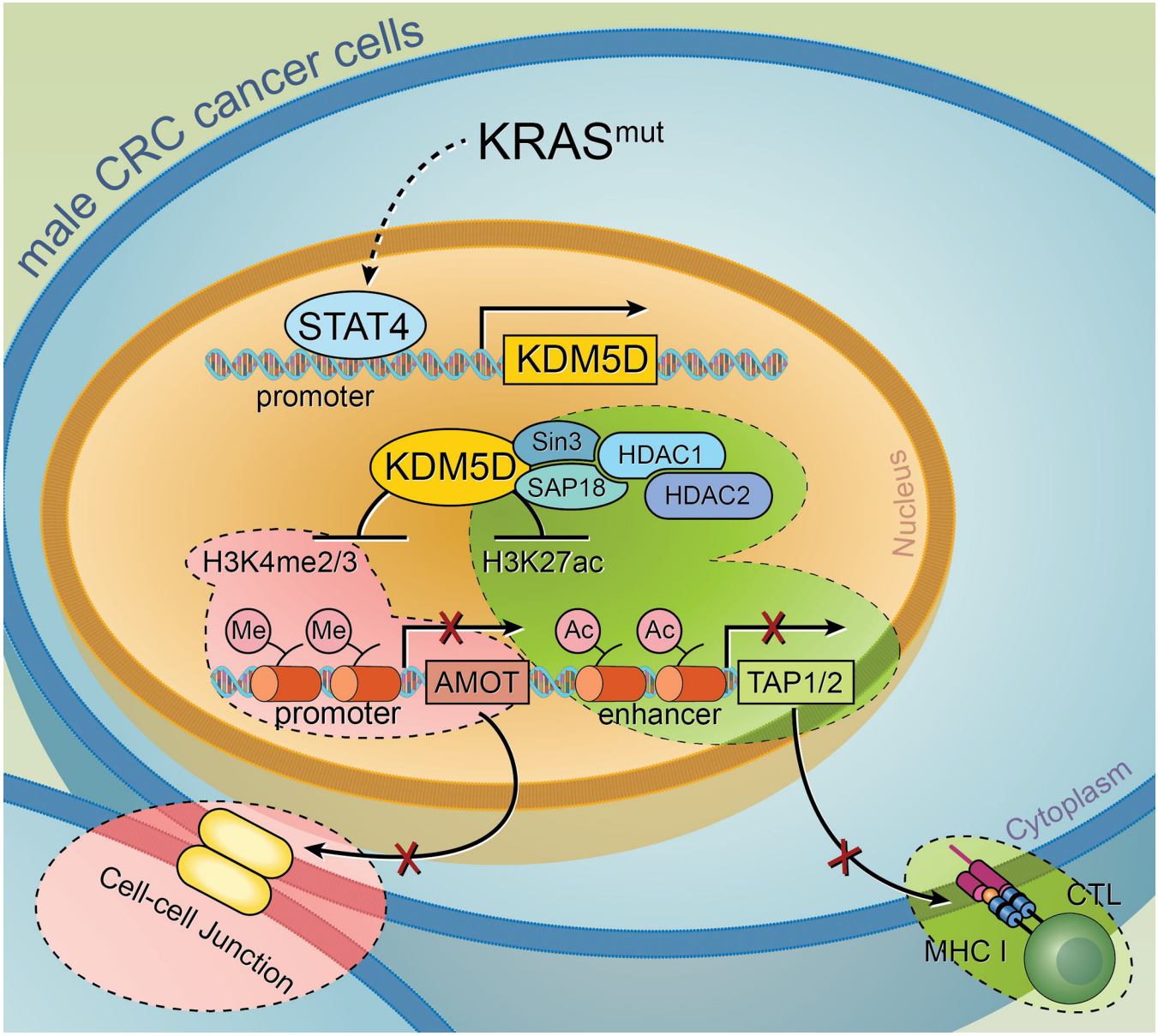
a, qPCR of iKAP cells. Relative expression was calculated by normalizing to '3346' (n=3 wells per group). Fold changes between L2–3 and KO of all three genes were < 1.5, indicating no expression correlation between genes tested and *Kdm5d*. **b**, qPCR of iKAP cells. Relative expression was calculated by normalizing to '3346' (n=3 wells per group). **c**, qPCR of human male metastatic CRC cells LoVo and two independent *KDM5D* knockout clones ('KO1' and 'KO2'). Relative expression was calculated by normalizing to 'LoVo' (n=3 wells per group). **d**, qPCR of organoids from male iKAP primary ('Pri') and metastatic ('Mets') tumours. Relative expression was calculated by normalizing to 'Pri organoids' (n=3 wells per group). **e**, qPCR of female colonoids of indicated genotypes and corresponding colonoids with mouse *Kdm5d* overexpression. Relative expression was calculated by normalizing to 'Ctrl' (n=3 wells per group). **f**, **g**, RNA microarray data of *TAP1* (**f**) and *TAP2* (**g**) in CRC primary and metastatic tumours from patients in Tsuji Colorectal Cancer data set in Oncomine (box plots, center lines denote medians; box limits denote 25th–75th percentile [Q1-Q3]; whiskers are drawn up to the smallest or largest observed value that was still within 1.5 times the interquartile range below the first quartile or above the third quartile, respectively; all other observed points are plotted as outliers). **h**, **i**, Kaplan-Meier OS analysis of TCGA CRC patients with low and high expression of *TAP1* (**h**) or *TAP2* (**i**) (log-rank (Mantel-Cox) test). The optimal cutoff values of *TAP1* and *TAP2* expression were determined with Cutoff Finder application. All bar plots (**a**-**e**) show

mean value ± s.d.. For **a**-**g**, *P* was derived with two-tailed unpaired *t* test; no adjustment for multiple comparisons.



Extended Data Figure 9. KDM5D interacts with Sin3-HDAC complex.

a, **b**, Western blot for co-IP experiment in HEK293T cells. '∅' indicates control vectors. WT, wild type, ED, enzymatic dead. **c**, qPCR of 2806 cell line and 2806 with overexpression of WT or ED form of mouse *Kdm5d*. *Kdm5d* relative expression was calculated by normalizing to '2806 mKDM5D^{WT} OE'; *Amot, Tap1 and Tap2* relative expression was calculated by normalizing to '2806' (n=3 wells per group; two-tailed unpaired *t* test; mean value ± s.d.; no adjustment for multiple comparisons).



Extended Data Figure 10. Schematic representation of the KRAS*-STAT4-KDM5D axis and the mechanisms of KDM5D driving metastasis in male-specific CRC.

CTL: $CD8^+$ cytotoxic T lymphocytes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Lucinette Marasigan for assistance with maintenance of the mouse colonies and Yutao Qi for supplying organoid culture medium. We thank Dr. Dipen Maru for help with tumour staging. We thank Drs. Kunal Rai, Michael Curran, Scott Kopetz, Guillermina Lozano, Jian Hu, Yejing Ge, Dihua Yu and Michelle Barton for insightful suggestions and criticisms. We thank Dr. Peiwen Chen, Dr. Pingping Hou, Dr. Yan Xia, Pingna Deng, Ko-Chien Chen, and all members of the DePinho laboratory for helpful discussions and support. We thank Dave

Aten, MA, CMI in the Department of Strategic Communications at The University of Texas MD Anderson Cancer Center, for the preparation of Extended Data Fig. 10. This study made use of MD Anderson's Research Histology, Pathology, and Imaging Core, the High-Resolution Electron Microscopy Facility; and the Advanced Cytometry & Sorting Core Facility supported by the National Institutes of Health (NIH) Cancer Center Support Grant P30 CA016672, the Advanced Technology Genomics Core supported by both NIH P30 CA016672 and NIH 1S10OD024977–01, and the Advanced Microscopy Core supported by NIH S10 RR029552. We thank the University of Texas Southwestern Medical Center (UTSW) Proteomics Core Facility for conducting the IP-MS. J.L. was supported by the Cancer Prevention and Research Institute of Texas (CPRIT) Research Training Program (RP210028). W.L. was supported by the National Natural Science Foundation of China (Nos. 81872401 and 82173289). K.A.L. was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under awards TL1TR003169 and UL1TR003167 and the CPRIT Research Training Program (RP210028). W.-H.H. was supported by the CPRIT Research Training Program (RP210028) and NIH F99/K00 (grant no. CA274661–01). R.L. was supported by NIH T32 (5T32CA186892–04). D.C. was supported by the Triumph postdoctoral training program at MD Anderson supported by CPRIT (RP170067) and by pilot funding through the Digestive Disease Center-National Institute of Diabetes and Digestive and Kidney Diseases award (P30CA16672). Work in R.A.D.'s laboratory was supported by NIH R01 CA231360, CPRIT (RP220364), the MD Anderson SPORE in Gastrointestinal Cancer-DRP Award, and the Harry Graves Burkhart III Distinguished University Chair in Cancer Research.

Data availability

The sequencing data generated in this study were deposited in the National Center for Biotechnology Information (NCBI) Bioproject database under accession number PRJNA773740 and GEO database under accession number GSE232459. Source data are provided in this paper. TCGA data from patients with CRC, including RNA-seq, mutation and survival data, were downloaded from cBioPortal Colorectal Adenocarcinoma Firehose Legacy (n=640) and the Human Protein Atlas (https://www.proteinatlas.org/). Integrated CRC patient data set by Colorectal Cancer Subtyping Consortium (CRCSC)[14] were downloaded from https://www.synapse.org (Synapse ID: syn2623706), including patient information such as sex, survival, mutation status and TNM (tumour, node, metastasis) stage. RNA-seq for iKAP mouse tumours (primary, metastatic, dox-on, and dox-off) and cell lines (dox-on and dox-off) were previously performed and deposited in NCBI Sequence Read Archive (SRA) SRP097890. Tsuji Colorectal Cancer RNA microarray data from primary and metastatic tumours were downloaded from Oncomine[56] (https://www.oncomine.org, also deposited in GEO under GSE28702). CRC cell line RNA-seq and mutation data from Cancer Cell Line Encyclopedia (CCLE) were downloaded from Depmap portal (https://depmap.org/portal/ccle/). Peptide identification in IP-MS used human protein database UniProt (https://www.uniprot.org/).

References

1. Haupt S, Caramia F, Klein SL, Rubin JB & Haupt Y Sex disparities matter in cancer development and therapy. Nat Rev Cancer 21, 393–407, doi:10.1038/s41568-021-00348-y (2021). [PubMed: 33879867]
2. Boutin AT et al. Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. Genes Dev 31, 370–382, doi:10.1101/gad.293449.116 (2017). [PubMed: 28289141]
3. Rubin JB et al. Sex differences in cancer mechanisms. Biol Sex Differ 11, 17, doi:10.1186/s13293-020-00291-x (2020). [PubMed: 32295632]
4. Mervic L Time course and pattern of metastasis of cutaneous melanoma differ between men and women. PLoS One 7, e32955, doi:10.1371/journal.pone.0032955 (2012). [PubMed: 22412958]
5. Kim SE et al. Sex- and gender-specific disparities in colorectal cancer risk. World J Gastroenterol 21, 5167–5175, doi:10.3748/wjg.v21.i17.5167 (2015). [PubMed: 25954090]

6. Li J, Ma X, Chakravarti D, Shalapour S & DePinho RA Genetic and biological hallmarks of colorectal cancer. Genes Dev 35, 787–820, doi:10.1101/gad.348226.120 (2021). [PubMed: 34074695]
7. Wood LD et al. The genomic landscapes of human breast and colorectal cancers. Science 318, 1108–1113, doi:10.1126/science.1145720 (2007). [PubMed: 17932254]
8. Liao W et al. KRAS-IRF2 Axis Drives Immune Suppression and Immune Therapy Resistance in Colorectal Cancer. Cancer Cell 35, 559–572 e557, doi:10.1016/j.ccell.2019.02.008 (2019). [PubMed: 30905761]
9. Sakai E et al. Combined Mutation of Apc, Kras, and Tgfbr2 Effectively Drives Metastasis of Intestinal Cancer. Cancer Res 78, 1334–1346, doi:10.1158/0008-5472.CAN-17-3303 (2018). [PubMed: 29282223]
10. Blair LP, Cao J, Zou MR, Sayegh J & Yan Q Epigenetic Regulation by Lysine Demethylase 5 (KDM5) Enzymes in Cancer. Cancers (Basel) 3, 1383–1404, doi:10.3390/cancers3011383 (2011). [PubMed: 21544224]
11. Li HT, Lu YY, An YX, Wang X & Zhao QC KRAS, BRAF and PIK3CA mutations in human colorectal cancer: relationship with metastatic colorectal cancer. Oncol Rep 25, 1691–1697, doi:10.3892/or.2011.1217 (2011). [PubMed: 21424126]
12. Modest DP et al. Clinical characterization of patients with metastatic colorectal cancer depending on the KRAS status. Anticancer Drugs 22, 913–918, doi:10.1097/CAD.0b013e3283493160 (2011). [PubMed: 21795973]
13. Mannan A & Hahn-Stromberg V K-ras mutations are correlated to lymph node metastasis and tumour stage, but not to the growth pattern of colon carcinoma. APMIS 120, 459–468, doi:10.1111/j.1600-0463.2011.02852.x (2012). [PubMed: 22583358]
14. Guinney J et al. The consensus molecular subtypes of colorectal cancer. Nat Med 21, 1350–1356, doi:10.1038/nm.3967 (2015). [PubMed: 26457759]
15. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM & Wallace MB Colorectal cancer. Lancet 394, 1467–1480, doi:10.1016/S0140-6736(19)32319-0 (2019). [PubMed: 31631858]
16. Banerjee S et al. Comparative analysis of clonal evolution among patients with right- and left-sided colon and rectal cancer. iScience 24, 102718, doi:10.1016/j.isci.2021.102718 (2021). [PubMed: 34258553]
17. Hugen N, van de Velde CJH, de Wilt JHW & Nagtegaal ID Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. Ann Oncol 25, 651–657, doi:10.1093/annonc/mdt591 (2014). [PubMed: 24504447]
18. Fornes O et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 48, D87–D92, doi:10.1093/nar/gkz1001 (2020). [PubMed: 31701148]
19. Visconti R et al. Importance of the MKK6/p38 pathway for interleukin-12-induced STAT4 serine phosphorylation and transcriptional activity. Blood 96, 1844–1852 (2000). [PubMed: 10961885]
20. Mehra-Chaudhary R, Matsui H & Raghow R Msx3 protein recruits histone deacetylase to down-regulate the Msx1 promoter. Biochem J 353, 13–22 (2001). [PubMed: 11115394]
21. Gao H et al. VentX, a novel lymphoid-enhancing factor/T-cell factor-associated transcription repressor, is a putative tumour suppressor. Cancer Res 70, 202–211, doi:10.1158/0008-5472.CAN-09-2668 (2010). [PubMed: 20028861]
22. Cobaleda C, Perez-Caro M, Vicente-Duenas C & Sanchez-Garcia I Function of the zinc-finger transcription factor SNAI2 in cancer and development. Annu Rev Genet 41, 41–61, doi:10.1146/annurev.genet.41.110306.130146 (2007). [PubMed: 17550342]
23. Harrison DA The Jak/STAT pathway. Cold Spring Harb Perspect Biol 4, doi:10.1101/cshperspect.a011205 (2012).
24. Palmer DC & Restifo NP Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. Trends Immunol 30, 592–602, doi:10.1016/j.it.2009.09.009 (2009). [PubMed: 19879803]
25. Yi C et al. A tight junction-associated Merlin-angiomotin complex mediates Merlin's regulation of mitogenic signaling and tumour suppressive functions. Cancer Cell 19, 527–540, doi:10.1016/j.ccr.2011.02.017 (2011). [PubMed: 21481793]

26. Dusek RL & Attardi LD Desmosomes: new perpetrators in tumour suppression. Nat Rev Cancer 11, 317–323, doi:10.1038/nrc3051 (2011). [PubMed: 21508970]
27. Bratt A et al. Angiomotin regulates endothelial cell-cell junctions and cell motility. J Biol Chem 280, 34859–34869, doi:10.1074/jbc.M503915200 (2005). [PubMed: 16043488]
28. Zihni C, Mills C, Matter K & Balda MS Tight junctions: from simple barriers to multifunctional molecular gates. Nat Rev Mol Cell Biol 17, 564–580, doi:10.1038/nrm.2016.80 (2016). [PubMed: 27353478]
29. McNeil E, Capaldo CT & Macara IG Zonula occludens-1 function in the assembly of tight junctions in Madin-Darby canine kidney epithelial cells. Mol Biol Cell 17, 1922–1932, doi:10.1091/mbc.e05-07-0650 (2006). [PubMed: 16436508]
30. Tsukita S, Yamazaki Y, Katsuno T, Tamura A & Tsukita S Tight junction-based epithelial microenvironment and cell proliferation. Oncogene 27, 6930–6938, doi:10.1038/onc.2008.344 (2008). [PubMed: 19029935]
31. Whyte WA et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319, doi:10.1016/j.cell.2013.03.035 (2013). [PubMed: 23582322]
32. Loven J et al. Selective inhibition of tumour oncogenes by disruption of super-enhancers. Cell 153, 320–334, doi:10.1016/j.cell.2013.03.036 (2013). [PubMed: 23582323]
33. Mumbach MR et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods 13, 919–922, doi:10.1038/nmeth.3999 (2016). [PubMed: 27643841]
34. Morrison BJ, Steel JC & Morris JC Reduction of MHC-I expression limits T-lymphocyte-mediated killing of Cancer-initiating cells. BMC Cancer 18, 469, doi:10.1186/s12885-018-4389-3 (2018). [PubMed: 29699516]
35. Seliger B, Maeurer MJ & Ferrone S Antigen-processing machinery breakdown and tumour growth. Immunol Today 21, 455–464, doi:10.1016/s0167-5699(00)01692-3 (2000). [PubMed: 10953098]
36. Lankat-Buttgereit B & Tampe R The transporter associated with antigen processing: function and implications in human diseases. Physiol Rev 82, 187–204, doi:10.1152/physrev.00025.2001 (2002). [PubMed: 11773612]
37. Kaklamanis L et al. Loss of major histocompatibility complex-encoded transporter associated with antigen presentation (TAP) in colorectal cancer. Am J Pathol 145, 505–509 (1994). [PubMed: 8080034]
38. Barbosa AM et al. Increased CD3(+), CD8(+), or FoxP3(+) T Lymphocyte Infiltrations Are Associated with the Pathogenesis of Colorectal Cancer but Not with the Overall Survival of Patients. Biology (Basel) 10, doi:10.3390/biology10080808 (2021).
39. Mlecnik B et al. The tumour microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. Sci Transl Med 8, 327ra326, doi:10.1126/scitranslmed.aad6352 (2016).
40. Li N et al. ZMYND8 Reads the Dual Histone Mark H3K4me1-H3K14ac to Antagonize the Expression of Metastasis-Linked Genes. Mol Cell 63, 470–484, doi:10.1016/j.molcel.2016.06.035 (2016). [PubMed: 27477906]
41. Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P & Reinberg D Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell 89, 357–364, doi:10.1016/s0092-8674(00)80216-0 (1997). [PubMed: 9150135]
42. Alland L et al. Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. Nature 387, 49–55, doi:10.1038/387049a0 (1997). [PubMed: 9139821]
43. Laherty CD et al. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell 89, 349–356, doi:10.1016/s0092-8674(00)80215-9 (1997). [PubMed: 9150134]
44. Martin TA The role of tight junctions in cancer metastasis. Semin Cell Dev Biol 36, 224–231, doi:10.1016/j.semcdb.2014.09.008 (2014). [PubMed: 25239399]
45. Cromme FV et al. Differences in MHC and TAP-1 expression in cervical cancer lymph node metastases as compared with the primary tumours. Br J Cancer 69, 1176–1181, doi:10.1038/bjc.1994.231 (1994). [PubMed: 8198988]
46. Garcia-Lora A, Martinez M, Algarra I, Gaforio JJ & Garrido F MHC class I-deficient metastatic tumour variants immunoselected by T lymphocytes originate from the coordinated downregulation

of APM components. Int J Cancer 106, 521–527, doi:10.1002/ijc.11241 (2003). [PubMed: 12845647]

47. Garrido F, Aptsiauri N, Doorduijn EM, Garcia Lora AM & van Hall T The urgent need to recover MHC class I in cancers for effective immunotherapy. Curr Opin Immunol 39, 44–51, doi:10.1016/j.coi.2015.12.007 (2016). [PubMed: 26796069]
48. Setiadi AF et al. Epigenetic enhancement of antigen processing and presentation promotes immune recognition of tumours. Cancer Res 68, 9601–9607, doi:10.1158/0008-5472.CAN-07-5270 (2008). [PubMed: 19047136]
49. Klose RJ, Kallin EM & Zhang Y JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet 7, 715–727, doi:10.1038/nrg1945 (2006). [PubMed: 16983801]
50. Li N et al. JARID1D Is a Suppressor and Prognostic Marker of Prostate Cancer Invasion and Metastasis. Cancer Res 76, 831–843, doi:10.1158/0008-5472.CAN-15-0906 (2016). [PubMed: 26747897]
51. Heijmans J et al. ER stress causes rapid loss of intestinal epithelial stemness through activation of the unfolded protein response. Cell Rep 3, 1128–1139, doi:10.1016/j.celrep.2013.02.031 (2013). [PubMed: 23545496]
52. Drost J et al. Sequential cancer mutations in cultured human intestinal stem cells. Nature 521, 43–47, doi:10.1038/nature14415 (2015). [PubMed: 25924068]
53. Terranova C et al. An Integrated Platform for Genome-wide Mapping of Chromatin States Using High-throughput ChIP-sequencing in Tumour Tissues. J Vis Exp, doi:10.3791/56972 (2018).
54. Servant N et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16, 259, doi:10.1186/s13059-015-0831-x (2015). [PubMed: 26619908]
55. Budczies J et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One 7, e51862, doi:10.1371/journal.pone.0051862 (2012). [PubMed: 23251644]
56. Rhodes DR et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 6, 1–6, doi:10.1016/s1476-5586(04)80047-2 (2004). [PubMed: 15068665]

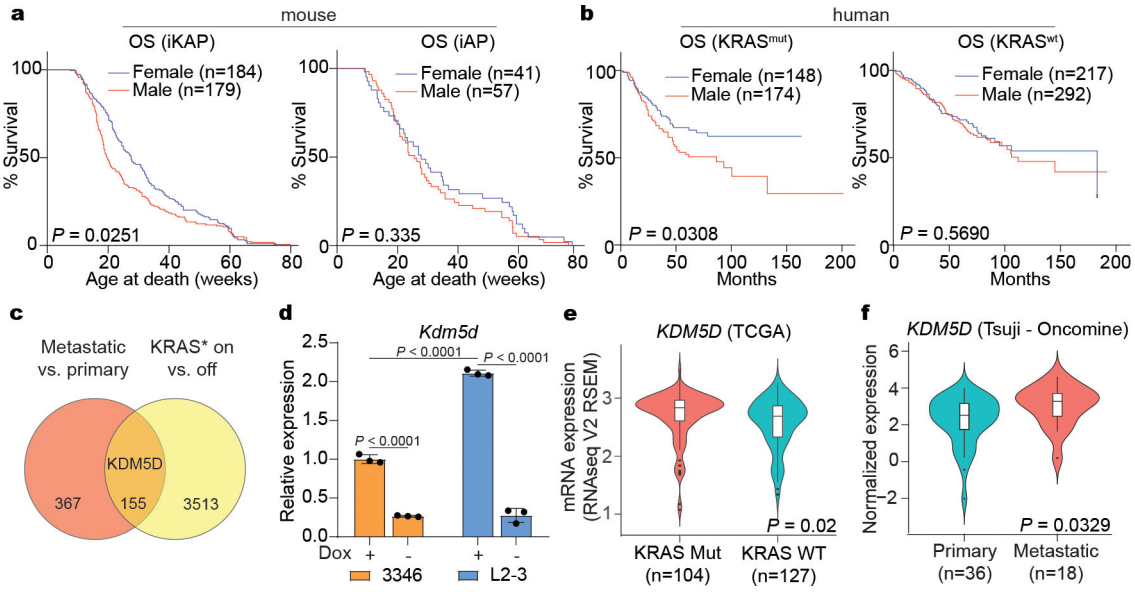


Figure 1. KDM5D is correlated with sex-specific KRAS* CRC metastasis.

a, Kaplan-Meier overall survival (OS) analyses of iKAP and iAP male and female mice. **b**, Kaplan-Meier OS analyses of CRC patients with wild-type (WT) or mutant (mut) *KRAS* in an integrated data set from Colorectal Cancer Subtyping Consortium (CRCSC) (for **a** and **b**, Log-rank (Mantel-Cox) test). **c**, Intersection analysis of differentially expressed genes in two comparisons. Differential genes were filtered with fold change > 1.5 and $P < 0.05$. KRAS*: oncogenic KRAS mutation. Database: SRP097890. **d**, Quantitative PCR (qPCR) measuring expression of *Kdm5d* in cell lines derived from male iKAP distal primary CRC (3346) and matched liver metastasis (L2–3) cultured in medium with doxycycline (dox) on or off. Relative expression was calculated by normalizing to '3346 dox on' (n=3 wells per group; two-tailed unpaired t test; mean value ± s.d.; no adjustment for multiple comparisons). **e**, Expression of *KDM5D* in TCGA RNA-seq data of male CRC patients. All *KRAS* mutations were included (the most common ones were G12D, G13D and G12V); patients were not filtered based on *APC* and *TP53* mutations. **f**, Expression of *KDM5D* in male CRC patients in Tsuji Colorectal Cancer RNA microarray data set downloaded from Oncomine database. Metastasis sites included liver (n=14), lung (n=1) and peritoneum (n=3) (for **e** and **f**, two-tailed unpaired t test; box plots, center lines denote medians; box limits denote 25th–75th percentile [Q1-Q3]; whiskers are drawn up to the smallest or largest observed value that was still within 1.5 times the interquartile range below the first quartile or above the third quartile, respectively; all other observed points are plotted as outliers).

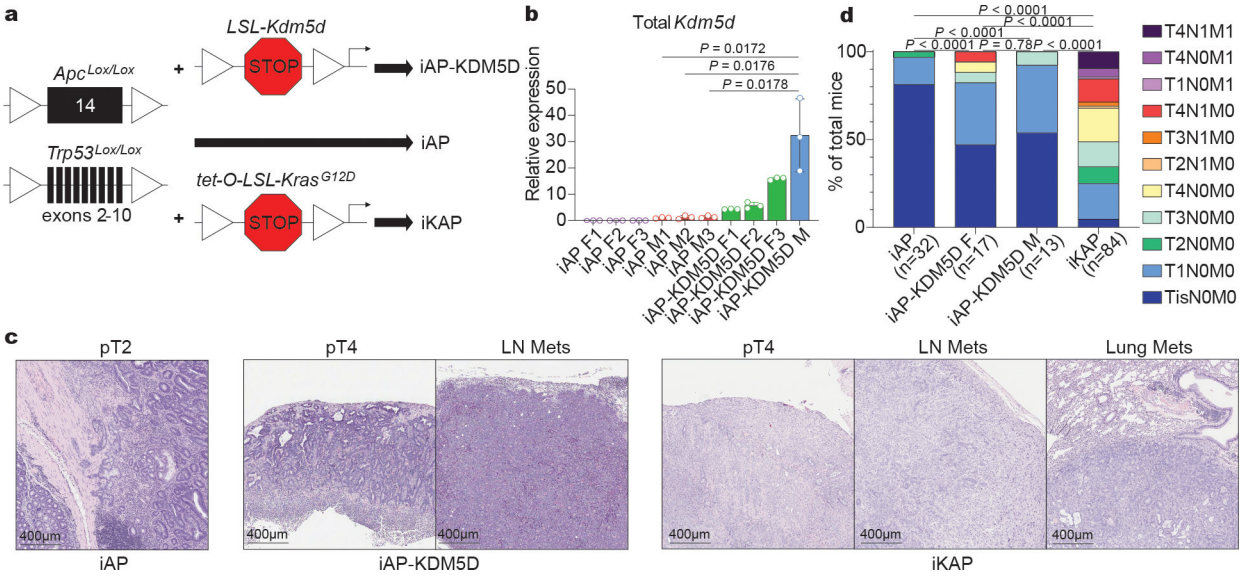


Figure 2. *Kdm5d* transgene expression promotes tumour invasiveness and metastasis.

a, Schemes of three genetically engineered mouse models. **b**, qPCR of total *Kdm5d* expression in formalin-fixed paraffin-embedded (FFPE) tissues of primary tumours. Relative expression was calculated by normalizing to the 'iAP M1' sample and housekeeping gene *Gapdh* (*n*=3 female iAP mice, *n*=3 male iAP mice, *n*=3 female iAP-KDM5D mice, *n*=1 male iAP-KDM5D mouse; unpaired two-tailed *t* test; mean value ± s.d.; no adjustment for multiple comparisons). **c**, Representative images of the most invasive primary tumours and, if any, lymph node (LN) and distal metastases. iAP and iAP-KDM5D, female; iKAP, male. **d**, Percentage of TNM (tumour, node, metastasis) stages of tumours from iAP (male and female combined), iAP-KDM5D male and female and iKAP (male and female combined) mice (two-tailed Chi-square test).

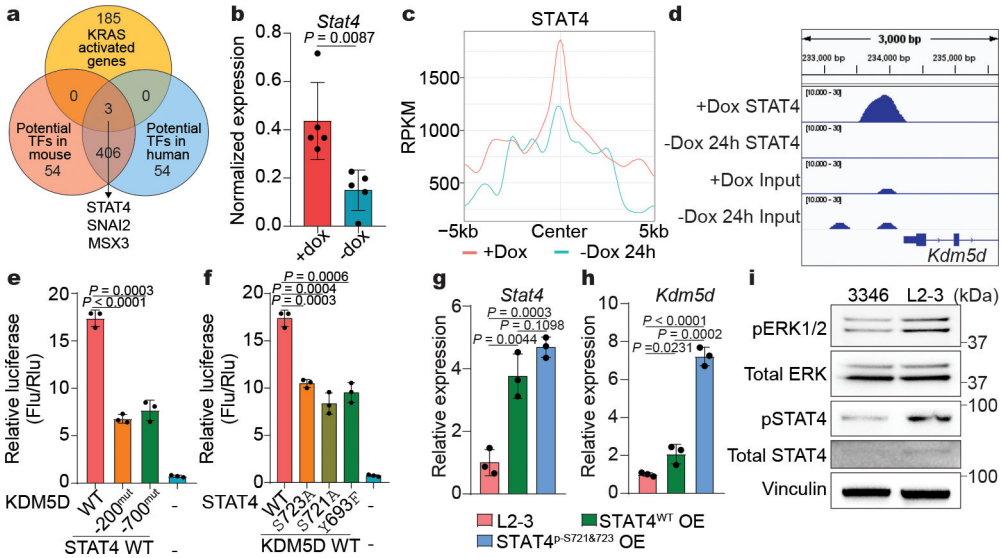


Figure 3. KRAS* drives KDM5D expression through STAT4.

a, Intersection of genes upregulated in KRAS* on versus off iKAP cell RNA-seq and the predicted transcription factors (TFs) bound to mouse *Kdm5d* or human *KDM5D* promoters. Differential expression in RNA-seq was filtered with fold change > 1.5, $P < 0.05$. RNA-seq data were retrieved from SRP097890. **b**, *Stat4* expression (z-score) in RNA-seq data of **a** (n=5 pairs of matched dox (KRAS*) on or off iKAP cells; two-tailed paired t test). **c**, Metaplots of STAT4 ChIP-seq in male iKAP cells with dox on/off. Average signal in the window of 5 kb up-/down-stream transcription start sites (TSSs) was used for two-tailed Mann-Whitney test. $P < 0.0001$. RPKM, reads per kilobase of transcript per million mapped reads. **d**, ChIP-seq tracks of STAT4 IP and Input around *Kdm5d* promoter in male iKAP cells with dox on/off. bp, base pair. **e**, **f**, Luciferase assays of WT mouse STAT4 and luciferase reporter driven by WT or mutant *Kdm5d* promoters (mutations at putative STAT4 binding site ~ 200bp and 700bp upstream of TSS) (**e**), and phospho-dead (S723A, S721A, Y693F) or WT mouse STAT4 and luciferase reporter driven by WT *Kdm5d* promoter (**f**) (n=3 wells per group; two-tailed unpairs t test; no adjustment for multiple comparisons). Flu, firefly luciferase activity; Rlu, Renilla luciferase activity. **g**, **h**, qPCR of *Stat4* (**g**) and the STAT4 driven *Kdm5d* expression (**h**). Relative expression was calculated by normalizing to ‘L2–3’ (n=3 wells per group; two-tailed unpaired t test; no adjustment for multiple comparisons). **i**, Western blot of iKAP cells. The samples were derived from the same experiment and gel/blots were processed in parallel. For all bar plots (**b**, **e**, **f**, **g**, **h**), data are mean value ± s.d..

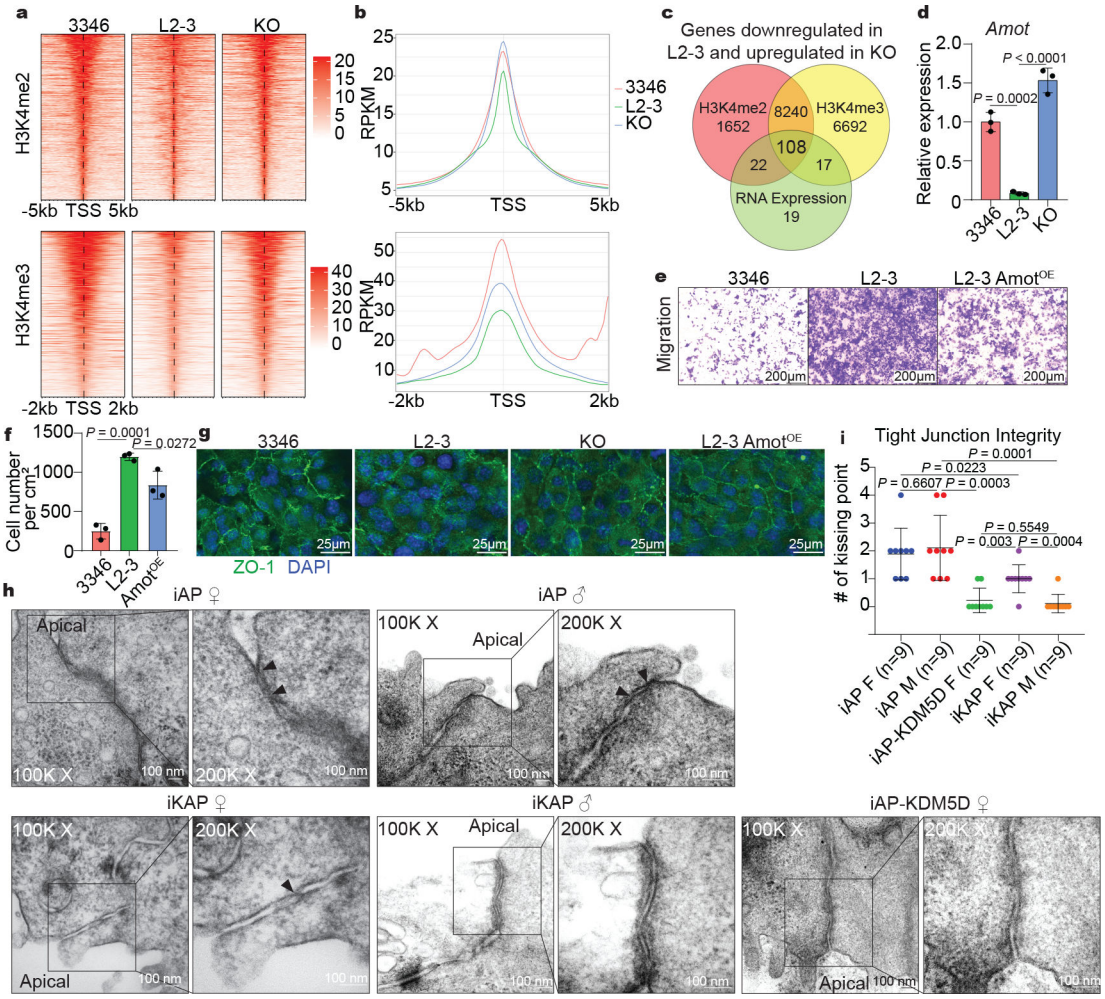


Figure 4. KDM5D impairs tight junctions through epigenetically repressing AMOT.

a, Heatmaps of H3K4me2/3 in iKAP cells. KO, knockout. **b**, Metaplots of H3K4me2/3 corresponding to **a**. Average signal in the window of 1kb up-/down-stream of TSSs (gene promoter area) was used for two-tailed Mann-Whitney test. H3K4me2: 3346 versus L2–3, $P = 0.002$, L2–3 versus KO, $P = 0.0051$, 3346 versus KO, $P = 0.968$; H3K4me3: 3346 versus L2–3, $P < 0.0001$, L2–3 versus KO, $P = 0.0005$, 3346 versus KO, $P = 0.0041$. **c**, Intersection analysis. **d**, qPCR of *Amot* in iKAP cells. Relative expression was calculated by normalizing to '3346' (n=3 wells per group). **e**, Representative images of migration assay of iKAP cells. OE, overexpression. **f**, Quantitation of **e** (n=3 wells per group). **g**, Representative confocal imaging of immunofluorescence staining. **h**, Representative transmission electron microscopy imaging of primary iKAP tumours. The apical side of intestinal epithelial cell was labeled. The number of images taken is indicated in **i**. Scale bars, 100 nm. **i**, Number of visible 'kissing points' within tight junction structure (n=9 tumours per group). All bar/dot plots (**d**, **f**, **i**) show mean value ± s.d.; P was derived with two-tailed unpaired t test; no adjustment for multiple comparisons.

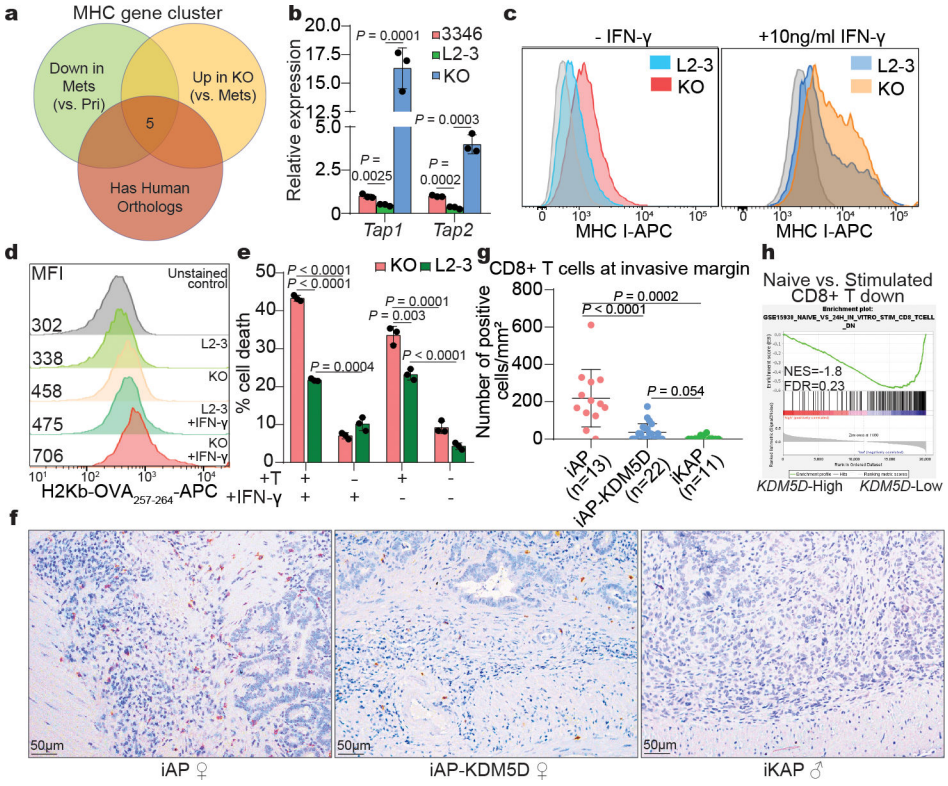


Figure 5. KDM5D represses MHC class I antigen processing and presentation.

a, Intersection analysis of all MHC cluster genes that were downregulated in L2–3 versus 3346 and upregulated in KO versus L2–3 and that have human orthologs. **b**, qPCR of *Tap1* and *Tap2* in iKAP cells. Relative expression was calculated by normalizing to '3346' (n=3 wells per group). **c**, Flow cytometry measurement of cell surface MHC-I complex in iKAP cells +/− 10 ng/ml interferon-gamma (IFN**γ**) overnight treatment. Median fluorescence intensity (MFI): - IFN**γ**: unstained (grey): 527, L2–3 (blue): 765, KO (red): 1297; + IFN**γ**: unstained (grey): 2048, L2–3 (dark blue): 3731, KO (orange): 5824. **d**, Flow cytometry measurement of cell surface OVA 257–264 (SIINFEKL) peptide bound MHC-I complex in iKAP cells treated with 2 mg of ovalbumin protein and +/− 10 ng/mL IFN**γ**. In **c** and **d**, the x-axis is the fluorescence intensity, the y-axis is the count of events scaled as the percentage of the maximum count. **e**, Percentage of dead iKAP cells (GFP+ for iKAP cells and SYTOX Blue+ for dead cells) in $CD8^+$ T cell killing assay (Effector: target = 10: 1) (n=3 wells per group). **f**, Representative images of immunohistochemistry (IHC) for CD8 at the tumour invasive front in iAP, iAP-KDM5D and iKAP primary tumours with comparable invasiveness (pTis and pT1). **g**, Quantitation of the number of $CD8^+$ cells at the tumour invasive front detected by IHC. **h**, GSEA of TCGA male CRC RNA-seq data using gene set 'c7.all.v7.4.symbols.gmt [Immunologic signatures]'. *KDM5D*-High or -Low were determined following the same criteria as in Extended Data Fig. 3l. For **b**, **e**, **g**, Data are mean value ± s.d.; P was derived with two-tailed unpaired t test; no adjustment for multiple comparisons. FDR, false discovery rate; NES, normalized enrichment score.