LETTER

doi:10.1038/nature25492

TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis

Daniele V. F. Tauriello[1,2], Sergio Palomo-Ponce[1,2], Diana Stork[1], Antonio Berenguer-Llergo[1], Jordi Badia-Ramentol[1], Mar Iglesias[2,3,4,5], Marta Sevillano[1,2], Sales Ibiza[1], Adrià Cañellas[1], Xavier Hernando-Momblona[1,2], Daniel Byrom[1], Joan A. Matarin[1], Alexandre Calon[1]†, Elisa I. Rivas[1]†, Angel R. Nebreda[1,6], Antoni Riera[1,7], Camille Stephan-Otto Attolini[1] & Eduard Batlle[1,2,6]

Most patients with colorectal cancer die as a result of the disease spreading to other organs. However, no prevalent mutations have been associated with metastatic colorectal cancers[1,2]. Instead, particular features of the tumour microenvironment, such as lack of T-cell infiltration[3], low type 1 T-helper cell (T_H1) activity and reduced immune cytotoxicity[2] or increased TGFβ levels[4] predict adverse outcomes in patients with colorectal cancer. Here we analyse the interplay between genetic alterations and the tumour microenvironment by crossing mice bearing conditional alleles of four main colorectal cancer mutations in intestinal stem cells. Quadruple-mutant mice developed metastatic intestinal tumours that display key hallmarks of human microsatellite-stable colorectal cancers, including low mutational burden[5], T-cell exclusion[3] and TGFβ-activated stroma[4,6,7]. Inhibition of the PD-1–PD-L1 immune checkpoint provoked a limited response in this model system. By contrast, inhibition of TGFβ unleashed a potent and enduring cytotoxic T-cell response against tumour cells that prevented metastasis. In mice with progressive liver metastatic disease, blockade of TGFβ signalling rendered tumours susceptible to anti-PD-1–PD-L1 therapy. Our data show that increased TGFβ in the tumour microenvironment represents a primary mechanism of immune evasion that promotes T-cell exclusion and blocks acquisition of the T_H1-effector phenotype. Immunotherapies directed against TGFβ signalling may therefore have broad applications in treating patients with advanced colorectal cancer.

Progression of colorectal cancer (CRC) generally coincides with successive alterations in four signalling pathways: WNT, EGFR, p53 and TGFβ[5,8]. Mice bearing compound mutations in these four pathways were recently shown to enable the study of CRC metastasis[9–11]. We crossed mice bearing conditional alleles in homologues of four key human CRC mutations: $Apc^{fl/fl}$, $Kras^{LSL\text{-}G12D}$, $Tgfbr2^{fl/fl}$ and $Trp53^{fl/fl}$ (designated A, K, T and P, respectively)[12–15], and targeted gene recombination to intestinal stem cells (ISCs) by means of the $Lgr5^{eGFP\text{-}creERT2}$ driver[16], which we designated L. We generated eight mouse strains bearing combinations of these mutations (Fig. 1a). Histopathological scoring demonstrated a stepwise increase in prevalence and severity of invasive adenocarcinomas along the linear progression sequence (Fig. 1b, c and Extended Data Fig. 1c–i). Ninety per cent of LAKTP mice developed carcinomas, more than half of which breached all intestinal layers (Fig. 1c). These cancers displayed a histology similar to human tumours, with mostly medium to high degrees of differentiation and abundant desmoplastic reaction (Extended Data Fig. 1d–j). Forty per cent of LAKTP mice developed metastases in the liver or lungs, or as carcinomatosis (Fig. 1d and Extended Data Fig. 1k–n), with a median latency of 66 days. Notably, mice bearing triple-mutant genotypes (LAKT, LATP or LAKP) presented with similarly invasive cancers but not metastasis (Fig. 1c). In LAKTP mice, T cells extensively infiltrated the stroma of normal mucosa and adenomas but were largely excluded from adjacent invasive cancers (Fig. 1e and Extended Data Fig. 2a). This exclusion phenotype intensified along the CRC mutational sequence (Fig. 1e). Invasive margins of compound-mutant cancers displayed high levels of stromal TGFβ activity, as indicated by the presence of phosphorylated SMAD3 (pSMAD3) (Fig. 1f and Extended Data Fig. 2b) and expression of CALD1 and IGFBP7 (Extended Data Fig. 2c, d)—two TGFβ-induced genes expressed in cancer-associated fibroblasts (CAFs) that predict poor prognosis[4]. These genetic CRC models therefore reproduce key features of the tumour microenvironment (TME) in advanced human CRCs.

We collected fresh samples from multiple primary mouse tumours or metastases and established a mouse tumour organoid (MTO) biobank (Fig. 1g). Quadruple-mutant LAKTP MTOs expanded *in vitro* independently of factors that stimulate WNT, EGF and TGFβ pathways (Fig. 1h and Extended Data Fig. 3). When injected into the caecum wall of syngeneic C57BL/6J recipients, these MTOs engrafted with a success rate of 10 out of 32 (31%) and progressed to fully invasive T3–T4 tumours (Extended Data Fig. 4a, b). Forty per cent of engrafted MTOs produced overt liver metastatic disease. We observed prominent T-cell exclusion and increased TGFβ activity in the TME in primary tumours from orthotopically transplanted LAKTP MTOs (Extended Data Fig. 4c–f). TGFβ-activated stroma is a defining feature of poor prognosis consensus molecular subtype 4 (CMS4)[7]. Transcriptomic classification indicated that LAKTP MTOs displayed the CMS4 phenotype when transplanted in the caecum of syngeneic mice, but not when cultured *in vitro* (Fig. 1i and Extended Data Fig. 5). We obtained similar results with triple-mutant MTOs. These findings confirm that stromal gene expression is required to identify poor prognosis molecular subtypes[4,17–19] (see Supplementary Discussion).

Similar to microsatellite-stable (MSS) human CRCs[20], triple- and quadruple-mutant MTOs accumulated between 0.5 and 3.5 non-synonymous coding mutations per megabase (Extended Data Fig. 6a, b), which indicates that the genomes of these tumours evolved extensively. The most prevalent mutational signature in MTOs was signature 1, characteristic of MSS CRCs[21] (Extended Data Fig. 6c). MTOs and human MSS CRCs also displayed similar numbers of predicted high-affinity MHC-I-binding neoantigens, whereas mouse CRC cell lines CT26 and MC-38 exhibited around 40 times this number (Fig. 1j). Experimental liver colonization by MTOs caused an increased metastatic burden in the *nu/nu* background compared to wild-type C57BL/6J mice (Fig. 1k), indicating that the tumours were susceptible to T-cell-mediated adaptive immunosurveillance. Indeed,

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri i Reixac 10, 08028 Barcelona, Spain. [2]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Barcelona, Spain. [3]Department of Pathology, Hospital del Mar, 08003 Barcelona, Spain. [4]Cancer Research Programme, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, Spain. [5]Autonomous University of Barcelona (UAB), Spain. [6]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain. [7]Department of Organic Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain. †Present address: Cancer Research Programme, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, Spain (A.C. and E.I.R.).

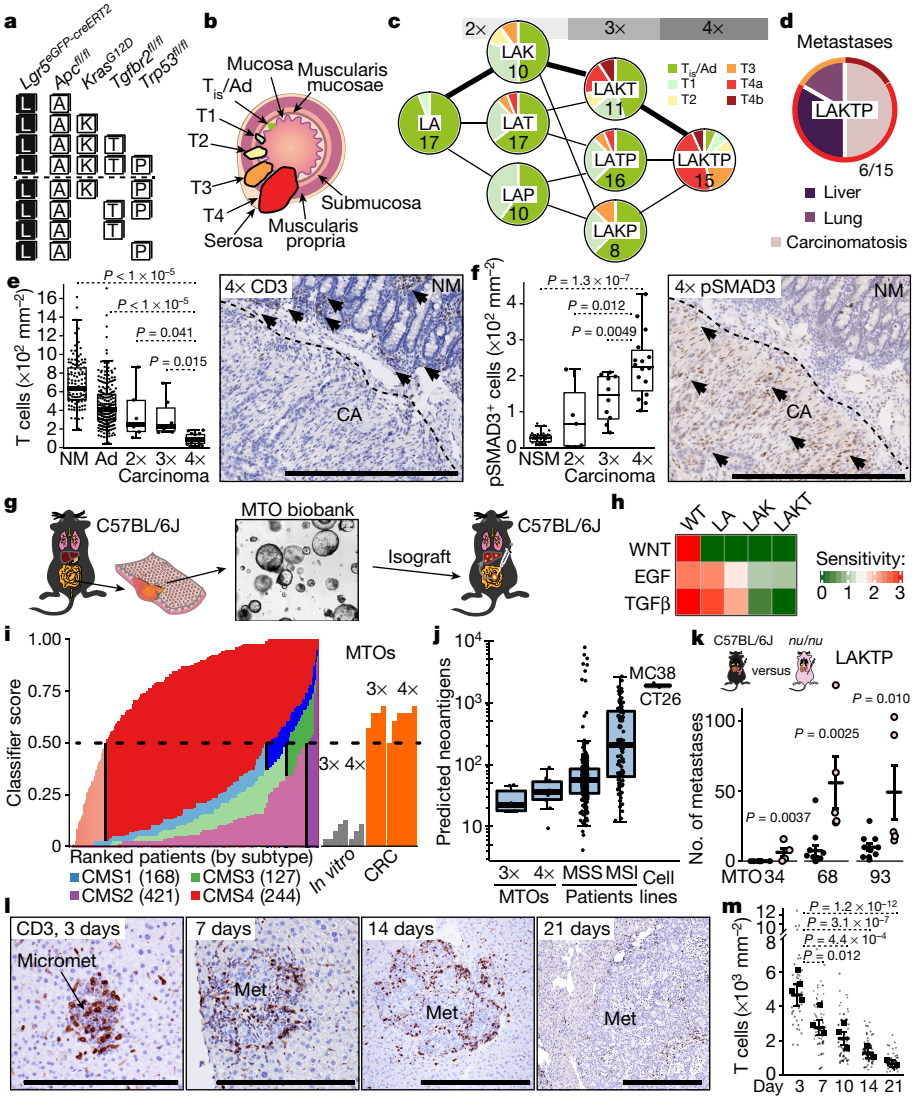


Figure 1 | Analysis of compound mutant mouse models and MTOs. **a**, Combination of alleles used to generate mouse models; dashed line distinguishes strains along the linear progression model[8] of A–K–T–P from other combinations of mutations. **b**, Classification of tumours according to the AJCC–TNM system, ranging from carcinoma *in situ* or adenoma (T_{is}/Ad) to four levels of increasingly invasive cancers (T1–T4, see Methods). **c**, Worst T-phenotype diagnosis per mouse, number of mice is indicated in the circles. The thick lines indicate the linear progression sequence. **d**, Metastasis in LAKTP mice. Outer ring, associated type of primary cancer. **e**, Box plot of T-cell density in normal mucosa (NM) and adenoma (Ad), compared to double-mutant (LAK, LAT and LAP, 2×), triple-mutant (LAKT, LAKP and LATP, 3×) and quadruple-mutant (LAKTP, 4×) cancers. $n = 119$, 242, 7, 8 and 16 regions of interest (ROIs), left to right. Right, CD3 immunohistochemistry of a LAKTP T4 carcinoma (CA), arrowheads indicate $CD3^+$ cells. Box plots have whiskers of maximum 1.5 times the interquartile range; the boxes represent first, second (median) and third quartiles. **f**, Box plot of $pSMAD3^+$ cell density in normal submucosal tissue (NSM), compared to 2×, 3× and 4× cancers, $n = 37$, 5, 10 and 16 ROIs, left to right. Right, pSMAD3 immunohistochemistry of a LAKTP T4 carcinoma, arrowheads indicate $pSMAD3^+$ cells. Two-sided Welch's *t*-test (**e**, **f**). **g**, MTO biobank and isograft technology. **h**, Mean sensitivity of MTOs to niche factors. WT, wild type. **i**, Left, cross-validation of consensus molecular subtype 4 (CMS4) classifier on human CRCs. Patients (shown as vertical bars) are ranked by subtype on the basis of classifier score. Right, classifier applied to 3× and 4× mutant MTOs *in vitro* (grey, $n = 5$ and 3, respectively) and to orthotopic isografts of those MTOs (orange, $n = 2$ and 3). **j**, Predicted MHC-I neoantigens in MTOs ($n = 6$ and 10, left and right, respectively), human CRCs ($n = 266$ and 112, left and right, respectively) and mouse CRC cell lines ($n = 2$). **k**, Liver metastasis generated by 4× MTOs in C57BL/6J or *nu/nu* mice ($n = 10$, 5, 11, 5, 10 and 5 mice, left to right; mean ± s.e.m.; two-sided Mann–Whitney *U*-test). **l**, CD3 immunohistochemistry on MTO129 liver metastases (Met) at the indicated days after intrasplenic injection. Micromet, micrometastasis. **m**, $CD3^+$ cell densities (individual metastases in grey and mean for mice in black; bars, group means ± s.e.m.); analysed with a mixed-effects linear model; $n = 4$, 3, 3, 2 and 2 mice, and $n = 39$, 58, 71, 56 and 56 tumours, left to right. Scale bars, 500 μm.

micrometastases were characterized by abundant $CD3^+$ T cells intermingled with tumour cells (Fig. 1l). Notably, T cells were progressively excluded at later time points (Fig. 1l, m).

Cell-population profiling of human or mouse CRC samples revealed that CAFs were the main contributors to TGFβ production (Extended Data Fig. 7a, b). We made use of the TGFBR1-specific inhibitor galunisertib[22] to inhibit TGFβ signalling in the TME. Treatment with galunisertib starting 11 days after transplantation of LAKTP MTOs in the caecum of syngeneic mice reduced primary tumour size, reduced the extent of carcinomatosis and blocked the appearance of liver metastases (Fig. 2a). Immunohistochemical quantification showed that galunisertib reduced the number of $pSMAD3^+$ cells (Fig. 2b) and gene expression profiling demonstrated decreased levels of TGFβ-response signatures[6] in fibroblasts (F-TBRS) and T cells (T-TBRS) (Fig. 2c).

To directly test therapeutic effects on liver metastatic disease, we inoculated LAKTP MTOs derived from either primary CRCs or liver

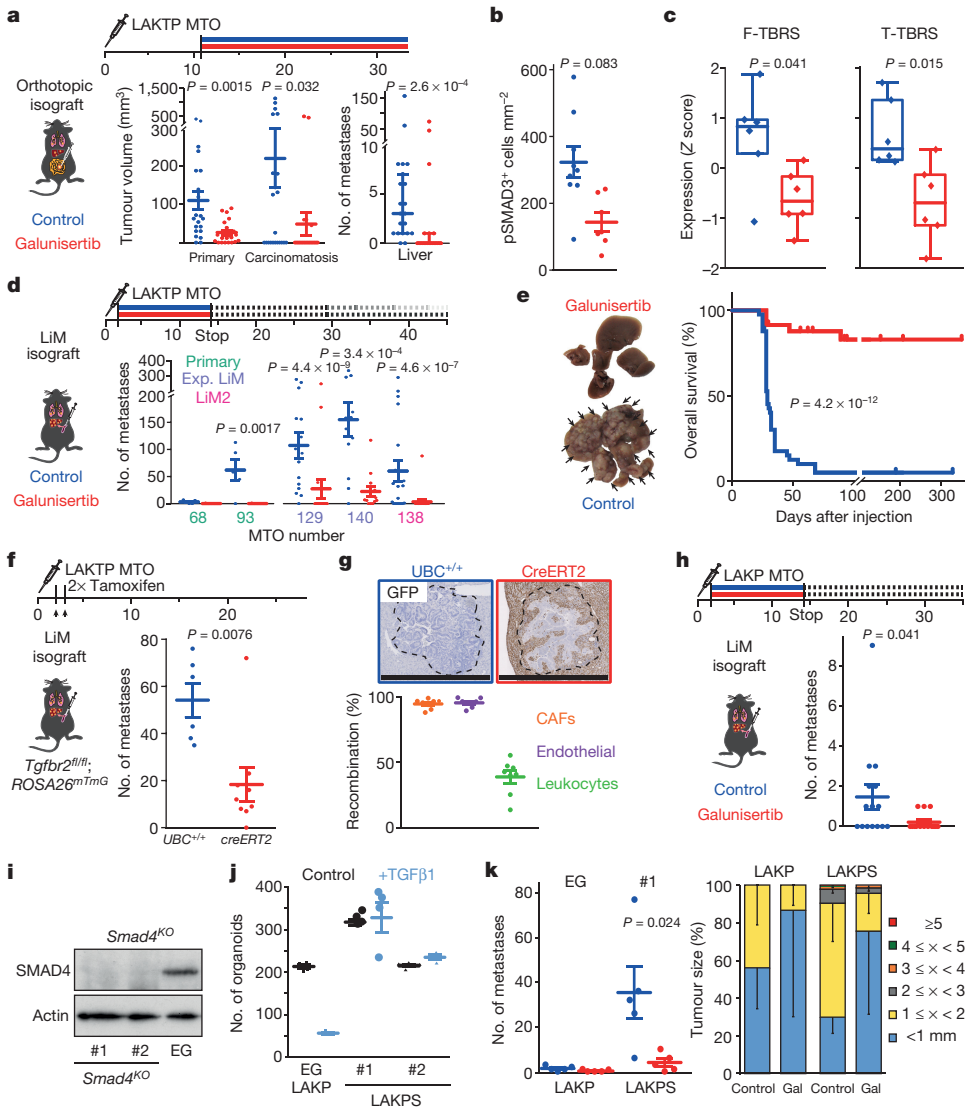


Figure 2 | Therapeutic effect of TGFβ inhibition. a, Galunisertib reduces tumour burden and metastasis in orthotopic MTO140 isografts. Middle, mean tumour volume ± s.e.m.; right, median number ± 95% confidence interval; $n = 21$ mice per condition, two-sided Mann–Whitney *U* test. **b**, pSMAD3$^+$ cell density (mean ± s.e.m.) in primary carcinomas ($n = 9, 7$); two-sided Student's *t*-test. **c**, Expression levels of TBRS in primary CRCs. Tukey box plots, $n = 6$ tumour samples, two-sided Mann–Whitney *U* test. **d**, Liver metastases (mean ± s.e.m.) generated by MTOs, treated with vehicle or galunisertib; $n = 5, 5, 5, 5, 15, 17, 12, 12, 20$ and 25 mice, left to right; analysed with a mixed-effects linear model. Exp. LiM, experimental liver metastasis; LiM2, second-round experimental liver metastasis. **e**, Representative livers at end point, arrows point to metastases. Kaplan–Meier survival curve of mice treated as in **d**. Control, $n = 41$ mice; galunisertib, $n = 35$ mice; Mantel–Cox test. **f**, Liver metastases (mean ± s.e.m.) in $UBC^{creERT2}$;$Tgfbr2^{fl/fl}$ mice, recombined ($n = 9$) or not ($n = 6$). **g**, Liver tumours from **f**, stained for the recombination marker GFP, representative of three experiments. Scale bars, 1 mm. The graph shows the percentage of GFP$^+$ cells (mean ± s.e.m.). **h**, Liver metastases (mean ± s.e.m.) generated by 3× LAKP MTOs. $n = 15$ and 14 mice, left and right, respectively; two-sided Mann–Whitney *U* test. **i**, Western blot for SMAD4 and actin of LAKP empty guide (EG)- or CRISPR-targeted (#1 and #2) $Smad4^{KO}$ organoids. **j**, LAKP or LAKPS organoids. (mean ± s.e.m., $n = 4$ technical replicates) in control medium or treated with TGFβ1. **k**, Liver metastases (mean ± s.e.m.) from LAKP (EG) and LAKPS (guide #1) MTOs. Gal, galunisertib; *x*, diameter. $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. Right, tumour diameters with s.e.m., $n = 10$ mice per condition.

metastases in the portal circulation; the latter exhibited enhanced metastatic capacity (Extended Data Fig. 6e). Treatment with galunisertib markedly decreased metastatic burden for all MTOs (Fig. 2d and Extended Data Fig. 6f), effectively curing a large proportion of mice (Fig. 2e). Galunisertib blocked TGFβ signalling in the TME of metastases (Extended Data Fig. 7c–h). We also transplanted LAKTP MTOs in mice bearing floxed *Tgfbr2* alleles that express a ubiquitous CreERT2 recombinase ($UBC^{creERT2}$;$Tgfbr2^{fl/fl}$). Tamoxifen treatment induced recombination in fibroblasts, endothelial cells and leukocytes, and inhibited metastasis formation by isografted MTOs (Fig. 2f, g and Extended Data Fig. 8a). To analyse triple-mutant MTOs *in vivo*, we inoculated large numbers of cells. Four out of six tested MTOs (two LAKP and two LAKT) gave rise to metastases. However, these liver tumours remained very small (Extended Data Fig. 6g), possibly owing to niche factor dependencies[9,11,19]. Nevertheless, galunisertib treatment reduced liver metastases generated by LAKP MTOs (Fig. 2h and Extended Data Fig. 8b, c). We also introduced loss of function *Smad4* mutations (S) in LAKP MTOs using CRISPR–Cas9 technology. *In vitro*, LAKPS MTOs were resistant to the cytostatic action of TGFβ (Fig. 2i, j). *In vivo*, LAKPS formed larger numbers of liver tumours that were also larger in size than those formed by parental LAKP MTOs. Galunisertib effectively abrogated the metastatic capacity of LAKPS MTOs and, importantly,

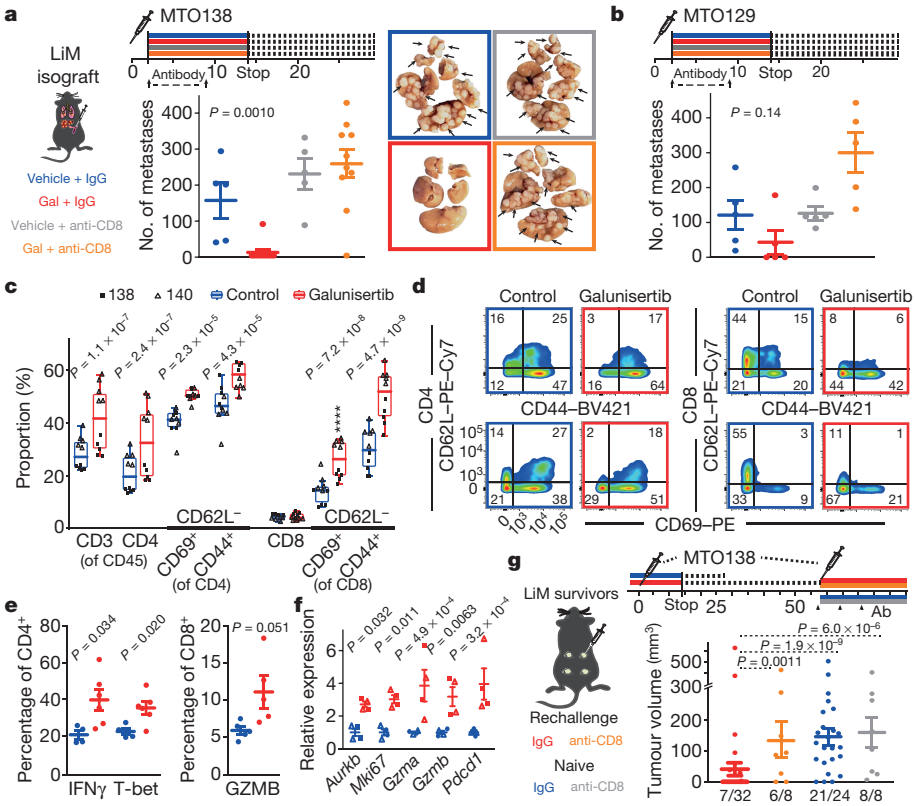


Figure 3 | TGFβ mediates immune evasion. a, **b**, Liver metastases (mean ± s.e.m.), $n = 5$, 10, 5, 10 mice, left to right (**a**), and $n = 5$ mice per condition (**b**). Two-sided Mann–Whitney *U* test. **c**–**e**, Immunophenotyping in whole livers from mice injected with MTO138 or MTO140 ($n = 5$ per condition). Tukey box plots, $n = 5$ mice per condition; analysed with a mixed-effects linear model. **d**, Flow cytometry plots from **c**, representative of two experiments. **e**, Intracellular cytokine expression (mean percentage ± s.e.m.), $n = 4$, 6, 4, 6, 5 and 5 mice, left to right; two-sided Student's *t*-tests. **f**, Relative mRNA expression (mean ± s.e.m.) on sorted CTLs from the experiment in **c**; $n = 4$ mice per condition; two-sided Student's *t*-test. **g**, Rechallenge experiment in liver metastasis survivors compared to naive hosts. Tumour volumes (mean ± s.e.m.) at end point. *x* axis, number of grown tumours/number of injections.

did not enhance growth of metastases generated by LAKP MTOs with wild-type TGFβ pathway (Fig. 2k and Extended Data Fig. 8d).

To study the kinetics of metastasis, we transduced MTOs with luciferase, which did not modify the therapeutic efficacy of galunisertib (Extended Data Fig. 8e). Bioluminescence revealed that galunisertib markedly enhanced cell killing at the onset of the exponential growth phase. However, this did not occur in *nu/nu* mice (Extended Data Fig. 8f–h). These kinetics are suggestive of an adaptive anti-tumour immune response. Indeed, the therapeutic effect of galunisertib was abolished upon depletion of $CD8^+$ cytotoxic T lymphocytes (CTLs) (Fig. 3a, b) or $CD4^+$ T-helper (T_H) cells (Extended Data Fig. 8i). We also transplanted MTOs in mice from the colony used to generate the original compound mutant mice. Galunisertib decreased metastatic burden in these hosts (Extended Data Fig. 8j), indicating that immunological rejection was not due to expression of exogenous antigens in MTOs such as those encoded by *eGFP* and *creERT2* transgenes.

We next assessed the status of the adaptive immune system in a transcriptomic dataset of human MSS CRC samples ($n = 981$) by measuring levels of expression signatures specific for naive or T_H1-activated $CD4^+$ T cells. Microsatellite-instable (MSI) CRCs displayed increased ratios of T_H1-to-naive T-cell genes, consistent with previous observations[23] (Extended Data Fig. 9a). In MSS CRCs, the ratio of T_H1 to naive cells was inversely correlated to the mean expression of *TGFB1*, *TGFB2* and *TGFB3* genes or the CAF-specific gene expression program (Extended Data Fig. 9b–f) and predicted disease relapse (Extended Data Fig. 9d, g). Therefore, abrogated T-cell differentiation, and increased TGFβ and CAF gene expression characterize a substantial subset of patients with MSS CRC and a poor prognosis.

Flow-cytometry analyses on whole livers colonized by MTOs for 7–10 days showed that there was increased recruitment of $CD3^+$ and $CD4^+$ cells after galunisertib treatment (Fig. 3c). This finding was confirmed by immunohistochemistry-based quantifications of liver sections, which also revealed an increased presence of T-bet, the master transcription factor regulating T_H1 cell differentiation (Extended Data Fig. 9h, i). Equivalent observations were made in treated liver metastases from parental MTOs, in experiments with orthotopic implantation, and during genetic abrogation of TGFβ signalling in the TME (Extended Data Fig. 10a–c). Importantly, both T_H cells and CTLs underwent activation, indicated by the increase in $CD44^+CD62L^-$ and $CD69^+CD62L^-$ populations (Fig. 3c, d), increased levels of T-bet and IFNγ effector molecules in T_H cells, and increased GZMB production and expression of proliferation genes in CTLs (Fig. 3e, f). We hypothesized that this robust immune response might immunize mice against the tumour cells. When survivors were rechallenged with the same MTO, most tumours were rejected within two weeks in the continued absence of treatment, whereas the MTOs grew efficiently in naive C57BL/6J recipients. Concurrent depletion of $CD8^+$ CTLs inhibited tumour rejection (Fig. 3g). We conclude that increased TGFβ levels in the TME limit adaptive immune responses by inhibition of the T_H1-effector cell phenotype.

Finally, we assessed the therapeutic action of galunisertib in mice with overt metastatic disease. Galunisertib treatment 14 days after MTO inoculation reduced metastatic burden but resulted in few complete remissions (Fig. 4a). Nevertheless, we observed increased infiltration of $CD4^+$ T cells and of T-bet^+ lymphocytes immediately after initiation of therapy (Extended Data Fig. 10d). Exploring resistance mechanisms to explain this mild response, we discovered that galunisertib-activated T cells exhibited marked surface expression of programmed cell-death protein 1 (PD-1) (Fig. 4b), and that metastases generated by MTOs progressively recruited stromal PD-1 ligand 1 (PD-L1)$^+$ cells—most

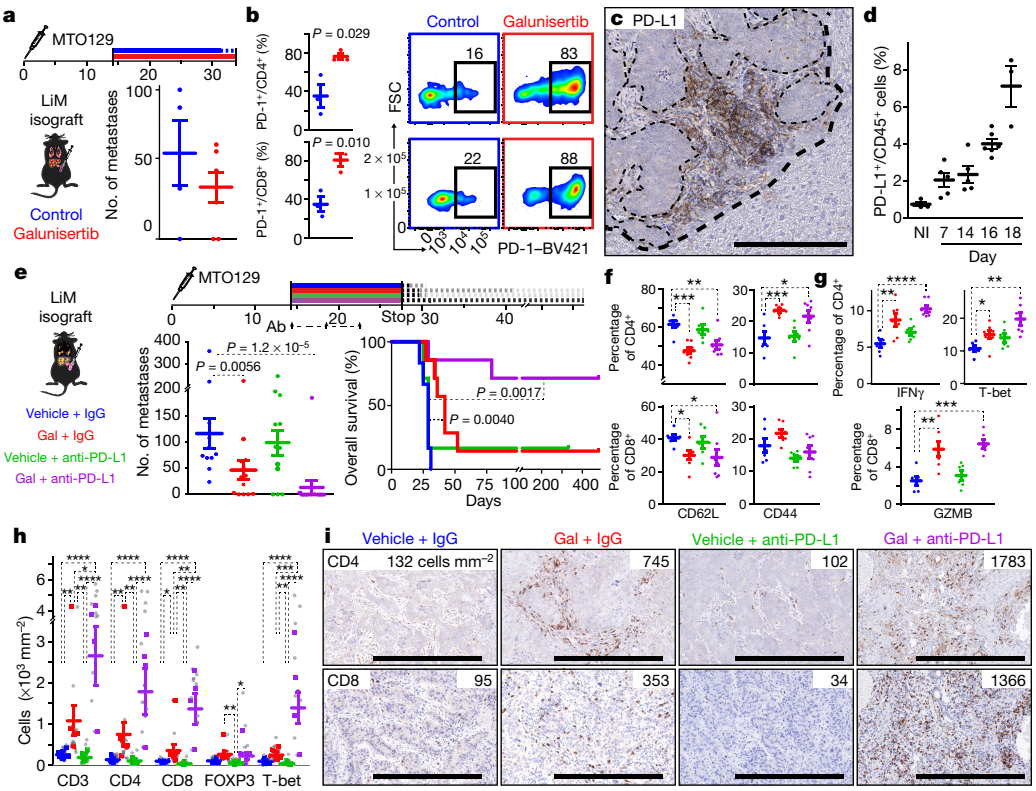


Figure 4 | Dual immunotherapy cures established metastases. **a**, Liver metastases (mean ± s.e.m.) in animals treated with galunisertib from day 14 after injection with MTO129, $n = 4$ and 6 mice, left and right, respectively. **b**, PD-1$^+$ T cells (mean ± s.e.m.) in microdissected liver metastases four days after the start of treatment (day 18), $n = 3$ per condition; two-sided Student's t-test. Right, representative density plots. **c**, PD-L1 immunohistochemistry on MTO129 liver metastases. **d**, PD-L1$^+$ leukocytes in liver metastases (mean ± s.e.m.) at indicated days after injection, or in non-injected (NI) liver, $n = 4$, 5, 5, 6 and 3, left to right. **e**, Treatment from day 14 of established liver metastases. Left, mean number of liver metastases ± s.e.m., $n = 11$, 13, 12 and 14 mice, left to right; two-sided Mann–Whitney U test; right, Kaplan–Meier survival curve; $n = 6$ (vehicle + IgG), 7 (galunisertib + IgG), 6 (vehicle + anti-PD-L1), 7 (galunisertib + anti-PD-L1), Mantel–Cox test. **f**, **g**, Surface activation markers (**f**) and intracellular markers (**g**) in mice euthanized two days after start of treatment (day 16). Mean percentage ± s.e.m., $n = 6$, 7, 6, 7 mice, left to right; two-sided Student's t-test. **h**, Cell densities quantified by immunohistochemistry from mice in **e**. Individual metastases in grey, mouse means in coloured squares; bars show group means ± s.e.m., $n = 5$, 4, 4 and 4 mice, left to right; analysed with a mixed-effects linear model. **i**, Images from the analysis in **h**; mean cell densities are indicated. Scale bars, 500 μm. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

prominently tumour-associated macrophages—as they expanded in size (Fig. 4c, d and Extended Data Fig. 10e, f). Treatment of mice bearing overt metastatic disease with anti-PD-L1 therapy alone had a very modest therapeutic effect (Fig. 4e). However, combined treatment with galunisertib and anti-PD-L1 antibodies induced a pronounced immune response, with increased T-bet and IFNγ levels in CD4$^+$ T_H cells and increased GZMB production in CTLs, which eradicated most metastases and prolonged recurrence-free survival for over a year after cessation of treatment (Fig. 4e–g). This response was associated with a marked, synergistic increase in infiltrating lymphocytes and T-bet expression (Fig. 4h, i), indicating both a disruption of the T-cell-exclusion phenotype characteristic of progressed metastatic disease and prominent T_H1 immune activation.

It has been hypothesized that MSS CRCs are immunologically 'cold', that is, scarcely T-cell infiltrated and possibly non-immunogenic, and that they are therefore unlikely to benefit from immune therapies[24] (Supplementary Discussion). By contrast, our data reveal that this class of CRCs can be killed effectively by the adaptive immune system through a CTL-dependent process, which CRC cells avert by increasing TGFβ levels. Consistent with the well-established role of TGFβ signalling in suppressing differentiation and activity of T cells[25–28], we observed that a TGFβ-activated TME antagonizes the T_H1-effector cell phenotype. We also show that such a TME excludes T cells from tumours, a phenomenon associated with poor outcomes across cancer types[3,29,30]. Enabling immune infiltration using TGFβ inhibitors is sufficient to confer susceptibility to anti-PD-1–PD-L1 checkpoint-based therapies, a strategy that may have broad applications for treatment of cancers that grow in a TGFβ-rich environment. These results strongly suggest that inhibition of TGFβ signalling could be promising as immunotherapy for patients with MSS, stroma-rich CRCs and a poor prognosis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2016; accepted 2 January 2018.
Published online 14 February 2018.

1. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105,** 4283–4288 (2008).
2. Mlecnik, B. *et al.* The tumor microenvironment and immunoscore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8,** 327ra26 (2016).
3. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313,** 1960–1964 (2006).
4. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).
5. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
6. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22,** 571–584 (2012).
7. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21,** 1350–1356 (2015).

8. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767 (1990).
9. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc. Natl Acad. Sci. USA* **114,** E2357–E2364 (2017).
10. O'Rourke, K. P. *et al.* Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nat. Biotechnol.* **35,** 577–582 (2017).
11. de Sousa e Melo, F.*et al.* A distinct role for $Lgr5^+$ stem cells in primary and metastatic colon cancer. *Nature* **543,** 676–680 (2017).
12. Colnot, S. *et al.* Liver-targeted disruption of *Apc* in mice activates β-catenin signaling and leads to hepatocellular carcinomas. *Proc. Natl Acad. Sci. USA* **101,** 17216–17221 (2004).
13. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* **15,** 3243–3248 (2001).
14. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in *p53*-null mutant mice by somatic inactivation of *Rb* in the external granular layer cells of the cerebellum. *Genes Dev.* **14,** 994–1004 (2000).
15. Levéen, P. *et al.* Induced disruption of the transforming growth factor beta type II receptor gene in mice causes a lethal inflammatory disorder that is transplantable. *Blood* **100,** 560–568 (2002).
16. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
17. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47,** 312–319 (2015).
18. van de Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161,** 933–945 (2015).
19. Fujii, M. *et al.* A colorectal tumor organoid library demonstrates progressive loss of niche factor requirements during tumorigenesis. *Cell Stem Cell* **18,** 827–838 (2016).
20. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
21. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
22. Rodón, J. *et al.* First-in-human dose study of the novel transforming growth factor-β receptor I kinase inhibitor LY2157299 monohydrate in patients with advanced cancer and glioma. *Clin. Cancer Res.* **21,** 553–560 (2015).
23. Llosa, N. J. *et al.* The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov.* **5,** 43–51 (2015).
24. Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372,** 2509–2520 (2015).
25. Gorelik, L. & Flavell, R. A. Abrogation of TGFβ signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12,** 171–181 (2000).
26. Marie, J. C., Liggitt, D. & Rudensky, A. Y. Cellular mechanisms of fatal early-onset autoimmunity in mice with the T cell-specific targeting of transforming growth factor-β receptor. *Immunity* **25,** 441–454 (2006).
27. Thomas, D. A. & Massagué, J. TGF-β directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* **8,** 369–380 (2005).
28. Mehal, W. Z., Sheikh, S. Z., Gorelik, L. & Flavell, R. A. TGF-β signaling regulates $CD8^+$ T cell responses to high- and low-affinity TCR interactions. *Int. Immunol.* **17,** 531–538 (2005).
29. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541,** 321–330 (2017).
30. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348,** 74–80 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Sancho for critical reading of this manuscript, all members of the Batlle laboratory for support and discussions, and members of the López-Bigas laboratory for help with revising the manuscript. We are grateful for the assistance of the IRB Barcelona core facilities for histopathology, functional genomics, mouse mutant and advanced digital microscopy; the flow cytometry and animal facilities of the UB/PCB; and the CRG genomic unit. D.V.F.T. held a Juan de la Cierva postdoctoral fellowship from MINECO. This work was supported by grants from the Doctor Josef Steiner Foundation, ERC advanced grant 340176, Instituto de Salud Carlos III, Olga Torres Foundation, BBVA Foundation, grant SAF-2014-53784 (MINECO) and by Fundación Botín. IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO.

Author Contributions D.V.F.T., S.P.-P., D.S. and A.Cal. performed animal husbandry and genotyping; M.S. performed immunohistochemistry; D.V.F.T. and M.I. analysed histopathology. D.V.F.T. generated MTOs, which D.V.F.T. and D.S. characterized *in vitro*; D.S. performed CRISPR experiments; C.S.-O.A. and A.B.-L. performed exome and RNA-seq analyses, other bioinformatics (CMS classifier and patient data), and statistical analyses. S.P.-P., D.V.F.T., J.B.-R., A.Cañ. and X.H.-M. performed mouse isografting; D.V.F.T., D.S. and J.B.-R. quantified immunohistochemistry. D.B., J.A.M. and A.R. synthesized galunisertib. D.V.F.T. coordinated and performed animal treatments and analysed the data. D.V.F.T., J.B.-R., S.I., E.I.R. and A.R.N. performed immunophenotyping experiments. E.B. and D.V.F.T. conceived the study, coordinated experiments and wrote the manuscript. E.B. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.B. (eduard.batlle@irbbarcelona.org).

Reviewer Information *Nature* thanks L. Vermeulen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Ethics and animal maintenance. All experiments with mouse models were approved by the Animal Care and Use Committee of Barcelona Science Park (CEEA-PCB) and the Catalan government. Mice were maintained in a specific-pathogen-free (SPF) facility with a 12-h light–dark cycle and given *ad libitum* access to standard diet and water. All mice were closely monitored by the authors (D.V.F.T. and S.P.-P), facility technicians (during treatments) and by an independent veterinary scientist responsible for animal welfare.

Genetically modified mice. *Lgr5-EGFP-IRES-creERT2* (B6.129P2-*Lgr5*$^{tm1(cre/ERT2)Cle}$/J; stock 008875), *LSL-K-ras*^{G12D} (B6.129S4-*Kras*^{tm4Tyj}/J; stock 008179), *p53*^{LoxP} (B6.129P2-*Trp53*^{tm1Brn}/J; stock 008462) and *Tgfbr2*^{LoxP} (B6.129-*Tgfbr2*^{tm1Karl}/J; stock 012603) mouse strains[13–16] were obtained from The Jackson Laboratory, and *Apc*^{LoxP} mice were obtained from C. Perret[12]. Mice were inbred on a C57BL/6J background and successive crosses were performed to combine alleles. In this study, we have abbreviated the alleles: L (Lgr5$^{eGFP\ creERT2}$), A (*Apc*^{LoxP}), K (*LSL-Kras*^{G12D}), T (*Tgfbr2*^{LoxP}) and P (*p53*^{LoxP}) and generated the following strains: LAK, LATK, LAPK, LATPK. For practical reasons, the strains were maintained as homozygous (*lox/lox*) when possible (A, T, P) and we used both K-heterozygous and wild-type littermates. -*UBC*^{creERT2};*Tgfbr2*$^{fl/fl}$(B6.Cg-Tg(*UBC-creERT2*)1Ejb/2J; stock 008085) and Rosa26^{mTmG} (B6.129(Cg)-Gt(ROSA)26Sortm4 (*ACTB-tdTomato-eGFP*)Luo/J; stock 007676) mice have been described previously[31,32]. The latter allele allowed us to detect recombination using the shift from tdTomato to eGFP expression.

Recombination of genetic models. To induce tumorigenesis from mutated intestinal stem cells (ISC) with preferred localization in the distal part of the intestine, adult mice (at 12–13 weeks of age) were given drinking water (*ad libitum*) containing 2.5–3% (w/v) dextran sodium sulfate (DSS) for five days, and two intraperitoneal injections of diluted (4 mg/kg) tamoxifen (Sigma; dissolved in sunflower oil with 0.5% ethanol) on days 0 and 5. Thirteen per cent of mice treated in this way died in the first two weeks, most likely owing to DSS toxicity. These mice were excluded from the analysis. Animal weight, stool type and overall appearance were scored two times per week. With increasing morbidity, mice were evaluated more frequently until the experimental end point: progressive or rapid weight loss or emaciation and poor physical appearance, characterized by anaemia, hunched posture, ungroomed appearance and lethargy. Mice were then euthanized and dissected; the wall of the peritoneum was inspected and intestines, mesentery, liver, spleen, kidneys, diaphragm and lungs were collected. After macroscopic and/or binocular analysis, tissues were washed in phosphate-buffered saline (PBS), fixed in 10% formalin solution (Sigma) overnight and embedded in paraffin. Survival statistics were analysed using GraphPad Prism software (v.7.03). Median latency of metastasis was obtained by taking the median survival time of the animals in which metastases were detected.

Although the DSS-treatment-mediated induction of inflammation is acute and subsides after 2–3 weeks, we also induced gene recombination with tamoxifen without DSS treatment. Without this acute inflammation, induction still gave rise to tumours equivalent to those described here (including invasive carcinomas). However, tumour burden tended to concentrate to the distal ileum (small intestine), leading to adenomatous carpets, causing serious complications for the mouse, including reduced life span. The benefit of DSS is that it helps to target the colon. The experiments without DSS are not included in the survival and tumour assessment of this study.

Recombination on the *UBC*^{creERT2};*Tgfbr2*$^{fl/fl}$ background was performed with two shots of 80 mg/kg tamoxifen on days two and three. One mouse was excluded from the analysis shown in Fig. 2f because it exhibited no recombination upon tamoxifen treatment.

Histology and immunohistochemistry. Standard haematoxylin and eosin and antibody staining were performed on 4-μm tissue sections using standard procedures, as described previously[4]. Antibodies against CALD1 (rabbit; Sigma HPA008066; 1:250), IGFBP7 (rabbit; Sigma HPA002196; 1:200), pSMAD2 (rabbit; Cell Signaling 3108; 1:50), CD4 (rabbit; Sino Biological 50134-R001; 1:1000), CD8 (rabbit; Biorbyt orb10325-200; 1:200), FoxP3 (rabbit; Abcam ab54501; 1:1000), T-bet (Santa Cruz sc-21003; 1:500), pSMAD3 (rabbit; Abcam ab52903; 1:500) and PD-L1 (Cell Signalling 16764988S; 1:25) were used for staining overnight at 4 °C. Antibodies against GFP (rabbit; Life Technologies A11122; 1:500) and CD3 (rabbit; DAKO IS50330; 1:30) were used for staining for 2 h at room temperature. Images of histology and immunohistochemistry were taken with a Nikon Eclipse E600 and Nikon DS-Ri1 camera or with a Hamamatsu NanoZoomer Digital Slide Scanner (20× magnification).

Histopathological quantifications. Haematoxylin- and eosin-stained sections of intestines were scored blindly for T (tumour) status by an expert pathologist (M.I.) and a second observer (D.V.F.T.) using the AJCC–TNM system: scoring for T_{is} (carcinoma *in situ*), T1 (adenocarcinoma with submucosal invasion), T2 (intramuscular invasion), T3 (transmuscular invasion; serosa intact or cannot be evaluated) or T4 (transserosal invasion; T4b when local metastasis is evident, otherwise T4a). Across all genotypes, 156 tissue samples were scored (both small intestine and large intestine: caecum, colon and rectum) from 68 mice across eight genotypes. Given the high expression of *Lgr5* in the distal ileum and caecum, these tissues were in some cases overgrown with a carpet of dysplastic tissue, making adenoma counting impossible. However, we observed no obvious correlation between the percentage of dysplastic small-intestine surface and genotype. All invasive adenocarcinomas (small and large intestine) and large intestine adenomas were counted individually, reaching a total count of 1,477 tumours.

Mouse tumour organoid generation and culture. Before tissue fixation, tumours were dissected to include potential submucosal invasion and washed in PBS. Dissected material was roughly chopped with razor blades, followed by enzymatic digestion with 200 U/ml collagenase IV in DMEM (Life Technologies) for 20 min at 37 °C. Tissue fragments were subsequently treated by mechanical disruption in DMEM with 10% FBS (Life Technologies), washed with cold PBS and filtered through 100- and 40-μm meshes. Single-cell preparations were stained with propidium iodide (PI, a membrane integrity marker) and GFP^+PI^- cells were sorted in a FACS Aria flow cytometer (BD Biosciences). Typically, around 1,000 cells were obtained and plated in a drop of cold basement-membrane extract (BME) medium (Cultrex BME Type 2, Amsbio): 40 μl in a prewarmed (37 °C) well of a standard 24-well plate (Corning). After 5 min, mouse tumour organoid culture medium (advanced DMEM/F12, supplemented with 10 mM HEPES, Glutamax, B-27 without retinoid acid (all Life Technologies), 50 ng/ml recombinant human EGF (Peprotech), 100 ng/ml recombinant human NOGGIN and 1 μM galunisertib (LY2157299, see below)) was added. NOGGIN was produced in house; expressed as a His-tagged protein in HEK293-6E cells and purified with a 5-ml nickel-affinity column and a 5-ml HisTrap HP column in an ÄKTAxpress at 4 °C. Protein fractions were pooled and desalted (HiPrep 26/10 column), and tested on mouse *Apc*-mutant adenoma organoids or on bone morphogenic protein (BMP)-sensitive patient-derived organoids, analysing *ID1* and *ID3* gene expression by quantitative PCR. In initial passages, MTO medium was supplemented with the antimicrobial reagent Normocin (InvivoGen). MTOs were cultured at 37 °C with 85–90% humidity, atmospheric O_2 and 5% CO_2.

For MTO passaging, basement-membrane extract drops were washed once with HBSS (Lonza) and treated with Trypsin-EDTA (Sigma) for 20 min at 37 °C, followed by mechanical disaggregation of organoid fragments (by pipetting) until a single-cell suspension was obtained. Trypsin was quenched with FBS, cells were washed with HBSS and replated in cold BME medium on warm plates. MTOs were frozen as trypsinized organoids (single cells or small clusters) in DMEM with 50% FBS and 10% DMSO (Sigma). Cultures were checked bimonthly for mycoplasma contamination.

CRISPR genome editing. For gene knockout, guides were designed and cloned into px330-U6-Chimeric BB-CBh-hSpCas9 (Addgene: #42230), which was modified by introducing a SV40 promoter–IRFP expression cassette downstream of hSpCas9. Guide sequences (sgRNA indicated in upper case, restriction overhang in lower case):

Smad4: #1 forward, caccgAGACAGGCATCGTTACTTGT and reverse, aaacACAAGTAACGATGCCTGTCTc; *Smad4*: #2 forward, caccgAGTTTGATGTG TCATAGACA and reverse, aaacTGTCTATGACACATCAAACTc; *mCherryLuc*: forward, caccGCGCATGAACTCCTTGATGA and reverse aaacTCATCAAGG AGTTCATGCGC.

Organoids were nucleofected using a Nucleofector 2b (Lonza) in combination with the Cell Line Nucleofector Kit V (Lonza). Organoids were trypsinized for 15 min at 37 °C to generate single cells ($1.0–1.5 \times 10^6$ cells per guide), which were resuspended in 100 μl nucleofection buffer mix containing 3 μg sgRNA and nucleofected using program A32. Subsequently, cells were plated in BME medium and cultured in full growth medium. For *Smad4* knockout mutants, three days after nucleofection, growth medium was exchanged for selection medium (+ TGFβ1). Selection of *mCherry-Luc* knockout mutants was achieved by FACS sorting of *IRFP-mCherry* double-negative cells twice.

Western blotting. Trypsinized MTO cell pellets were resuspended in lysis buffer (1 mM EDTA, 1 mM EGTA, 1% SDS) containing protease inhibitor cocktail (Sigma). Protein concentration was determined using a standard Bradford assay (BioRad) and 30 μg per sample was separated by SDS–PAGE and transferred to a PVDF membrane (Millipore). Membranes were incubated with antibodies in TBS-Tween (0.2%) containing 5% milk at room temperature for 1 h. Antibodies against SMAD4 (B-8; Santa Cruz sc7966) and actin (Abcam ab20272) were used. The secondary antibody was horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (Pierce 31430).

Immunohistochemical quantifications. Scanned CD3, CD4, CD8, FoxP3, T-bet and pSMAD3 immunohistochemistry stainings were analysed in QuPath (v.0.1.2) using the positive cell detection feature with empirical parameters. Several ROIs (tumours) were taken per section. In cases in which multiple sections per mouse or liver were considered, care was taken to avoid quantifying the same tumour

more than once. Data were processed and visualized with R and RStudio[33] (v.3.4.2 and v.1.1.383, respectively) and the ggplot2[34] package (v.2.2.1) (see Statistics and reproducibility).

Neoantigen prediction. Single-nucleotide variants that were annotated as non-synonymous or nonsense and in-frame insertions or deletions were selected for each sample. Peptides including ten bases upstream and downstream from the alteration were extracted from the canonical protein associated with the corresponding gene and were discarded if they matched another annotated protein. Human CRC: all mutations and predicted neoantigens for human CRC samples were downloaded from https://tcia.at/neoantigens. Samples were classified as MSS or MSI according to the TCGA consortium[5].

For both mouse and human data, neoantigens were predicted using netMHCpan software (v.2.8). See Supplementary Methods for a detailed description.

Mouse injections. For all injections, C57BL/6J (or athymic BALB/c *nu/nu*) mice were purchased from Janvier Labs at six weeks of age and injected at seven weeks. Sex was matched with the origin of the tumour. Intrasplenic or portal vein injections were used for liver colonization by the introduction of dissociated organoids (single cells) into the portal circulation; MTOs were cultured in standard six-well plates for four days and trypsinized as described in 'Mouse tumour organoid generation and culture'. The resulting cell suspension was filtered through 100- and 40-μm meshes (to remove clumps of cells and aggregated debris). Cells were counted and suspended in HBSS for injection, using $2–5 \times 10^5$ cells in 70 μl per mouse. Intrasplenic injections were performed as previously described[4,6,35]. For portal vein injections, a 30*G* syringe was used to inject 100 μl of cell suspension directly into the portal vein. Mice were euthanized at 3–5 weeks, visible liver metastases were counted and data were analysed using GraphPad Prism software (v.7.03). For subcutaneous injections, MTOs were cultured in standard six-well plates for three days and recovered with Cell Recovery Solution (BD Biosciences) for 40 min on ice. Organoids were washed in HBSS and a 10% fraction was trypsinized for cell counting (Neubauer or TC20 automated cell counter, Bio-Rad). Organoids were suspended in HBSS with 30% BME, an equivalent of 1.5×10^5 cells was injected in 2–4 locations on the flanks of each mouse. Growth was scored by palpation and volume was measured with a caliper (multiplying three orthogonal diameters and dividing by two) twice per week. In accordance with our IACUC protocol, mice with tumours over 300 mm^3 in volume were euthanized. Intracaecum injections were used for orthotopic isograft formation: MTOs were recovered as for subcutaneous injection, and an equivalent of 5×10^5 cells in 11 μl was injected per mouse. Intracaecum injections were performed using a 30*G* needle under binocular guidance, adapting the protocol used in ref. 36. Mice were euthanized ten weeks after injection and metastasis was scored macroscopically as well as histologically. Alternatively, orthotopic primary tumours were generated by transplantation of a 1-mm^3 piece of subcutaneously grown MTO-derived tumour (for example, for 3× MTO orthotopic transplantation); the piece was sutured on the tip of the caecum, which was folded over itself to mitigate carcinomatosis. Post-injection, all mice received analgesia (buprenorphine). The caliper technique was used to measure orthotopic tumour sizes after mice were killed. After measurement, pieces were dissected from fresh tumours and lysed in TRIzol (Life Technologies). RNA was purified with the Ambion PureLink kit (Life Technologies).

Mouse treatments. Galunisertib (also known as LY2157299) was synthesized in house and prepared as previously described[6]. Galunisertib or vehicle control was administered by gavage in a 0.15 ml volume, twice per day, starting two days after cell injection for metastasis-initiation experiments, unless otherwise indicated (treatments starting two days after injection give identical results to pre-treating two days before injection). A 800 mg/kg dose per treatment was used for all experiments except for the treatment of MTO138, or for the prolonged treatment of orthotopically implanted mice, in which 400 mg/kg doses were used. For *in vivo* $CD8^+$ T-cell depletion, rat CD8α (YTS 169.4; BioXCell BE0117) or rat IgG2b (LTF-2; BioXCell BE0090) isotype-control antibodies were diluted to 1 mg/ml in PBS and 200 μl was administered per mouse by intraperitoneal injection on days −1, 6 and 13 (unless otherwise indicated) relative to the day of MTO injection. For $CD4^-$ T_H cell depletion, rat anti-CD4α (GK1.5; BioXCell BE0003-1) or rat IgG2b isotype-control antibodies were used. For checkpoint immunotherapy or dual immunotherapy, we used rat PD-L1 (10F.9G2; BioXCell BE0101) or rat IgG2b (LTF-2; BioXCell BE0090) isotype-control antibodies.

***In vivo* study design.** Experimental group sizes were practically associated to cage sizes (five mice per cage) and treatment experiments were designed to have $n \geq 5$ mice per group (one or more cages). No mice were excluded from the analysis, unless explicitly indicated in the Methods. For gavage (galunisertib) treatment, as vehicle control and galunisertib are visually distinguishable, the only randomization we performed was the order of injecting mice: the researcher performing the injections was blinded to the treatment group. End-point criteria are equivalent to those described in 'Recombination of genetic models'. For genetic models, littermates were used as controls and cohorts were accumulated over time for logistical reasons. Mice were excluded if no tumours were observed after a year of survival post-induction.

Statistics and reproducibility. When no statistical control for confounders was needed, numbers of metastases in each condition were compared using a two-sided Mann–Whitney *U* test with exact computation of the null hypothesis, as implemented in the wilcox_test function in the R[33] coin package[37] or in GraphPad Prism. For Fig. 2d, where data were derived from different experimental runs and/or the same treatment performed on multiple MTOs, a mixed-effect linear model was fitted in which experiment run and/or MTO were included as a random effect.

In immunohistochemical experiments and the FACS experiment shown in Fig. 3c, group mean differences were assessed using a linear model. When the data involved more than one MTO (Fig. 3c and Extended Data Figs 9h and 10a) or more than one experiment run (Extended Data Figs 9h and 10c), these variables were included in the models as covariates. When multiple observations per mouse were derived, a mixed-effects linear model was fitted that included the mouse as a random effect (Figs 1m, 4h and Extended Data Figs 9h, 10a, c, d); otherwise, a standard linear regression model (only fixed effects) was used for analysis (Extended Data Fig. 10b).

In all cases, a fourth root was used to transform the outcomes in order to fulfil the assumptions of the linear models. Immunohistochemical quantifications: besides plotting individual tumours (grey), results were represented at the mouse level (coloured or black squares or MTO-dependent symbols) and at group level using the adjusted means and standard error of the means derived by the models in their original scale, after undoing the transformation. To achieve this, standard error of the means were computed by simulation from the corresponding model[38]. Wald tests derived from the linear models were used to perform pairwise comparisons between experimental conditions. A 5% level was chosen for statistical significance. All the analyses were performed using R[33]. Mixed-effects models were fitted using the R packages lme4[39] and lmerTest[40].

For the Kaplan–Meier survival curve after galunisertib treatment (Fig. 2e), all censored mice were metastasis-free when killed. Some mice were euthanized at the end point of control mice (severe morbidity), in order to count nodules at the same time point. Metastasis-burdened mice (≥1 tumour) were scored as death events, whereas disease-free mice were censored from the plot (ticks, zeroes in Source Data).

Tukey box plots in Figs 1e, f, j, 2c 3c and Extended Data Figs 3d, 7a and 9a–c, have whiskers of maximum 1.5 times the interquartile range; the boxes represent first, second (median) and third quartiles. In Fig. 1c, *P* values for comparison of phenotype severity, defined as severe (T3 or higher) versus less severe (<T3), are 0.0002 for LAKTP versus combined 3× genotypes, and <0.0001 for LAKTP versus combined 2× genotypes and for LAKTP versus LA (Fisher's exact test). We also tested for the presence (≥T1) or absence of carcinomas and the *P* values for these comparisons are: 0.0013 for LAKTP versus 3×, 0.0005 for LAKTP versus 2×, and <0.0001 for LAKTP versus LA (Fisher's exact test). The immunohistochemistry images in Fig. 1e, f are representative of four independent experiments. In Fig. 1i, numbers of patients are indicated; MTOs *in vitro*: 3×, $n = 5$ individual MTOs; and 4×, $n = 3$; plotted values are means of biological duplicates. MTOs orthotopically transplanted: 3×, $n = 2$ individual MTOs (MTO54 and 207); 4×, $n = 3$ individual MTOs (MTO68, 93 and 140); plotted values are biological duplicates ($2n$). The experiment in Fig. 1l, m was performed once. Besides providing individual *P* values, this model was used for the overall assessment of galunisertib treatment effect: control mean (95% confidence interval), 24.0 (10.5–47.7) versus galunisertib: 0.084 (0.00–0.81); $P = 1.2 \times 10^{-12}$. Data are from ten independent experiments. In Fig. 2e, data are from seven independent experiments, using either MTO129 or MTO138. In Fig. 2h, data are from MTO54 and MTO220, combined. Figure 2i, j, using MTO220, is representative of two independent experiments. Western blot source data are shown in Supplementary Fig. 1. In Fig. 2k, right, the values are fraction of total ± s.e.m. Data are from experiment with empty guide or from guides #1, #2 LAKP and S combined, from two independent experiments. The experiment in Fig. 3a was performed with MTO138, and the experiment in Fig. 3b was performed with MTO129. In Fig. 3g, *P* values refer to the comparison between: anti-CD8 and IgG in the survivors ($P = 0.0043$); rechallenge IgG and naive IgG, and between rechallenge IgG and naive anti-CD8 ($P < 0.0001$); two-sided Mann–Whitney *U* test. Figure 4b, c are representative of two experiments. In Fig. 4d, MTO129 is injected. In Fig. 4f, *P* values are for CD4 and CD62L: 0.0002 (control versus galunisertib) and 0.0051 (control versus double); CD44: 0.0008 (control versus galunisertib) and 0.0297 (control versus double); CD8 and CD62L: 0.0159 (control versus galunisertib) and 0.0181 (control versus double). In Fig. 4g, *P* values are for CD4 and IFNγ: 0.0099 (control versus galunisertib) and <0.0001 (control versus double); T-bet: 0.0208 (control versus galunisertib) and 0.0017 (control versus double); CD8 and GZMB: 0.0054 (control versus galunisertib) and 0.0001 (control versus double). For Fig. 4h, all *P* values are listed in Source Data. The experiment in Fig. 4i was performed once.

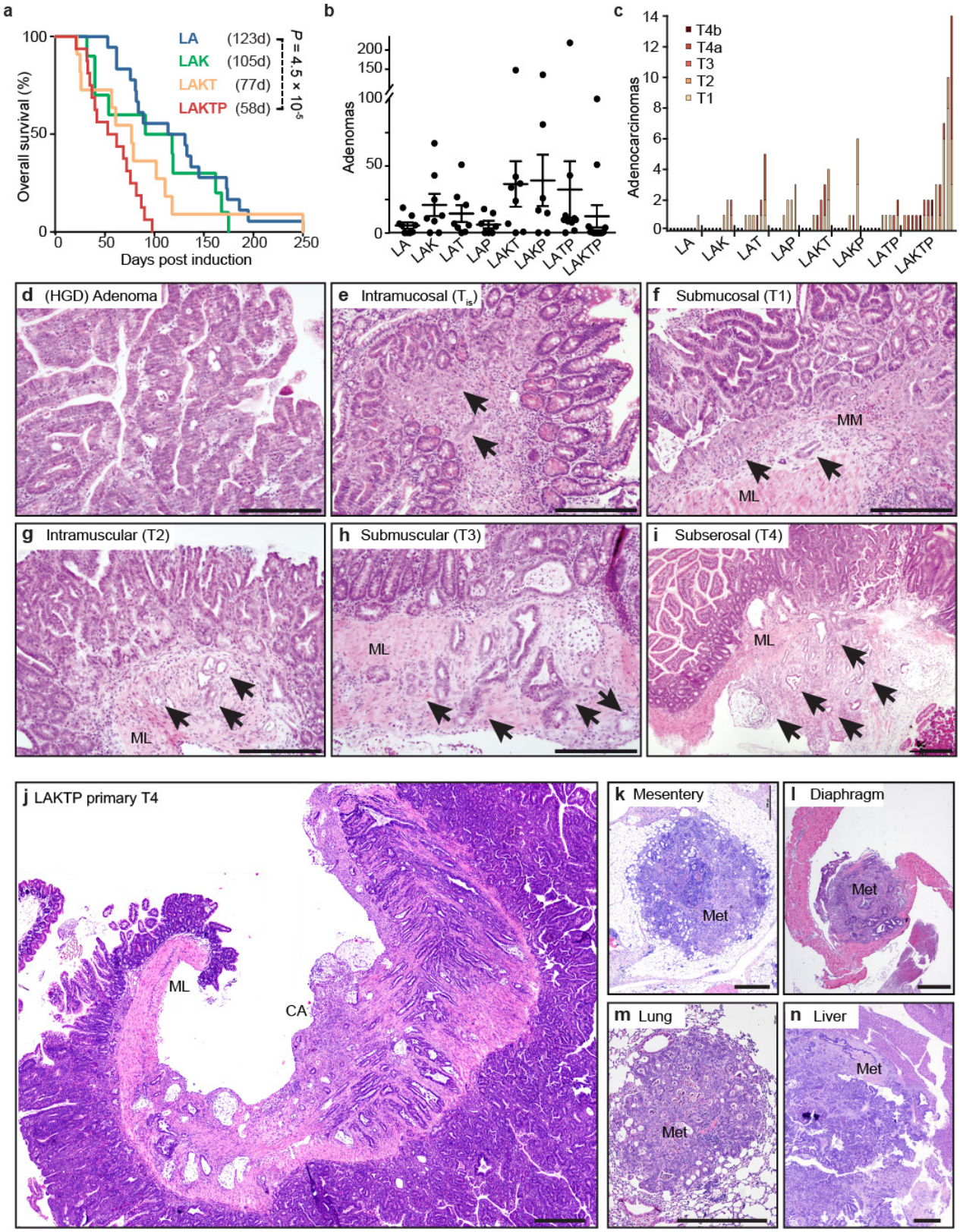
Patient data analysis. We used five publicly available Affymetrix microarray datasets from the NCBI GEO repository, which include gene expression and

clinical information from a total of 1,194 patients with CRC. In order to obtain CD4-naive and $CD4^+$ T_H1 profiles, we carried out a differential expression analysis on gene expression dataset GSE22886[41]. We defined the CD4 naive signature to include genes that are upregulated at least threefold in both CD4-naive versus T_H1 and in CD4-naive versus Th2; Benjamini–Hochberg false discovery rate[42] (FDR) $< 5\%$ in both comparisons. Similarly, the CD4-T_H1 signature included genes that were upregulated at least threefold compared to CD4-naive, and limma FDR $< 5\%$. Association with metastasis was evaluated using a frailty Cox proportional hazards model. Statistical significance was assessed by means of a log-likelihood ratio test, and Wald tests were used for pairwise comparisons when necessary. Association of expression intensities was evaluated as continuous variables assuming a linear relationship with the logarithm of the relative risk. Sample groups of low, medium and high expression levels were defined using the tertiles of the intensity distribution after correction for technical effects. For visualization purposes, Kaplan–Meier curves were estimated for groups of tumours showing low, medium and high gene or signature expression. Only samples from patients diagnosed in stages I, II and III were taken into consideration for these analyses. The threshold for statistical significance was set at 5%. All analyses were carried out using R[33,43]. See Supplementary Methods for a detailed description.

Classification of mouse tumours according to consensus molecular subtypes. We used the *k*-top scoring pairs (*k*-TSP) algorithm and trained the classifier for a binary outcome (CMS4 versus not-CMS4) in 960 samples. See Supplementary Methods and Extended Data Fig. 5 for a detailed description.

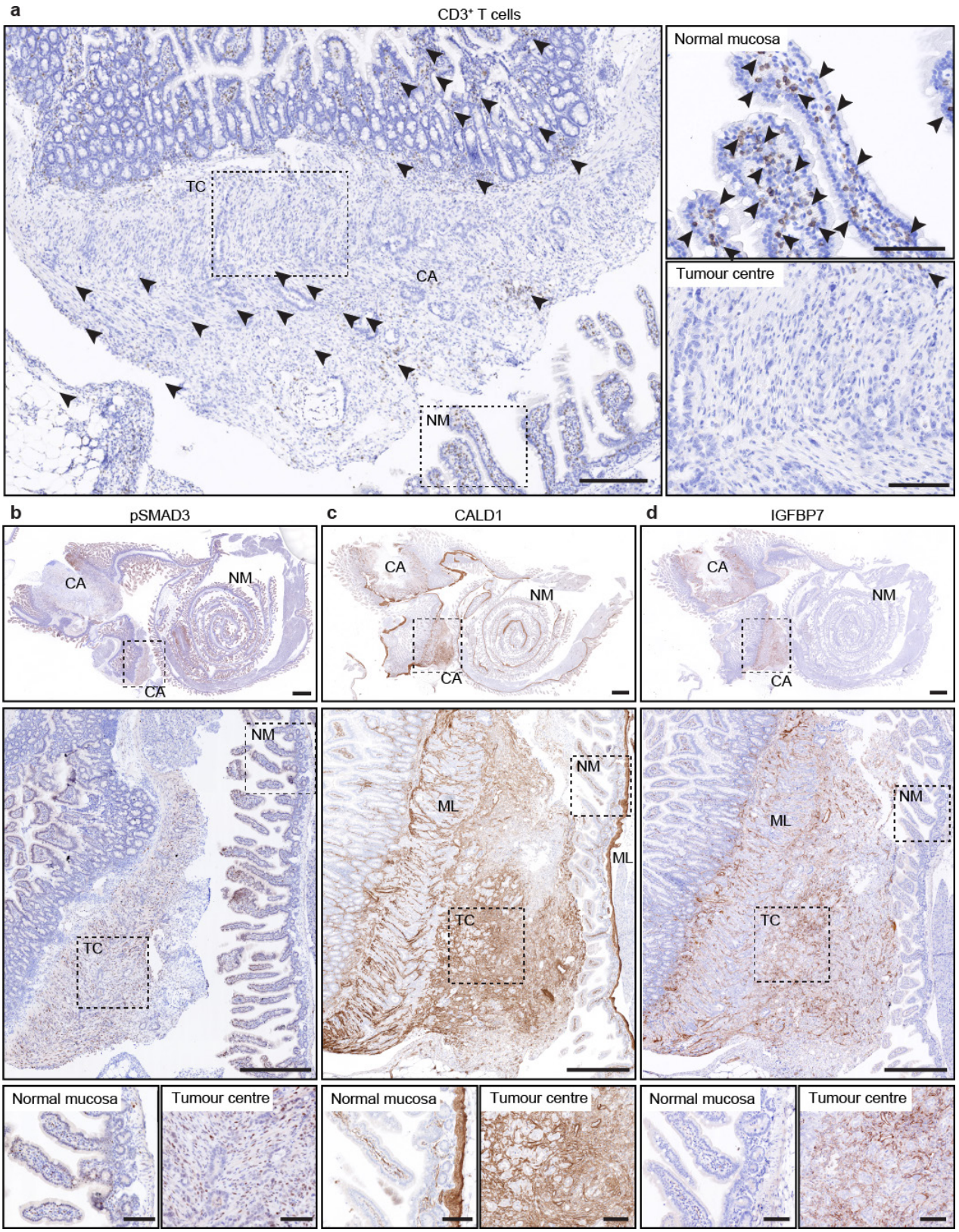
Data availability. MTO whole-exome sequencing and RNA sequencing analysis (RNA-seq) data have been deposited in the ENA archive with accession number PRJEB22559. Expression array data are available at GEO with accession number GSE103562. Computer code is available upon request. Source Data are available in the online version of the paper.

31. Ruzankina, Y. *et al.* Deletion of the developmentally essential gene *Atr* in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1,** 113–126 (2007).
32. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45,** 593–605 (2007).
33. R Core Team. *R: a Language and Environment for Statistical Computing* http://www.r-project.org (R Foundation for Statistical Computing, 2017).
34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
35. Warren, R. S., Yuan, H., Matli, M. R., Gillett, N. A. & Ferrara, N. Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. *J. Clin. Invest.* **95,** 1789–1797 (1995).
36. Céspedes, M. V. *et al.* Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *Am. J. Pathol.* **170,** 1077–1085 (2007).
37. Hothorn, T., Hornik, K., Van der Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60,** 257–263 (2006).
38. Gelman, A. & Hill, J. Data analysis using regression and multilevel/hierarchical models. *J. Stat. Softw.* **30,** 1–5 (2009).
39. Bates, D., Maechler, M., Bolker, B. M. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).
40. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in linear mixed effects models *J. Stat. Softw.* **82,** 1–26 (2015).
41. Abbas, A. R. *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6,** 319–331 (2005).
42. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300(1995).
43. Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S. (eds) *Bioinformatics and Computational Biology Solutions using R and Bioconductor.* https://doi.org/10.1007/0-387-29362-0 (Springer, 2005).

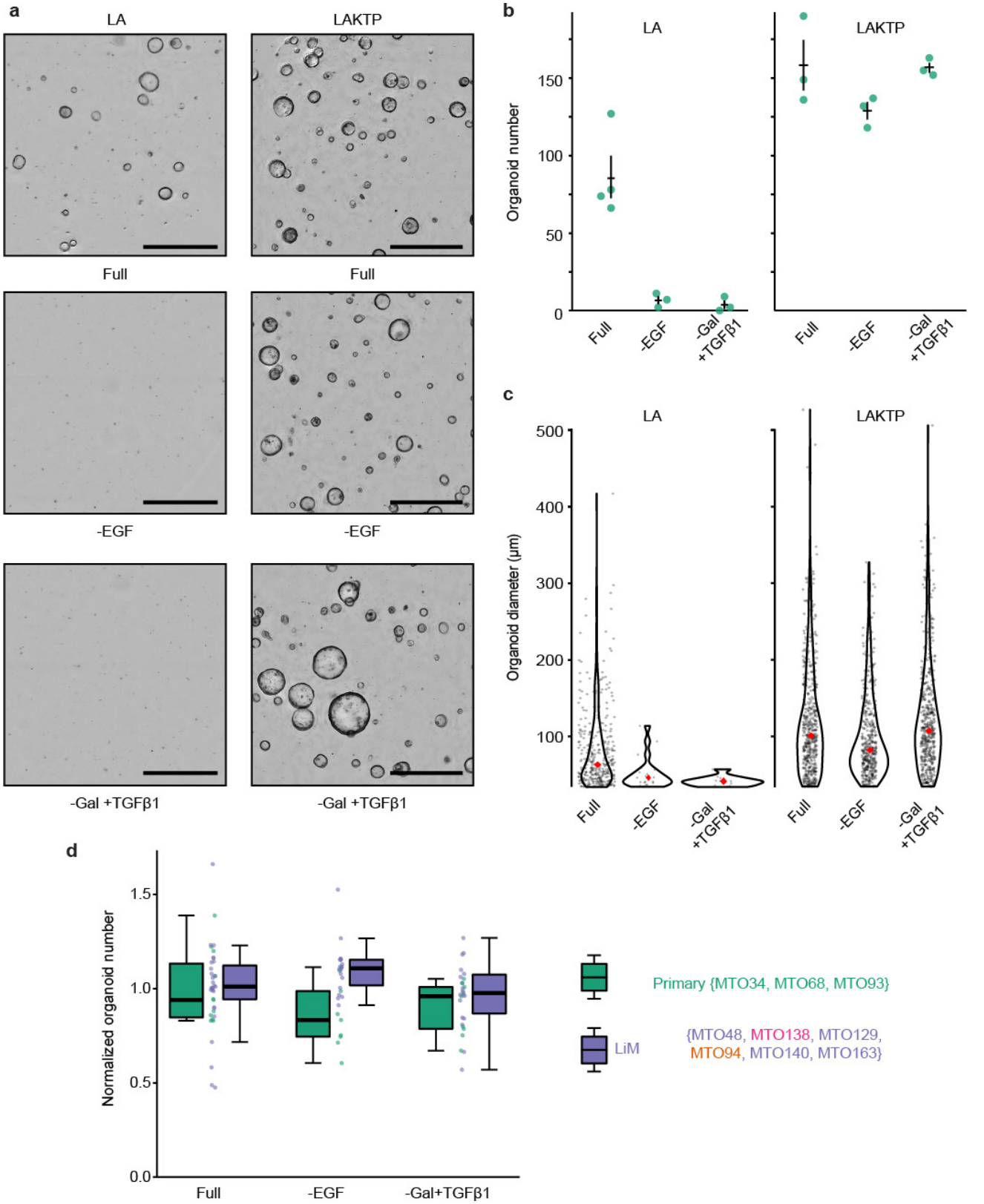


Extended Data Figure 1 | See next page for caption.

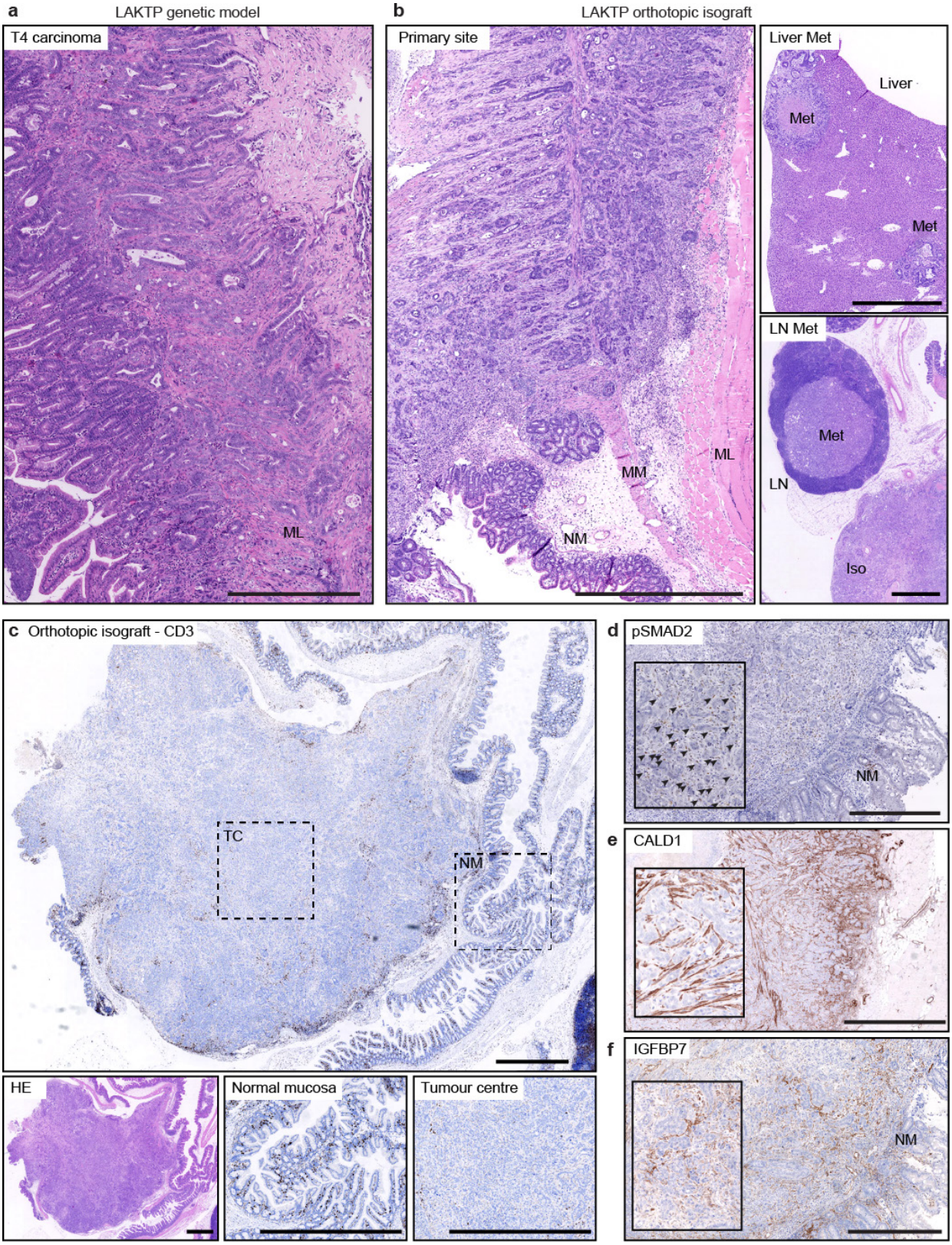
Extended Data Figure 1 | Additional description of the genetic mouse models. a, Kaplan–Meier survival curves for LA ($n = 18$ mice), LAK ($n = 10$), LAKT ($n = 11$) and LAKTP ($n = 16$) mice, as days post-tamoxifen induction of recombination in intestinal stem cells. Median survival is indicated in parentheses; Mantel–Cox test. **b**, Number of tumours (adenomas and carcinomas) in the large intestine, by genotype. Numbers for individual mice ($n = 9$ (LA), 8 (LAK, LAT and LAKT), 7 (LAP and LAKP), 10 (LATP), and 13 (LAKTP)) are indicated, as well as mean ± s.e.m. **c**, Number and grade of adenocarcinomas per individual mouse (bars), by genotype. *P* value for pairwise comparisons of tumour number: LA versus 2×, 0.0198; LA versus 3×, 0.0337; LA versus LAKTP, 0.0001; 2× versus LAKTP, 0.0051; and 3× versus LAKTP, 0.0014; $n = 9$ mice (LA), 8 (LAK, LAT, LAKT), 7 (LAP, LAKP), 10 (LATP), 13 (LAKTP); two-sided Mann–Whitney *U* test. **d**–**i**, Example images of tumour types scored by the TNM criteria, representative of 23 independent experiments. Arrows indicate invasive cells or glands, absent in an adenoma (**d**), restricted to the mucosa in an *in situ* tumour (Tis, **e**), breaching the muscularis mucosae (MM) in a submucosal invasive tumour (T1, **f**), penetrating the muscle layer (ML) in an intramuscular invasive tumour (T2, **g**), reaching the serosa layer beyond the ML in a submuscular invasive tumour (T3, **h**), or infringing all layers including the serosa in a subserosal or T4 tumour (**i**). **j**, Representative micrograph of an haematoxylin- and eosin-stained invasive subserosal adenocarcinoma (T4) of an LAKTP mouse. Carcinoma (CA) glands are indicated, invading through the muscle layer (ML). **k**–**n**, Examples of spontaneous metastases (Met) in the mesentery, diaphragm, liver and lung of induced LAKTP mice, each image is representative of two experiments. Scale bars, 500 μm.



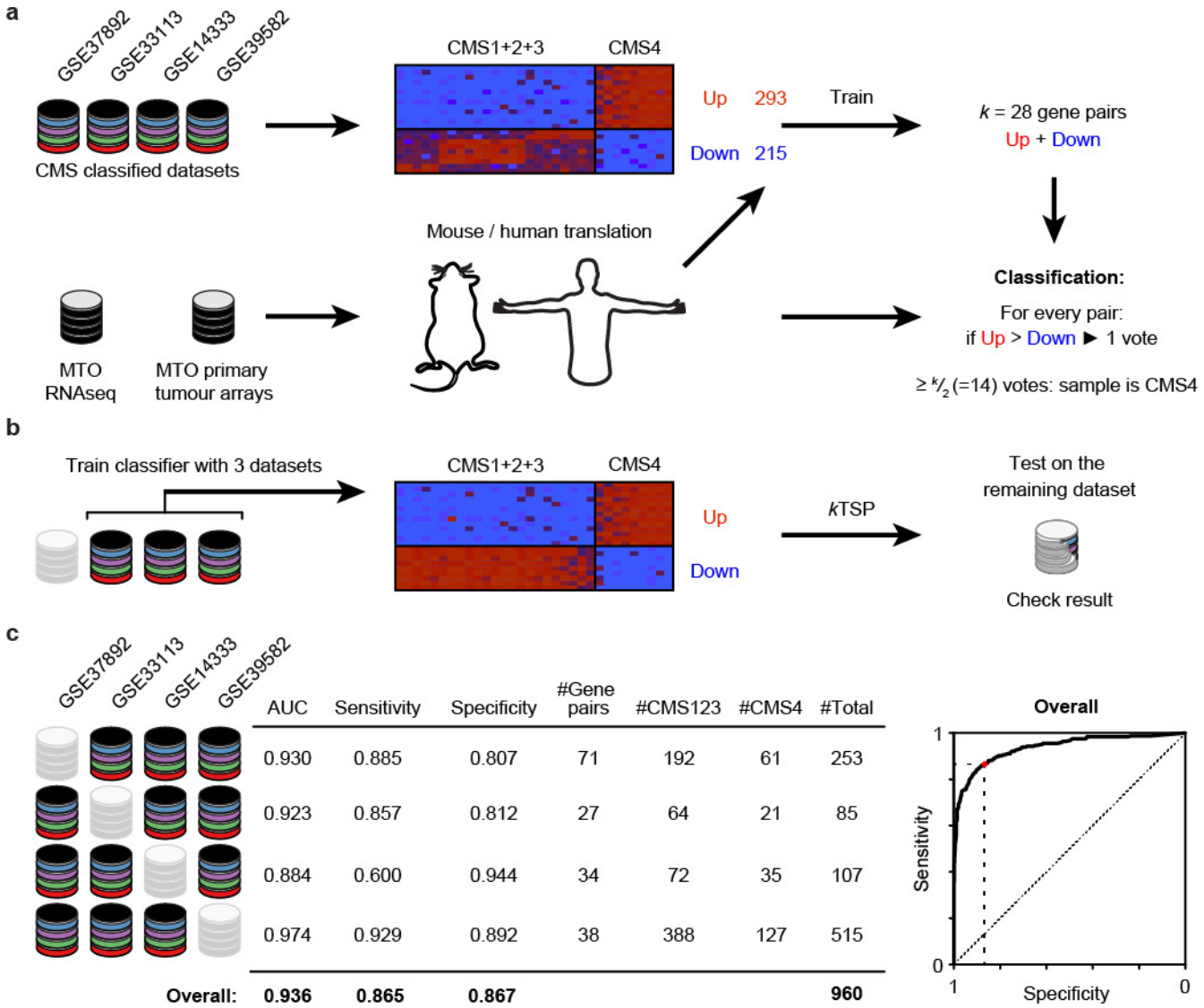
Extended Data Figure 2 | Analysis of the TME of LAKTP adenocarcinomas in the genetic model. a, Immune infiltration in LAKTP carcinomas. Immunohistochemistry for CD3, indicating infiltrating T cells, in an intestinal adenocarcinoma of an LAKTP mouse. Insets show a magnified view of normal mucosa and tumour centre (TC). Arrowheads indicate T cells. **b**–**d**, Markers of TGFβ activation and poor prognosis in LAKTP carcinomas in a representative section of intestine with two aggressive invasive adenocarcinomas. **b**, Staining for pSMAD3. Note nuclear staining in stromal cells of the tumour centre. **c**, Staining for CALD1. Note that it is expressed in the muscle layer throughout the intestine, and in a subset of cells in the stroma of the normal mucosa. In the tumour centre of invasive carcinomas, the staining intensity is clearly higher, especially in fibroblast-like cells. **d**, Staining for IGFBP7, which is barely detectable in the normal mucosa. Expression is strong in the TME of the tumour centre. Images are representative of four independent experiments Scale bars, 1 mm (**b**–**d**, top); 500 μm (**a**, left; **b**–**d,** middle); 100 μm (all other panels).



Extended Data Figure 3 | Niche factor analysis of MTOs *in vitro*. Organoid formation assay in different media with representative MTOs from LA and LAKTP mice. **a**, Images obtained by automated plate scanning, *z*-stack projection and stitching of the niche factor requirement assay. Shown are full medium, medium without EGF and medium without galunisertib (Gal) but with TGFβ1. **b**, Results from automated organoid detection for the MTOs shown in **a**, each data point shows the number of organoids in three replicate wells for representative analyses. Data are shown as mean ± s.e.m. **c**, Organoid diameter was calculated from detected pixel area. Data points are individual organoids, pooled from triplicate wells. Violin plots and medians are shown. **d**, Summary of all data obtained from different LAKTP MTOs analysed in triplicates, from both primary tumours and liver metastases. Data (individual values and Tukey box plots) are normalized to the mean of full medium. Scale bars, 1 mm.

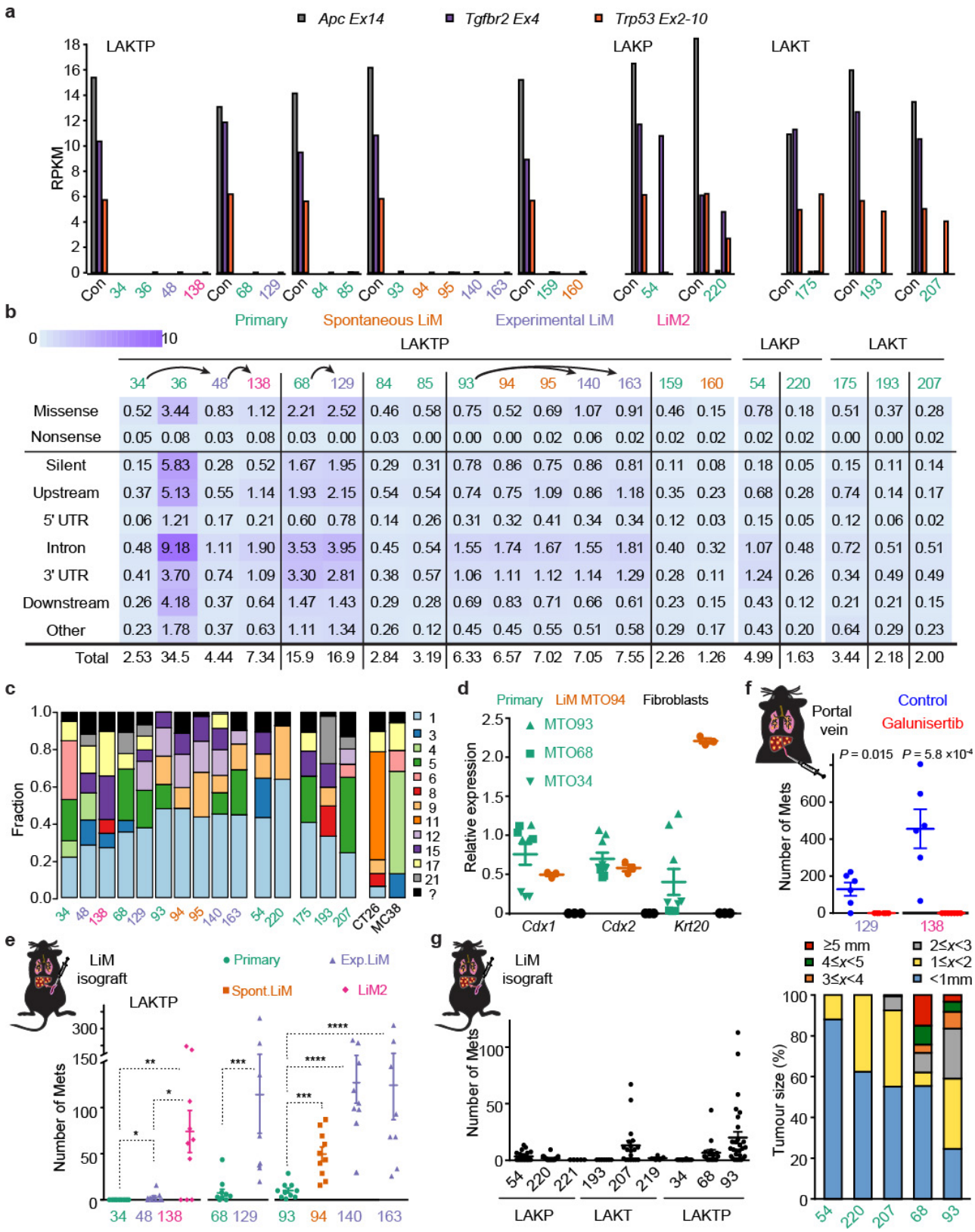


Extended Data Figure 4 | Histology and TME of primary tumours from orthotopically isografted LAKTP MTOs. a, **b**, Haematoxylin and eosin staining of a LAKTP-T4 carcinoma in the genetic model (**a**) and of a recapitulated adenocarcinoma developed from orthotopically isografted LAKTP MTO93 (**b**). Liver and lymph node (LN) metastasis observed in orthotopic isografted mouse (**b**, right panels). Iso, primary isograft. **c**–**f**, Markers of poor prognosis in isografted primary tumours. **c**, Immunohistochemistry for the T-cell marker CD3 (**c**, main panel and right, bottom right two panels) and haematoxylin and eosin stain (**c**, bottom left) of an orthotopic isograft, with typical glandular differentiation. Note that there are few infiltrating T cells in the tumour centre (**c**, bottom right), compared to the periphery and the normal mucosa (**c**, middle right). **d**–**f**, An orthotopically isografted, invasive primary tumour, stained for pSMAD2 (**d**, arrows indicate positive stromal cells), CALD1 (**e**) and IGFBP7 (**f**). Images represent three independent experiments. Scale bars, 1 mm (**c**, main and bottom left); 500 μm (all other panels).



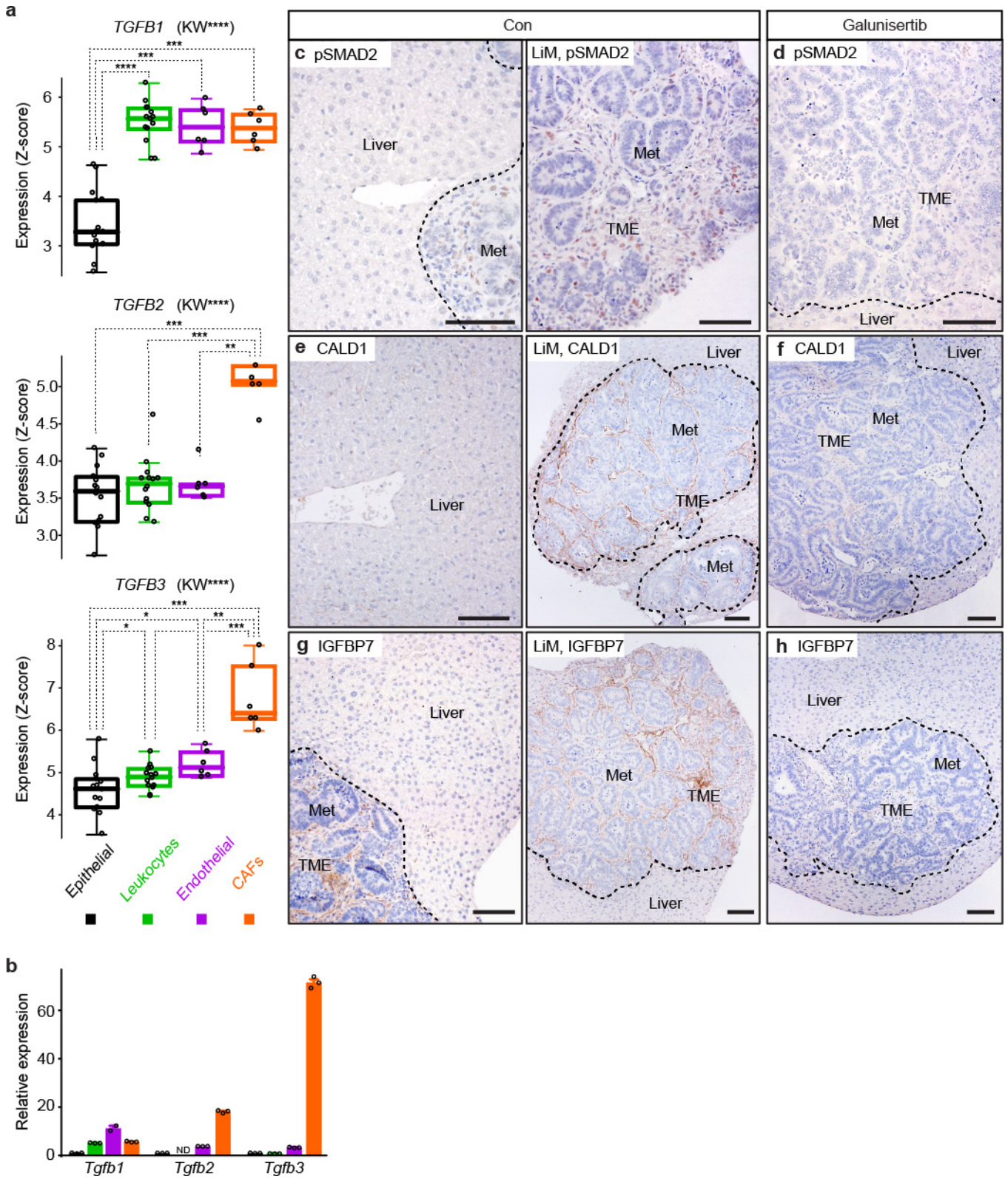
	AUC	Sensitivity	Specificity	#Gene pairs	#CMS123	#CMS4	#Total
	0.930	0.885	0.807	71	192	61	253
	0.923	0.857	0.812	27	64	21	85
	0.884	0.600	0.944	34	72	35	107
	0.974	0.929	0.892	38	388	127	515
Overall:	**0.936**	**0.865**	**0.867**				**960**

Extended Data Figure 5 | Schematic of the CSM4 classifier, using the *k*-TSP algorithm. a, A metacohort of four pooled human CRC datasets with CMS annotation is used to find upregulated and downregulated genes in the CMS4 subtype (434 with >1.5-fold change and 300 with <0.8-fold change). To ensure usability in mouse samples, candidate genes are filtered by high-confidence homology between the two species. The *k*-TSP algorithm selects from among the up- and down-regulated genes the optimal set of pairs that correctly classifies CMS4 samples. The resulting classifier is applied to mouse samples by comparing the expression of pairs of genes. For each sample, the score is increased by one unit if the expression of the 'Up' gene is greater than that of the 'Down' gene. The maximum score for CMS4 classification is 14. **b**, Cross-validation by a leave-one-out approach. To assess performance, we repeatedly trained a classifier on three out of the four human datasets and tested using the remaining dataset. **c**, Results of the leave-one-out analysis. Area under the curve (AUC) was computed after re-scaling the classifier scores to the range between zero and one. Right, receiver operating characteristic curve.



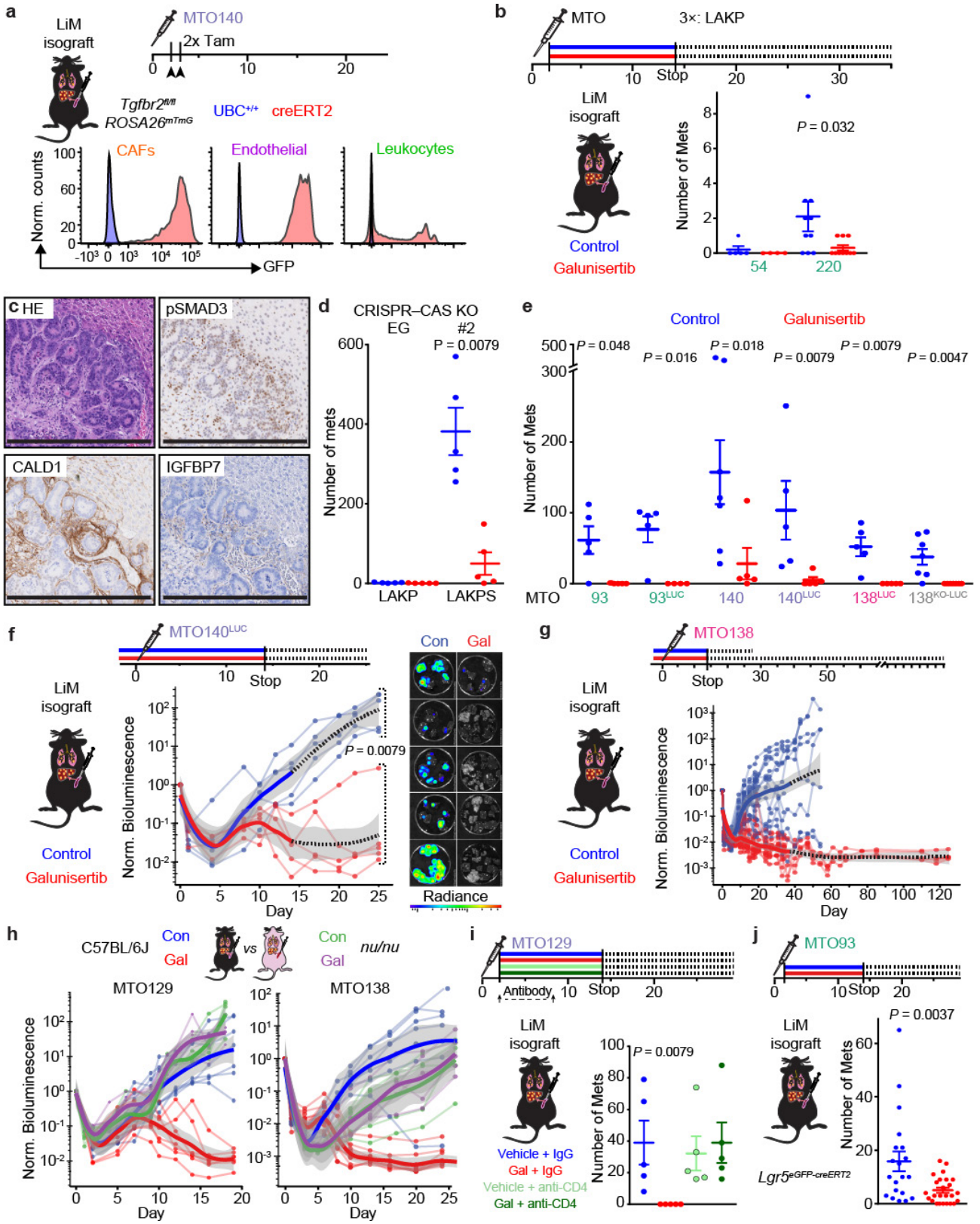
Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Genetic and functional analyses of MTOs. **a**, Whole-exome sequencing of MTOs. Number of reads for the recombined exons in reads per kilobase per million mapped reads (RPKM) in MTOs and their matched host tail DNA (Con). Mouse genotypes are indicated. **b**, Overview of the different LAKTP MTOs analysed, including experimental metastatic descendance (indicated by arrows), and heat map and table of the number of indicated type of somatic mutation per Mb of sequenced exome. **c**, Mutational signatures[21] in the MTOs and mouse monolayer cell lines CT26 and MC38. As in human MSS CRC, mutations in MTOs (but not in the cell lines) are predominantly of type 1. **d**, mRNA expression of intestinal epithelial genes *Cdx1*, *Cdx2* and *Krt20* in cultured MTOs derived from primary tumours MTO34, 68 and 93, and spontaneous liver metastasis MTO94, compared to mouse liver fibroblasts. Values are relative to $2^{-\Delta C_t}$ for *Ppia*, normalized to the value for MTO93; and are shown as three technical replicates with mean ± s.e.m. **e**, Number of liver nodules after intrasplenic injection of primary LAKTP tumour and spontaneous or experimental liver metastasis (LiM)-derived MTOs in syngeneic C57BL/6J mice. Data points represent individual mice; data are mean ± s.e.m. *P* values (95% confidence interval) are MTO34 ($n = 10$ mice) versus MTO48 ($n = 9$): 0.0325 (−2–0), MTO48 versus MTO138 ($n = 10$): 0.0305 (−108–0); MTO34 versus MTO138: 0.0031 (−108–0); MTO68 ($n = 11$) versus MTO129 ($n = 8$): 3.2×10^{-4} (−207 to −30); MTO93 ($n = 10$) versus MTO94 ($n = 10$): 1.1×10^{-4} (−60 to −19); MTO93 versus MTO140 ($n = 9$): 2.2×10^{-5} (−149 to −81); MTO93 versus MTO163 ($n = 8$): 9.1×10^{-5} (−220 to −31); two-sided Mann–Whitney *U* test. **f**, Number of liver nodules (mean ± s.e.m.) after inoculation of disaggregated MTO129 or MTO138 directly into the portal vein in syngeneic animals. Data are from individual mice, $n = 6$ per condition. 95% confidence interval: MTO129, (−223–0) and MTO138 (−805 to −66); two-sided Mann–Whitney *U* test. **g**, Number of liver nodules (mean ± s.e.m.) of intrasplenically injected 3× and 4× MTOs, $n = 22$ mice (MTO54), 20 (MTO220), 5 (MTO221), 7 (MTO193), 19 (MTO207), 5 (MTO219), 14 (MTO34), 16 (MTO68) and 29 (MTO93). Right, distribution of tumour diameters.



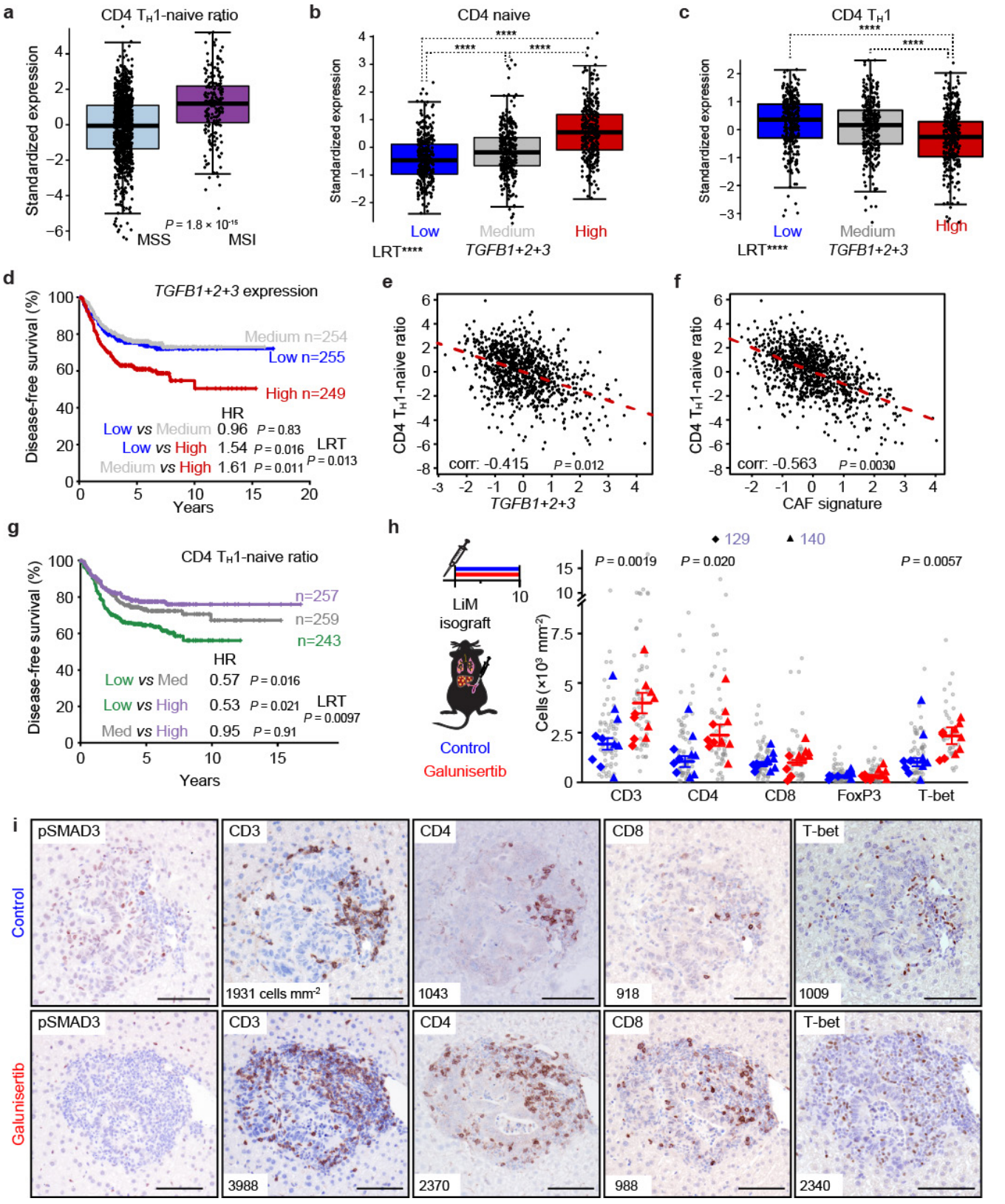
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Source of tumoral TGFβ and pathway blockade *in vivo*. **a**, **b**, *TGFB1*–*TGFB3* (**a**) *or Tgfb1*–*Tgfb3* (**b**) mRNA expression levels in sorted cell populations from disaggregated human CRCs (**a**) or mouse liver metastasis (**b**). First, epithelial cancer cells, immune cells, endothelial cells and CAFs were sorted with labelled antibodies against EPCAM, CD45 ($n = 14$ patients for each), CD31 and FAP ($n = 6$ for each) for human and PDGFRB for mouse. Subsequently, RNA from sorted cells was extracted and analysed by microarray (human) or reverse transcription–quantitative PCR (RT–qPCR) (mouse). **a,** Standardized expression data are shown as *z* score. Individual values and Tukey box plots; *P* values are for *TGFB1*: epithelial versus leukocytes, 7.5×10^{-6}; epithelial versus endothelial, 6.2×10^{-4}; epithelial versus CAFs, 6.2×10^{-4}; and the Kruskal–Wallis (KW) *P* value is 6.3×10^{-6}. For *TGFB2*: CAFs versus epithelial, 6.2×10^{-4}; CAFs versus leukocytes, 8.4×10^{-4}; CAFs versus endothelial, 0.0051; the Kruskal–Wallis *P* value is 0.0017. For *TGFB3*: epithelial versus CAFs, 6.2×10^{-4}; epithelial versus leukocyte, 0.0409; epithelial versus endothelial, 0.0150; leukocyte versus CAFs, 6.2×10^{-4}; endothelial versus CAFs, 0.0051; the Kruskal–Wallis *P* value is 8.5×10^{-5}. Direct comparisons using two-sided Mann–Whitney *U* test. Data are shown as mRNA expression levels normalized to epithelial cells, mean ± s.e.m. from technical triplicates; *Tgfb2* was undetectable (ND) in leukocytes. Note that in both settings, *TGFB1* is expressed by all stromal cell types, *TGFB2* and *TGFB3* are mainly expressed by CAFs. Expression in epithelial cancer cells is comparatively low. **c**–**h**, Immunohistochemistry for TGFβ target gene products in the TME of MTO138-derived liver metastases, representing two independent experiments (**c**, **e**, **g**), as well as for the inhibition of stromal TGFβ signalling by galunisertib (**d**, **f**, **h**). **c**, **d**, Micrographs of sections of liver metastases stained for pSMAD2. **e**, **f**, Staining reveals expression of the TGFβ target CALD1 exclusively in the metastatic TME. Treatment with galunisertib for three days abrogates expression of this cytoplasmic protein. **g**, **h**, Staining for the TGFβ target IGFBP7, which is exclusively expressed in the metastatic TME. Treatment with galunisertib for three days abrogates expression of this secreted protein. Scale bars, 100 μm.



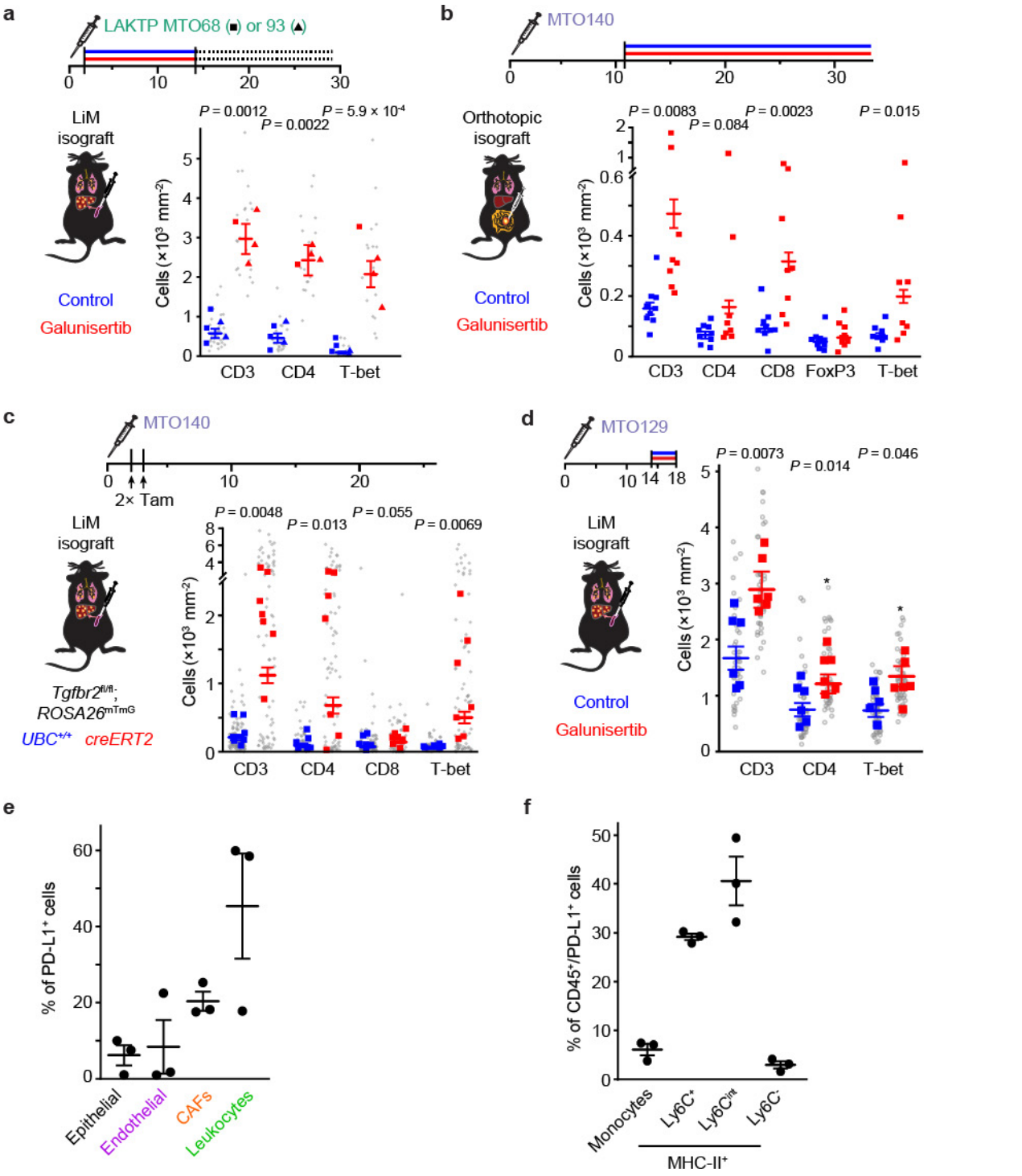
Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Characterization of MTOs in metastatic experiments. a, Representative examples of GFP positivity measured by flow cytometry, indicating genetic recombination in the $UBC^{creERT2}$;$Tgfbr2^{fl/fl}$ genetic model (representative of three independent experiments). **b**, Liver metastases counted five weeks after intrasplenic injection of LAKP MTO54 or MTO220 and treatment from day 2 to day 14 with galunisertib (red; $n = 4$ mice for MTO54 and $n = 10$ for MTO220) or control (blue; $n = 5$ mice for MTO54 and $n = 10$ for MTO220), mean ± s.e.m., 95% confidence interval for MTO220 is (−3–0), from a two-sided Mann–Whitney *U* test. **c**, Stromal TGFβ activation markers in LAKP MTO liver tumours, both supporting the classification of these tumours as CMS4 (Fig. 1i) and explaining the efficacy of galunisertib in **b**; representative of two independent experiments. **d**, Number of nodules five weeks after injection of LAKP empty guide or LAKPS (LAKP + S) MTOs and treatment with galunisertib (red) or control (blue), mean ± s.e.m., 95% confidence interval for LAKPS is (−554 to −176), $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. **e**, Number of liver nodules (mean ± s.e.m.) four weeks after intrasplenic injection of MTOs (either unlabelled or with mCherry–luciferase vector, or after CRISPR knockout of the mCherry–luciferase coding region for MTO138), after treatment. 95% confidence intervals are: for MTO93, (−113−0), $n = 5$ mice per condition; MTO93-LUC: (−104 to −4), $n = 5$ control and $n = 4$ galunisertib; MTO140, (−297 to −21), $n = 7$ control and $n = 5$ galunisertib; MTO140-LUC, (−251 to −21), $n = 5$ per condition; MTO138-LUC, (−86 to −8), $n = 5$ per condition; MTO138-KO-LUC, (−70 to −13), $n = 7$ per condition; two-sided Mann–Whitney *U* test. **f**, Longitudinal intravital bioluminescence imaging (BLI) quantification (photons s^{-1}, normalized to day 0) of an intrasplenic metastasis initiation experiment with MTO140 in C57BL/6J mice treated with galunisertib or control until day 14, $n = 5$ mice per condition. Points and lines represent individual mice, trend lines (bold) show a LOESS model with 95% confidence interval (grey band). 95% confidence interval for the difference at day 25: (−222.7 to −23.9), two-sided Mann–Whitney *U* test. **g**, BLI (as in **f**) of a metastasis-initiation experiment with MTO138 in C57BL/6J mice, $n = 24$ control mice, $n = 15$ galunisertib. **h**, BLI (as in **f**) of galunisertib treatment in intrasplenic liver colonization assays with MTO129 or MTO138, in C57BL/6J ($n = 7$ mice per condition, except for MTO138 + Gal, $n = 11$) or athymic *nu/nu* mice ($n = 5$ per condition). Note the complete absence of treatment effect in the mice lacking functional T cells. **i**, Number of liver nodules (mean ± s.e.m.) in mice intrasplenically injected with MTO129 and treated with galunisertib alone or with antibodies against CD4 or IgG control. 95% confidence interval for galunisertib versus control is (−79 to −8), $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. **j**, Number of liver nodules (mean ± s.e.m.) after intrasplenic injection of MTO93 in syngeneic, $Lgr5^{eGFP\text{-}creERT2}$ mice—from the colonies that gave rise to the 3× and 4× genetic models—treated with galunisertib (red) or control (blue). 95% confidence interval is (−13 to −1), $n = 20$ control mice and $n = 25$ for galunisertib; two-sided Mann–Whitney *U* test. Scale bars, 500 μm.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Immune modulation by TGFβ in CRC. A human meta-cohort was used to analyse the interaction of TGFβ expression levels and T-cell-activation state. **a**, Expression signature ratio between active CD4 T_H1 and naive CD4 T cells differs significantly between MSS and MSI patients. Standardized expression for individual patients and Tukey box plots; MSS, $n = 981$ samples; MSI, $n = 198$; two-sided Wald test. **b**, **c**, Expression signature for naive CD4 T cells (**b**) and T_H1-differentiated CD4 T cells (**c**) in patients with low, medium and high expression of *TGFB1–TGFB3*. Standardized expression for individual patients and Tukey box plots. *P* values (**b**): low versus medium ($n = 327$), 1.7×10^{-4}; low versus high, 2.2×10^{-16}; medium versus high, 2.2×10^{-16}. *P* values (**c**): low versus medium, 0.0503; low versus high, 6.8×10^{-14}; medium versus high, 8.1×10^{-8}; $n = 327$ patient samples per group; two-sided Wald test. The log-likelihood ratio test *P* value is 2.2×10^{-16} for **b** and 2.3×10^{-14} for **c**. **d**, *TGFB1–3* mRNA expression levels predict poor prognosis. Hazard risk (HR) 95% confidence intervals are: low versus medium, (0.65–1.42); low versus high, (1.08–2.20); medium versus high, (1.12–2.32); $n = 255$ patient samples (low), $n = 254$ (medium) and $n = 249$ (high); two-sided Wald test. **e**, **f**, Correlation between ratio of CD4 T_H1 versus naive CD4 signatures and *TGFB1–3* mRNA expression levels (**e**) or expression of a FAP^+ CAF signature (**f**) in $n = 981$ MSS patient samples. Pearson correlation coefficients are indicated; two-sided Wald test. **g**, Kaplan–Meier curves for relapse-free survival for patients according CD4 T_H1-naive signatures ratio. Hazard risk 95% confidence intervals are: low versus medium (0.40–0.81), low versus high (0.37–0.78), medium versus high (0.64–1.41); $n = 243$ (low), $n = 259$ (medium) and $n = 257$ (high); two-sided Wald test. **h**, Cell densities in micrometastases 10 days after injection, treated with vehicle (control, blue) or galunisertib (red); shown are individual MTO129 and MTO140 tumour values and mouse means (diamonds and triangles, respectively), together with group mean ± s.e.m. derived from a mixed-effects linear model; $n = 4$ mice per condition (MTO129) and $n = 6$ for MTO140. **i**, Representative images of immunohistochemistry of the MTO140 micrometastases quantified in **h**. Mean cell densities (cells mm^{-2}) are indicated. Images represent two independent experiments. Scale bars, 100 μm.



Extended Data Figure 10 | TGFβ inhibition and the tumour immune microenvironment. a–d, Densities of tumour-infiltrating lymphocytes in galunisertib (Gal, red) or control (blue) treated liver metastases derived from primary LAKTP MTOs (**a**), in treated orthotopically injected MTO-derived primary cancers (**b**), in liver metastases in the $UBC^{creERT2}$;$Tgfbr2^{fl/fl}$ background compared to $UBC^{+/+}$;$Tgfbr2^{fl/fl}$ control (**c**), or two days after start of treatment in established liver metastases (**d**; treatment started on day 14). Shown are individual tumours (grey circles, **a**, **c, d**), mouse means (squares, or triangles for MTO93, **a**), and group means ± s.e.m. Data are analysed with a mixed-effects linear model; in (**a**) $n = 5$ (control) or $n = 4$ (galunisertib) mice per condition; in (**b**), $n = 9$ (control) or $n = 8$ (galunisertib) mice per condition; in (**c**), $n = 5$ ($UBC^{+/+}$) or $n = 7$ ($UBC^{creERT2}$;$Tgfbr2^{fl/fl}$) mice per condition; and in (**d**), $n = 6$ mice per condition. **e**, Distribution of cell types within the PD-L1$^+$ population (mean ± s.e.m.) in microdissected tumours (euthanized at day 18, four days after treatment start); $n = 3$ mice. **f**, Myeloid cell types (gated for CD45/(CD11b and/or CD11c)) within the CD45$^+$ PD-L1$^+$ population. Mean ± s.e.m. percentage of $n = 3$ mice.

nature research

Corresponding author(s): Eduard Batlle

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Experimental group sizes were practically associated to cage sizes (5 mice/cage) and treatment experiments were designed to have n >= 5 per group (1 or more cages) .

2. Data exclusions

Describe any data exclusions.

In the DSS/tam induction of the genetic model, 13% of mice died in the first 2 weeks from DSS treatment and were excluded from analysis.
One mouse was excluded from the Fig. 2f because it showed no recombination upon tamoxifen treatment.
In the KM survival curve in Fig 2e, some cohorts of mice (Con and Gal) were sacrificed at endpoint of Con (metastasis-associated severe morbidity) for tumour nodule count. Fully cured Gal mice that were culled this way were censored from the survival plot (ticks).

3. Replication

Describe whether the experimental findings were reliably reproduced.

All experiments were reproducible.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Because daily drug treatments were given as one type per cage (to avoid confusion and errors), we found the best way to randomize was to inject mice with tumour organoids in semi-random order: balanced among the predetermined treatment arms so as to not have any difference in cell viability but otherwise unpredictable to the person doing the injection.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

T (tumour) status scoring from H&E stained sections was performed by two blinded observers.
As stated above, the investigator doing the injection was blinded to treatment arm. In addition, the technicians doing the daily drug administrations were blinded to the experiment design including groups. Data collection and analysis was not blinded.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used R (v3.4.2), RStudio (v1.1.383) and FIJI/ImageJ (v1.51d), and have custom codes that are available upon request. In addition, we used QuPath (v0.1.2), GraphPad Prism (v7.03), netWHCpan (v2.8), bwa (v bwa-0.7.4), Sambamba (v0.5.9), Picard (v1.128), GATK (v3.5), snpEff (v4.1), deconstructSigs (c1.8.0), Star (v2.3.0e), GeneAtlas, FlowJo (v10.4),

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	The MTO biobank established in this paper is available to the community upon request, subject to restrictions imposed by the original mouse strain MTAs. All other materials are readily available from indicated standard commercial sources.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Only commercial antibodies have been used:
Rat αCD8α (YTS 169.4; BioXCell BE0117) or Rat αCD4α (GK1.5; BioXCell BE0003-1) or Rat IgG2b (LTF-2, BioXCell BE0090) isotype control antibodies were used for in vivo T cell depletion. All routinely validated and confirmed by flow cytometry analysis on peripheral blood (by us).
Rat αPD-L1 (10F.9G2; BioXCell BE0101) or Rat IgG2b (LTF-2, BioXCell BE0090) isotype control antibodies were used for PD-1/PD-L1 therapy. Validation reported on supplier's website.

For IHC, we used: Antibodies against CALD1 (Rabbit, Sigma, ref HPA008066; 1:250), IGFBP7 (Rabbit, Sigma, HPA002196; 1:200), phopsho-SMAD2 (rabbit, Cell Signaling, ref 3108; 1:50), CD4 (Rabbit, Sino Biological, ref 50134-R001; 1:1000), CD8 (Rabbit, Biorbyt, ref orb10325-200; 1:200), FoxP3 (Rabbit, Abcam, ref ab54501, 1:1000), T-bet (Santa Cruz, ref sc-21003; 1:500), phospho-SMAD3 (Rabbit, Abcam, ref ab52903; 1:500) and PD-L1 (Cell Signalling, ref 16764988S; 1:25); staining o/n at 4ºC. We used anti-GFP (Rabbit, Life Technologies, ref A11122; 1:500), anti-CD3 (Rabbit, DAKO, ref. IS50330; 1:30), staining for 2h at room temperature. Antibodies were validated extensively by the provider (including westernblot analysis) and confirmed by the authors for specific localization and--in case of TGF-beta targets--on modulation upon pathway inhibition in vivo.

The following (routinely used and community-validated) antibodies were used for flow cytometry staining: anti-CD45 (clone 30-F11), anti-CD4 (clone GK1.5), anti-aCD8 (clone 53-6.7), anti-CD69 (clone H1.2F3), anti-CD104b (PDGFRb, clone APB5), and anti-MHCII (clone M5/114.15.2) were obtained from eBioscience; Epcam (clone EBA-1), anti-CD3e (clone 145-2C11), anti-CD44 (clone G44-26), CD62L (clone MEL-14), anti-IFNγ (clone XMG1.2), anti-CD274 (PD-L1, clone MIH-5), and anti-Ly6C (clone AL-32) were obtained from BD Pharmingen; anti-CD31 (clone 390) was obtained from Abcam; and anti-CD11b (clone M1/70), anti-CD11c (clone H418), anti-GZMB (GB11), anti-CD279 (PD-1, clone 29F.1A12), anti-CD8a (clone 53-6.7), anti-T-bet (clone 4B10), anti-Ly6G (clone 1A8) were obtained from BioLegend. Extra validation was performed by comparing expression in internal controls (known negative populations).

For Westernblot, we used antibodies against SMAD4 (B-8, Santa Cruz, ref: sc7966) and actin (Abcam, ref: ab20272), validated on protein size and by knockout (SMAD4).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Mouse tumour organoids were derived from genetic models in our lab.
MC38 and CT26 murine cell lines as well as L cells producing Wnt3a (L-Wnt3a) were obtained from the ATCC.

b. Describe the method of cell line authentication used.

Cancer cell lines used were exome sequenced. Wnt3a production was tested in a cellular reporter assay, sensitive to Wnt ligands.

c. Report whether the cell lines were tested for mycoplasma contamination.

Mycoplasma contamination was tested for (bi-monthly) and always negative.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

None were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Genetic models: We used inbred (backcrossed) C57BL/6J mice in which we induced recombination at age 12-15 weeks, observing up to 1 year after induction or sacrificing upon severe morbidity.
From tumours/polyps in this model we derived tumour organoids (MTOs) to study metastasis vastly more efficiently than otherwise possible with purely genetic (spontaneous) mouse models. Injection and transplantation of these MTOs was performed again in C57BL/6J hosts or in BALB/c Nude mice, aged 7-8 weeks. (This age limit was made more flexible (7-10 weeks) when injecting in cohorts of genetic mice such as the UBC-creERT2;Tgfbr2-fl/fl or Lgr5-creERT2 hosts, to reduce total numbers of litters necessary.) Experimental mice were typically observed for 3-5 weeks, but treatment survivors are being kept for over 1 year after injection.
All mice were closely monitored by authors, facility technicians (during treatments) and by an external veterinary scientist responsible for animal welfare.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human research subjects.

nature research

Corresponding author(s): Eduard Batlle

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Livers with tumours were removed, lobules (and in case of micro-dissection: individual liver metastases) were carefully dissected and finely minced with scalpels. The tissue was enzymatically digested in 10 ml of DMEM supplemented with 10% FBS, 1% HEPES, sodium pyruvate, glutamine, streptomycin and penicillin and 0.1% β -mercaptoethanol (Gibco) and containing 1 mg/ml Collagenase A (Roche), 0.2 mg/ml Dispase II (Sigma) and 0.2 mg/ml DNAse I (Roche), during 25 min at 37ºC with rotation. The enzymatic reaction was quenched by the addition of 30 ml of ice-cold DMEM (10% FBS, supplement). Cell suspension was filtered through a 70 µm cell strainer (BD). The filter was washed with 10 ml of ice-cold 10% FBS DMEM and the cells were pelleted at 280 g for 5 min at 4ºC. Lysis of erythrocytes was performed in Red Cell Lysis Buffer (RCLB, 155 mM NH4Cl, 12 mM NaHCO3, 0.1 mM EDTA) during 4 min at room temperature and immediately washed with ice-cold 10% FBS DMEM, followed by filtration through a 70 µm cell strainer and centrifugation. Sequentially, cells were purified by centrifugation 30 min at 2,400 rpm in 40/80 Percoll (Sigma) gradient. Cells were resuspended in DMEM (10% FBS, supplement; FACS buffer).
6. Identify the instrument used for data collection.	Flow cytometry analysis and cell separation were performed in a FACSAriaFusion flow cytometer (Beckton Dickinson).
7. Describe the software used to collect and analyze the flow cytometry data.	Data were analyzed using FlowJo software v10.4.
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Purity was not assessed
9. Describe the gating strategy used.	Figure 1g (MTO biobank generation): To sort Lgr5-GFP+ tumour cells, we gated: Cells/Single cells/Living cells/GFP+. Figures 2g, ED Figs 7b and 8a: To assess genetic recombination in the UbC-CreERT2; Tgfbr2fl/fl; R26mTmG model, or to sort tumour cell populations, we gated: Living cells/Cells/ Single cells/PDGFRb+ (CAFs). From the negative population, we followed: Endothelial cells CD31+/CD45- and Leukocytes CD31-/CD45+, further refined to small size and good viability. Figure 3c-d: T-cell activation markers were assessed inside: Living cells/Cells/Single cells, than CD45+/ CD3+, then CD4+/CD8- or CD4-/CD8+. Figure 3e and 4g: T-cell cytokine levels were assessed inside: Living cells/Cells/Single cells/CD45+/CD3+/

NK1.1-, then CD4+/CD8- or CD8+/CD4-. Figure 3f: CD8+ T cells were sorted using gates: Living cells/Cells/Single cells/CD45+/CD19-/MHCII-/CD3+/CD4-/CD8+. Figure 4b: PD-1 expression was assessed inside: Living cells/Cells/Single cells/CD45+/CD3+, then CD4+/CD8- or CD8+/CD4-. Figure 4d and ED Fig 10e: PD-L1 populations were determined using: Living cells/Cells/Single cells/PD-L1+. Within these we measured for CAFs (PDGFRb+/Epcam-) or Epithelial cells (PDGFRb-/Epcam+). From the double-negative population, we followed: Endothelial cells (CD31+/CD45-) and Leukocytes (CD45+/CD31-). Fig. 4d: PD-L1 was measured inside Living cells/Cells/Single cells/CD45+. Figure 4f: For T-cell activation markers we gated: Living cells/Cells/Single cells/CD45+/CD3+/NK1.1-, then CD4+/CD8- or CD8+/CD4-. ED Fig. 10f: Myeloid populations were assessed inside: Living cells/Cells/Single cells/CD45+/PD-L1+/Ly6G-/Ly6C+/CD11b+.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☐