

TGF-beta drives immune evasion in genetically reconstituted colon cancer metastasis

Daniele VF Tauriello^{1,2}, Sergio Palomo-Ponce^{1,2}, Diana Stork¹, Antonio Berenguer-Llargo¹, Jordi Badia-Ramentol¹, Mar Iglesias³, Marta Sevillano^{1,2}, Sales Ibiza¹, Adrià Cañellas¹, Xavier Hernando-Momblona^{1,2}, Daniel Byrom¹, Joan A Matarin¹, Alexandre Calon^{1¥}, Elisa I Rivas^{1¥}, Angel R Nebreda^{1,5}, Antoni Riera^{1,4}, Camille Stephan-Otto Attolini¹ and Eduard Batlle^{1,2,5}

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), Spain.

3. Department of Pathology, Hospital del Mar; Cancer Research Program, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, and Autonomous University of Barcelona (UAB), Spain

4. Department of organic chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain;

5. ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain.

¥ Present address: Cancer Research Program, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, Spain

Corresponding author: Eduard Batlle (eduard.batlle@irbbarcelona.org)

Most colorectal cancer (CRC) patients die as a result of dissemination of the disease to foreign organs. Prevalent mutations associated to metastatic CRC have not been identified^{1,2}. Instead, particular features of the tumour microenvironment (TME) such as lack of T-cell infiltration³, low Th1 activity and reduced immune cytotoxicity² or elevated TGF-beta levels⁴ predict adverse outcome in CRC patients. To analyse the interplay between genetic alterations and the TME, we crossed mice bearing conditional alleles for 4 main CRC mutations in intestinal stem cells. Quadruple mutant mice developed metastatic intestinal tumours that display key hallmarks of human microsatellite stable (MSS) CRCs including low mutational burden⁵, T-cell exclusion³ and a TGF-beta activated stroma^{4,6,7}. Anti-PD-1/PD-L1 checkpoint therapy provoked limited responses in this model system. In contrast, inhibition of TGF-beta unleashed a potent, enduring cytotoxic T-cell response against tumour cells that prevented metastasis. In mice with progressed liver metastatic disease, blockade of TGF-beta signalling rendered tumours susceptible to anti-PD-1/PD-L1 checkpoint therapy. Our data reveal that elevated TGF-beta in the TME represents a primary mechanism of immune evasion that promotes T-cell exclusion and blocks Th1 effector phenotype acquisition. Thus, anti-TGF-beta signalling-based immunotherapies may have broad applications to treat CRC patients at advanced stages of the disease.

CRC progression generally coincides with successive alterations in 4 signalling pathways: WNT, EGFR, p53 and TGF-beta^{5,8}. It has been recently shown that mice bearing compound mutations in these four pathways develop metastatic CRCs⁹⁻¹¹. Similarly, we crossed mice bearing conditional alleles for 4 key human CRC mutations: *Apc*^{fl/fl}, *Kras*^{L^{SL}-G12D}, *Trp53*^{fl/fl}, *Tgfr2*^{fl/fl} and targeted gene recombination to intestinal stem cells (ISCs) by means of the *Lgr5-GFP-CreER^{T2}* driver (refs¹²⁻¹⁶). We generated 8 increasingly compound mutant strains (Fig.1a). Histopathological scoring demonstrated a stepwise increased prevalence and severity of invasive adenocarcinomas along the linear progression sequence (Fig. 1b, c and Extended Data Fig. 1c-i). 90% of LAKTP mice bear carcinomas, more than a half of which breached all intestinal layers (Fig. 1c). These cancers displayed a human-like histology with mostly medium-high degree of differentiation and abundant desmoplastic reaction (Extended Data Fig. 1d-j). Remarkably, 40% of LAKTP mice developed metastases in liver or lungs, or as carcinomatosis (Fig. 1d and Extended Data Fig. 1k-n), with a median latency of 66 days. Of note, mice bearing triple (3x) mutant genotypes presented with similarly invasive cancers but not metastasis (Fig. 1c). 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 areas of compound mutant cancers displayed high levels of stromal TGF-beta activity as indicated by the presence of phospho-SMAD3 (Fig. 1f and Extended Data Fig. 2b) and expression of CALD1 and IGFBP7 (Extended Data Fig. 2c, d), two TGF-beta-induced genes in cancer-associated fibroblasts (CAFs) that predict poor prognosis⁴. Thus, these genetic CRC models reproduce key features of the TME in human advanced 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 TGF-beta 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 full-blown invasive T3-T4 tumours (Extended Data Fig. 4a, b). 40% produced overt liver metastatic disease. Prominent T-cell exclusion and elevated TGF-beta activity in the TME was present in primary tumours from orthotopically transplanted LAKTP organoids (Extended Data Fig. 4c-f). TGF-beta-activated stroma is a defining feature of poor prognosis consensus molecular subtype 4 (CMS4)⁷. Transcriptomic analysis indicated that LAKTP MTOs classified as CMS4 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 from triple mutant MTOs. These findings confirm that stromal gene expression is required to identify poor prognosis molecular subtypes (Supplementary Discussion)^{4,17-19}.

Similar to MSS human CRCs²⁰, 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 genome of these tumours evolved extensively. The most prevalent mutational signature in MTOs was signature 1, characteristic of MSS CRCs²¹ (Extended Data Fig. 6c). MTOs and human MSS CRCs also displayed similar numbers of predicted high affinity MHC-I binding neoantigens, whereas murine CRC cell lines CT26 and MC38 exhibited a ~40-fold higher number (Fig. 1j). Experimental liver colonization by MTOs demonstrated an increased metastatic burden in the *nu/nu* background compared to C57BL/6J (Fig. 1k), implying susceptibility to T-cell-mediated adaptive immunosurveillance. Indeed, micrometastases were characterized by abundant CD3⁺ T-cells intermingled with tumour cells (Fig. 1l). Importantly, T-cells were progressively excluded at subsequent

time points (Fig. 1l-m).

Cell population profiling of human or mouse CRC samples revealed CAFs as the main contributors to TGF-beta production (Extended Data Fig. 7a-b). We made use of the TGFBR1 specific inhibitor Galunisertib²² to inhibit TGF-beta 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 a reduction of pSMAD3+ cells (Fig. 2b) and gene expression profiling demonstrated decreased levels of TGF-response signatures⁶ of fibroblasts (F-TBRS) and T-cells (T-TBRS) (Fig. 2c).

To test therapeutic effects on liver metastatic disease directly, we inoculated in the portal circulation LAKTP MTOs derived from either primary CRCs or liver metastases (LiM), the latter displaying enhanced metastatic capacity (Extended Data Fig. 6e). Treatment with Galunisertib dramatically decreased metastatic burden for all MTOs (Fig. 2d and Extended Data Fig. 6f), effectively curing a large proportion of mice (Fig. 2e). Galunisertib blocked TGF-beta signalling in the TME of metastasis (Extended Data Fig. 7c-h). We also transplanted LAKTP MTOs in mice bearing floxed *Tgfr2* alleles that express a ubiquitous CreER^{T2} recombinase (*Ubc-CreERT2*; *Tgfr2^{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 high numbers of cells. 4 out of 6 tested MTOs (2 LAKP and 2 LAKT) gave rise to metastases yet 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 8b, c). We also introduced loss of function *Smad4* mutations in LAKP MTOs using CRISPR/Cas9 technology. *In vitro*, LAKP+S MTOs were resistant to the cytostatic action of TGF-beta (Fig. 2i, j). *In vivo*, LAKP+S formed more and larger liver tumours than parental LAKP MTOs. Galunisertib effectively abrogated metastatic capacity of LAKP+S MTOs and, importantly, did not enhance growth of metastases generated by *Tgfr2* wildtype LAKP MTOs (Fig. 2k and Extended Data Fig. 8d).

To study the kinetics of metastasis, we transduced MTOs with luciferase, which did not modify therapeutic efficacy of Galunisertib (Extended Data Fig. 8e). Bioluminescence revealed that Galunisertib markedly enhanced cell killing at the onset of exponential growth phase. Yet, this phenomenon did not occur in *nu/nu* mice (Extended Data Fig. 8f-h). These kinetics suggested an adaptive anti-tumour immune

response. Indeed, therapeutic effect of Galunisertib was abolished upon depletion of CD8⁺ cytotoxic T-lymphocytes (CTLs) (Fig. 3a, b) or CD4⁺ helper T-cells (Extended Data Fig. 8i). We also transplanted MTOs in mice from the colony utilized to generate the original compound genetic models. Galunisertib decreased metastatic burden in these hosts (Extended Data Fig. 8j), implying that immunological rejection was not due to expression of exogenous antigens in MTOs such as those encoded in GFP and CreER^{T2} 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 of naïve or Th1 activated CD4⁺ T-cells. Microsatellite-unstable (MSI) CRCs displayed elevated ratios of Th1-to-naïve T-cell genes, in line with previous observations²³ (Extended Data Fig. 9a). In MSS CRCs, the Th1-to-naïve ratio followed an inverse correlation with *TGFB* or with the CAF-specific gene expression programme (Extended Data Fig. 9b-e) and predicted disease relapse (Extended Data Fig. 9d, g). Therefore, abrogated T-cell differentiation, elevated TGF-beta and CAF gene expression characterize a substantial subset of poor prognosis MSS CRC patients.

Flow cytometry analyses on whole livers colonized by MTOs for 7-10 days revealed increased recruitment of CD3⁺ and CD4⁺ cells upon Galunisertib treatment (Fig. 3c). This finding was confirmed by IHC-based quantifications on liver sections, which also revealed increased positivity for T-bet, the master transcription factor regulating Th1 differentiation (Extended Data Fig. 9h, i). Equivalent observations were made in treated liver metastases from parental MTOs, in orthotopic implantation, and during genetic abrogation of TGF-beta signalling in the TME (Extended Data Fig. 10a-c). Importantly, both T-helper cells and CTLs underwent activation, as evidenced by the increase of CD44⁺/CD62L⁻ and CD69⁺/CD62L⁻ populations (Fig. 3c, d), elevated levels of T-bet and of effector molecules IFN γ in T-helper cells, and increased GzmB production and proliferation gene expression in CTLs (Fig. 3e, f). We hypothesized that this robust immune response might immunize mice against the tumour cells. When survivors were re-challenged with the same MTO, most tumours were rejected within 2 weeks in the continued absence of treatment, whereas MTOs grew efficiently in naïve C57BL/6J recipients. Concurrent depletion of CD8⁺ CTLs inhibited tumour rejection (Fig. 3g). We conclude that elevated TGF-beta levels in the TME limit adaptive immune responses by inhibition of the Th1 effector phenotype.

Finally, we assessed therapeutic action of Galunisertib in mice bearing overt metastatic disease. Treatment 14 days after MTO inoculation reduced metastatic burden but resulted in few complete remissions (Fig. 4a-c). Nevertheless, we observed

increased infiltration of CD4⁺ T-cells and of T-bet⁺ lymphocytes immediately after therapy initiation (Extended Data Fig. 10d). Exploring resistance mechanisms to explain this mild response, we discovered that Galunisertib-activated T-cells displayed a marked surface expression of PD-1 (Fig. 4b) and that metastases generated by MTOs progressively recruited stromal PD-L1⁺ cells, most 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-1/PD-L1 therapy alone had a very modest therapeutic effect (Fig. 4e). However, combined Galunisertib and anti-PD-L1 treatment induced a pronounced immune response, with elevated T-bet and IFN γ levels in CD4 T-helper cells and increased GZMB production in CTLs, which strikingly eradicated most metastases and prolonged recurrence-free survival for over a year after treatment cessation (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 Th1 immune activation.

It has been proposed that MSS CRCs are immunologically cold and thus that they are unlikely to benefit from immune therapies (Supplementary Discussion)²⁴. In contrast, our data reveal that this class of CRCs can be effectively killed by the adaptive immune system through a CTL-dependent process, which CRC cells avert by raising TGF-beta levels. In concordance with the well-established role of TGF-beta signalling in suppressing differentiation and activity of T-cells²⁵⁻²⁸, we observed that a TGF-beta-activated TME antagonizes the Th1 effector phenotype. We also show that such a TME excludes T-cells from tumours, a phenomenon associated to poor outcomes across cancer types^{3,29,30}. Enabling immune infiltration using TGF-beta inhibitors is sufficient to confer susceptibility to anti-PD-1/PD-L1 checkpoint-based therapies, a strategy that may have broad application for treatment of cancers that thrive in a TGF-beta-rich environment. These results immediately suggest the use of TGF-beta signalling inhibitors as immunotherapy for stromal-rich poor prognosis MSS CRC patients.

Acknowledgements

We thank all members of the Batlle laboratory for support and discussions, and the López-Bigas lab for help with the revision. We are grateful for the outstanding assistance by the IRB Barcelona core facilities for histopathology, functional genomics, mouse mutant, and advanced digital microscopy as well as the flow cytometry, animal facilities of the UB/PCB, and the CRG genomic unit. DT held a Juan de la Cierva postdoctoral fellowship from MINECO. This work has been supported by grants from 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 the MINECO.

Author contributions

DT, SPP, DS and ACL performed animal husbandry and genotyping; MS performed IHC; DT and MI analysed histopathology. DT generated MTOs, which DT and DS characterized *in vitro*; DS performed CRISPR experiments; CSOA and ABL performed exome/RNAseq analysis, other bioinformatics (CMS classifier and patient data) and statistical analyses. SPP, DT, JBR, ACN and XHM performed mouse isografting; DT, DS, and JBR quantified IHC. DB, JAM and AR synthesized Galunisertib. DT coordinated/performed animal treatments and analysed the data. DT, JBR, SI, EIR and ARN performed immunophenotyping experiments. EB and DT conceived the study, coordinated experiments and wrote the manuscript. EB supervised the study.

Author information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to eduard.batlle@irbbarcelona.org.

References

1. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4283–8 (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**, 1960–4 (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, T. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–7 (2012).
6. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-beta-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.* **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. Melo, F. de S. e *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 beta-catenin signaling and leads to hepatocellular carcinomas. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17216–21 (2004).
13. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–8 (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**, 560–8 (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**, 312–9 (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**, 214–8 (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**, 553–60 (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 TGF β signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12**, 171–81 (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-beta Receptor. *Immunity* **25**, 441–454 (2006).
27. Thomas, D. A. & Massague, J. TGF-beta 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-beta 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 (80-.).* **348**, 74–80 (2015).

Figure legends

Figure 1. Analysis of compound mutant mouse models and MTOs. **a)** Permutations of alleles used to generate mouse models. **b)** AJCC classification system. **c)** Worst T phenotype diagnosis per mouse, n mice indicated within the circles. **d)** Metastasis in LAKTP mice. Outer ring: associated type of primary cancer. **e)** T-cell density in Normal Mucosa, Adenoma, compared to 2x, 3x and 4x cancers, n=119, 242, 7, 8, 16 ROIs. Right: CD3 staining of an LAKTP T4 carcinoma. **f)** p-SMAD3⁺ cell density in normal submucosal tissue (NSM), compared to 2x, 3x and 4x cancers, n=37, 5, 10, 16 ROIs. Right: example of pSMAD3 IHC. **e-f)** Shown are Tukey boxplots; two-sided Welch's *t*-tests. **g)** MTO biobank and isograft technology. **h)** Average sensitivity of MTOs to niche factors. **i)** Left: Cross-validation of CMS4 classifier on human CRCs. Patients (vertical bars) are ranked by subtype on score. Right: classifier applied to MTOs *in vitro* (grey, n=5, 3) and orthotopically isografted (orange, n=2, 3). **j)** Predicted MHC-I neoantigens in MTOs (n=6, 10), human CRCs (n=266, 112) and mouse CRC cell lines (n=2). **k)** Liver metastasis generated by 4x MTOs in C57BL/6J or *nu/nu* mice (mean \pm SEM, n=10, 5, 11, 5, 10, 5 mice, two-sided MWW tests). **l)** CD3 IHC on MTO129 liver mets, at indicated days after intrasplenic injection. **m)** CD3⁺ cell densities (individual mets in grey and mouse mean in black, bars: group means \pm SEM), analysed by a mixed-effects linear model, n=4, 3, 3, 2, 2 mice and n=39, 58, 71, 56, 56 tumours. *Scale bars: 500 μ m.*

Figure 2. Therapeutic effect of TGB-beta inhibition. **a)** Galunisertib reduces tumour burden and metastasis in orthotopic MTO140 isografts. Mean volume \pm SEM; right: median number \pm 95%CI; n=21 mice per condition, two-sided MWW tests. **b)** pSMAD3⁺ cell density (mean \pm SEM) in primary carcinomas (n=9, 7); two-sided Student's *t* test. **c)** Expression levels of TBRS in primary CRCs. Tukey boxplots, n=6 tumour samples, two-sided MWW tests. **d)** Liver metastases (mean \pm SEM) generated by MTOs, treated with vehicle or Galunisertib; n=5, 5, 5, 5, 15, 17, 12, 12, 20, 25 mice; analysed by a mixed-effects linear model. **e)** Representative livers at endpoint; Kaplan-Meier survival curve of mice treated as in **(d)**, n=41 mice for Control, 35 for Galunisertib; Mantel Cox test. **f)** Liver metastases (mean \pm SEM) in *Tgfr2^{fl/fl}* mice, recombined (n=9) or not (n=6). **g)** Liver tumours from **(f)**, stained for recombination marker GFP, representative of 3 experiments; *Scale bars: 1mm.* Below: percentage of GFP⁺ (mean \pm SEM). **h)** Liver metastases (mean \pm SEM) generated by 3x LAKP MTOs, n=15, 14 mice; two-sided MWW test. **i)** Western blot for SMAD4 and Actin of LAKP (empty guide, eg) or CRISPR targeted *Smad4* KO organoids. **j)** LAKP or LAKP+S

organoids (mean \pm SEM, n=4 technical replicates) in control medium or treated with TGF β 1. **k**) Liver metastases (mean \pm SEM) from LAKP (eg) and LAKP+S (guide #1) MTOs, n=5 mice per condition; two-sided MWW test. Right: tumour diameters with SEM, n=10 mice per condition.

Figure 3. TGF-beta mediates immune evasion. a-b) Liver metastases (mean \pm SEM), n=5, 10, 5, 10 mice (**a**) and n=5 per condition (**b**); two-sided MWW tests. **c-e)** Immunophenotyping in whole livers from mice (n=5 per condition) injected with MTO138 or MTO140. Tukey boxplots, n=5 mice per condition; analysed by a mixed-effects linear model. **d)** Flow cytometry plots from (**c**), representative of 2 experiments. **e)** Intracellular cytokine expression (mean percentage \pm SEM), n=4, 6, 4, 6, 5, 5 mice; two-sided Student's *t* tests. **f)** Relative mRNA expression (mean \pm SEM) on sorted CTLs from the experiment in (**f**); n=4 mice per condition, two-sided Student's *t*-tests. **g)** Rechallenge experiment in liver metastasis survivors compared to naïve hosts. Tumour volumes (mean \pm SEM) at end point. Numbers of grown tumours (over injections) are indicated.

Figure 4. Dual immunotherapy cures established metastases. a) Liver metastases (mean \pm SEM) in animals treated from day 14, n=4, 6 mice. **b)** PD-1⁺ T-cells (mean \pm SEM) in microdissected liver mets 4 days after treatment start (d18), n=3 per condition; two-sided Student's *t* tests. Right: representative density plots. **c)** PD-L1 IHC on MTO129 liver mets. **d)** PD-L1⁺ leukocytes in liver mets (mean \pm SEM) at indicated days after injection, or in non-injected liver, n=4, 5, 5, 6, 3. **e)** Treatment from d14 of established liver metastases. Mean liver mets \pm SEM, n=11, 13, 12, 14 mice, two-sided MWW tests. Right: Kaplan-Meier curve for survival; n=6 (Con), 7 (Gal), 6 (α PD-L1), 7 (Dual), Mantel Cox test. **f-g)** Surface activation markers (**f**) and intracellular markers (**g**) in mice sacrificed 2 days after treatment start (d16), mean percentage \pm SEM, n=6, 7, 6, 7 mice; two-sided Students. **h)** Cell densities quantified by IHC from mice in (**e**). Individual metastases in grey, mouse means in coloured squares, bars are group means \pm SEM; n=5, 4, 4, 4 mice, analysed by a mixed-effects linear model. **i)** Images from the analysis (**h**), mean cell densities are indicated. *Scale bars: 500 μ m; **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001.*

Methods

Ethics and general animal study statements

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 are maintained in specific-pathogen-free (SPF) facility with a 12-hour light-dark cycle and fed with standard diet and water *ad lib*. All mice were closely monitored by the authors (DT and SPP), facility technicians (during treatments) and by an independent veterinary scientist responsible for animal welfare.

Genetic animal models

Lgr5-EGFP-IRES-CreER^{T2} (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 TβRII^{LoxP} (B6.129-*Tgfr2*^{tm1Kar/J}; Stock 012603) mouse strains^{13–16} were obtained from The Jackson Laboratory, and Apc^{LoxP} mice obtained from Christine Perret¹². Mice were all inbred in C57BL6/J background and successive crosses were performed to combine alleles. In this study, we have abbreviated the alleles: L (Lgr5-EGFP-IRES-CreER^{T2}), A (Apc^{LoxP}), K (LSL-K-ras G12D), T (TβRII^{LoxP}), P (p53^{LoxP}) and generated the following strains: LA/K, LAT/K, LAP/K, LATP/K. For practical reasons, the strains were kept in homozygous (lox/lox) status when possible (A, T, P) and used both K heterozygous and wildtype littermates. UbC-CreER^{T2} (B6.Cg-Tg(UBC-cre/ERT2)1Ejb/2J; Stock 008085) and Rosa26^{mTmG} (B6.129(Cg)-Gt(ROSA)26Sortm4 (ACTB-tdTomato,-EGFP)Luo/J; Stock 007676) mice were described before^{31,32}. The latter allele allowed us to detect recombination by 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 lib*) containing 2.5-3% (w/v) dextran sodium sulfate (DSS) for 5 days, while given 2 ip injections of diluted (4 mg/kg) tamoxifen (Sigma; dissolved in sunflower oil with 0.5% EtOH) on days 0 and 5. 13% of mice treated this way died in the first 2 weeks, most likely due to DSS toxicity. These losses were excluded from analysis. Animal weight, stool type and overall appearance were scored 2x per week. At increasing morbidity, mice were evaluated more frequently until reaching the experimental endpoint: progressive or rapid weight-loss/emaciation and poor physical appearance, characterized by anaemia, hunched posture, unkempt appearance and lethargy. Mice were then sacrificed and dissected, inspecting the peritoneum wall and

harvesting intestines, mesentery, liver, spleen, kidneys, diaphragm and lungs. After macroscopic/binocular analysis, tissues were washed in PBS, fixed in 10% formalin solution (Sigma) overnight and embedded in paraffin. Survival statistics were analysed using GraphPad Prism software (v7.03). Median latency of metastasis was obtained by taking the median survival of the animals where metastasis was detected.

Although the DSS-treatment-mediated induction of inflammation is acute and subsides after 2-3 weeks, we have also induced gene recombination with tamoxifen without DSS treatment. Without this acute inflammation, induction still gave rise to equivalent tumours as in this study (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 for DSS is in targeting the colon. The experiments without DSS are not included in the survival and tumour assessment of this study.

Recombination in the UbC-CreER^{T2} background were performed with 2 shots of 80 mg/kg tamoxifen at day 2 and 3. One mouse was excluded from analysis in Fig. 2f because it showed no recombination upon tamoxifen treatment.

Histology and immunohistochemistry (IHC)

Standard H&E and immune-stainings were performed on 4 µm tissue sections using standard procedures, as described before¹. Antibodies against CALD1 (Rabbit, Sigma, ref HPA008066; 1:250), IGFBP7 (Rabbit, Sigma, HPA002196; 1:200), phospho-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) were used for staining o/n at 4°C. Anti-GFP (Rabbit, Life Technologies, ref A11122; 1:500), anti-CD3 (Rabbit, DAKO, ref. IS50330; 1:30) were used for staining for 2h at room temperature. Photographs from histology/IHC were taken with a Nikon eclipse E600 and Nikon DS-Ri1 camera or with a Hamamatsu NanoZoomer Digital Slide Scanner (20x).

Histopathological quantifications

H&E stained sections of intestines were blindly scored for T (tumour) status by an expert pathologist (MI) and a second observer (DT), using the AJCC-TNM system: scoring for Tis (carcinoma in situ), T1 (adenocarcinoma with submucosal invasion), T2 (intramuscular invasion), T3 (transmuscular invasion; serosa intact or cannot be evaluated), T4 (transserosal invasion; T4b when local metastasis is evident, otherwise T4a). Across all genotypes, 156 tissue samples were scored (both small intestine (SI)

and large intestine (LI): caecum, colon and rectum) from 68 mice across 8 genotypes. Given the high expression of Lgr5 there, the distal ileum and caecum 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 SI surface and genotype. All invasive adenocarcinomas (SI + LI) as well as LI adenomas were counted individually, reaching a total count of 1477 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 minutes (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, membrane integrity marker) and GFP⁺PI⁻ cells were sorted in a FACS Aria flow cytometer (BD Biosciences). Typically, ~1000 cells were obtained and plated in a drop of cold basement membrane extract (Trevigen Cultrex BME Type 2): 40 µl in a prewarmed (37°C) standard 24^{well} plate (Corning) well. After 5 minutes, mouse tumour organoid (MTO) culture medium was added: 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 (Peprotech) and 1 µM Galunisertib (LY2175299, see below). (NOGGIN was also produced in-house: expressed as His-tagged protein in HEK293-6E cells and purified with a mammalian affinity 5 ml Ni column and a 5 ml HisTrap HP column in an ÄKTAexpress at 4 °C. Protein fractions were pooled and desalted (HiPrep 26/10 column), and tested on mouse Apc-mutant adenoma organoids or on BMP-sensitive PDOs, analysing *ID1/3* gene expression by qPCR.) In initial passages, MTO-medium was supplemented with antimicrobial reagent Normocin (InvivoGen). MTOs were cultured in 37°C incubators with 85-90% humidity, atmospheric O₂ and 5% CO₂.

For MTO passaging, BME 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 (pipetting) until a single cell suspension was obtained. Trypsin was quenched with FBS, cells were washed with HBSS and replated in cold BME 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 knock-out, 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 capitals):

Smad4: #1 F: caccgAGACAGGCATCGTTACTTGT and R: aaacACAAGTAACGATGCCTGTCTc

Smad4: #2 F: caccgAGTTTGATGTGTCATAGACA and R: aaacTGTCTATGACACATCAAACCTc

mCherryLuc: F: caccGCGCATGAACTCCTTGATGA and R: aaacTCATCAAGGAGTTCATGCGC

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 (1.0-1,5x10⁶ 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 and cultured in full growth medium. For Smad4 KO mutants, 3 days after nucleofection growth medium was exchanged for selection medium (+TGFB1). Selection of mCherry-Luc KO mutants was achieved by FACS sorting of IRFP/mCherry double negative cells 2 subsequent times.

Western blotting

Trypsinized MTO cell pellets were resuspended in lysis buffer (1 mM EDTA, 1 mM EGTA, 1 % SDS) containing protease inhibitor cocktail (Sigma Aldrich). Protein concentration was determined using standard Bradford assay (BioRad). 30 µg per sample was separated by SDS-PAGE and transferred to 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, ref: sc7966) and actin (Abcam, ref: ab20272) were used. The secondary antibody was goat anti-mouse IgG HRP conjugated (Pierce, ref: 31430).

Immunohistochemical quantifications

Scanned CD3, CD4, CD8, FoxP3, T-bet, and pSMAD3 stainings were analysed in QuPath (v0.1.2) using the Positive Cell Detection feature with empirical parameters. Several regions of interest (tumours) were taken per section. In case multiple sections per mouse/liver were considered, care was taken not to quantify the same tumour more than once. Data were processed and visualized with R/RStudio³³ (v3.4.2/v1.1.383) and the *ggplot2*³⁴ package (v2.2.1). See *Statistics and Reproducibility* (below) for more information.

Neoantigen prediction (abbreviated)

SNVs annotated as non-synonymous or nonsense and inframe insertions or deletions were selected for each sample. Peptides including 10 bases up and downstream from the alteration were extracted from the canonical protein associated with the corresponding gene and discarded if matching any other annotated protein. Human CRC: total number of mutations and predicted neoantigens for human colorectal cancer (CRC) samples were downloaded from [<https://tcia.at/neoantigens>]. Samples were classified as MSS or MSI according to the TCGA consortium⁵. Both for mouse and human data, neoantigens were predicted using the netMHCpan software (v2.8). 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 at 6 weeks of age and injected at 7 weeks. Sex was matched with the origin of the tumour. Intrasplenic (IS) or portal vein (PV) injections were used for liver colonization by the introduction of dissociated organoids (single cells) into the portal circulation: MTOs were cultured in standard 6^{well} plates for 4 days and trypsinized (see above). 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 to 5 x 10⁵ cells in 70 µl per mouse. IS injections were performed as previously described^{4,6,35}. For PV injections, a 30G syringe was used to inject 100 µl of cell suspension directly into the PV. Visible liver metastases were counted after sacrifice at 3-5 weeks and data were analysed using GraphPad Prism software (v7.03). For subcutaneous injections, MTOs were cultured in standard 6^{well} plates for 3 days and harvested with Cell Recovery Solution (BD Biosciences), 40 minutes on ice. Organoids were washed (HBSS) and a 10% fraction was trypsinized to count cells (Neubauer or TC20 automated cell counter, Bio-Rad). Organoids were suspended in HBSS with 30% BME, an equivalent of 1.5 x 10⁵ cells was injected in 2-4 flanks per mouse. Growth was scored by palpation and volume quantified with a caliper (multiplying 3 orthogonal diameters and dividing by 2) twice per week. In accordance with our IACUC protocol, mice with tumours over 300 mm³ in volume were sacrificed. Intracaecum (IC) injections were used for orthotopic isograft formation: MTOs were harvested as for subcutaneous injection, and an equivalent of 5 x 10⁵ cells in 11 µl was injected per mouse. IC injections were performed using a 30G needle under binocular guidance, adapted from³⁶. Mice were sacrificed 10 weeks after injection and metastasis was scored macroscopically as well as histologically. Alternatively, orthotopic primary tumours were generated by transplantation of a 1 mm³ piece of subcutaneously grown MTO-derived tumour (e.g. for 3x 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). At sacrifice, the caliper technique was used to measure orthotopic tumour sizes. 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 (aka LY2157299) was synthesized in-house and prepared as previously described⁶. Galunisertib or control vehicle was administered by gavage in a 0.15 ml volume, twice per day, starting 2 days after cell injection for metastasis initiation experiments, unless otherwise indicated (treatments starting 2 days after injection give identical results to pre-treating from day -2, before injection). A dose of 800 mg/kg per treatment was used for all experiments except for the treatment of MTO138, or for the prolonged treatment of orthotopically implanted mice, where 400 mg/kg doses were used. For *in vivo* CD8⁺ T cell depletion, Rat α CD8 α (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 μ l was administered per mouse by ip injection on day -1, 6 and 13 (unless otherwise indicated), respective to the day of cancer cell injection. For CD4 T helper cell depletion, Rat α CD4 α (GK1.5; BioXCell BE0003-1) or Rat IgG2b isotype control antibodies were used. For checkpoint immunotherapy or dual immunotherapy, we used Rat α 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 (5 mice/cage) and treatment experiments were designed to have $n \geq 5$ per group (1 or more cages). No mice were excluded from the analysis, unless explicitly indicated in the methods. For gavage (galunisertib) treatment, as control vehicle and galunisertib are visually distinguishable, the only randomization we performed was the order of injecting mice: researcher performing the injections was blinded to treatment group. End point criteria are equivalent to those described above. For genetic models, littermates were used as controls and cohorts accumulated over time for logistic reasons. Mice were excluded if after a year of survival post induction, not a single tumour was observed. End point criteria are described above.

Statistics and Reproducibility

When no statistical control for confounders was needed, number of metastases between conditions were compared using a two-sided Mann-Whitney-Wilcoxon test

(MWW) with exact computation of the null hypothesis, as implemented in the `wilcox_test` function from the R³³ *coin* package³⁷ or in GraphPad Prism. For Figs 2d and Extended Data Fig 8j, where data was 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 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 9e and 10a) or more than one experiment run (Extended Data Figs 9e 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 9e, 10a, and 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. IHC quantifications: besides plotting individual tumours (grey), results were represented at the mouse level (coloured/black squares or MTO-dependent symbols) and at group level using the adjusted means and standard error derived by the models in their original scale, after undoing the transformation. For doing so, standard errors were computed by simulation from the corresponding model³⁸. Wald tests derived from the linear models were used to perform pairwise comparisons between the experimental conditions. A 5% level was chosen for statistical significance. All the analyses were performed using R³³. 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 animals were metastasis-free at sacrifice. Parts of the experiment were sacrificed at the endpoint 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 boxplots in Figs. 1e, f and j, 2c, 3c and Extended Data Figs. 3d, 7a, and 9a-c, have whiskers of maximum 1.5 IQR; 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) vs less severe ($< T3$), are: 0.0002 for LAKTP vs 3x genotypes combined, and < 0.0001 for LAKTP vs 2x genotypes combined and for LAKTP vs LA (Fisher test). We also tested for the presence ($\geq T1$) or absence of carcinomas and the

P values for these comparisons are: 0.0013 for LAKTP vs 3x, 0.0005 for LAKTP vs 2x, and <0.0001 for LAKTP vs LA (Fisher test). The IHC images in Fig. 1e, f are representative of 4 independent experiments. In Fig. 1i, numbers of patients are indicated; MTOs in vitro: 3x n=5 individual MTOs, and 4x: n=3; plotted values are means of biological duplicates. MTOs orthotopically transplanted: 3x: n=2 individual MTOs (MTO54 and 207), 4x: n=3 individual MTOs (MTO68, 93 and 140); plotted values are biological duplicates (2n). Experiment in Figs. 1l, m was performed once. In Fig. 2d, a mixed-effect linear model was fitted in which experiment run and MTO were included as a random effect. Besides providing individual P values, this model was used for the overall assessment of Gal treatment effect: control mean [95%CI]: 24.0 [10.5, 47.7] vs Gal: 0.084 [0.00, 0.81]; $P=1.2 \times 10^{-12}$. Data are from 10 independent experiments. In Fig. 2e, data are from 7 independent experiments, using either MTO129 or 138. In Fig. 2h, data are from MTO54 and MTO220, combined. Fig. 2i, j, using MTO220, is representative of two independent experiments. For WB source data, see Supplementary Figure 1. In Fig. 2k, right, the values are fraction of total +SEM. Data are from experiment with empty guide or from guides #1+ #2 LAKP+S combined, from two independent experiments. Fig. 3a was performed with MTO138, 3b with MTO129. In Fig. 3c, a linear model was used as described above, considering the overall (MTO combined) treatment effect. In Fig. 3g, P values are: for the comparison between α CD8 and IgG in the survivors is 0.0043; between Rechallenge IgG and Naive IgG, and between Rechallenge IgG and Naive α CD8, the P value is < 0.0001; two-sided MWW test. Fig. 4b, c are representative for 2 experiments. In Fig. 4d, the MTO injected is MTO129. In Fig. 4f, P values are for CD4: CD62L: 0.0002 (Control vs Gal) and 0.0051 (Control vs Dual); CD44: 0.0008 (Control vs Gal) and 0.0297 (Control vs Dual); CD8: CD62L: 0.0159 (Control vs Gal) and 0.0181 (Control vs Dual). In Fig. 4g, P values are for CD4: IFN-gamma: 0.0099 (Control vs Gal) and <0.0001 (Control vs Dual); T-bet: 0.0208 (Control vs Gal) and 0.0017 (Control vs Dual); CD8 GZMB: 0.0054 (Control vs Gal) and 0.0001 (Control vs Dual). In Fig. 4h, statistics were performed as described above. All P values are available in the source data, online. The experiment in Fig. 4i was performed once.

Patient data analysis

In this study, we used five Affymetrix microarray datasets publicly available in the NCBI GEO repository, which include gene expression and clinical information from a total of 1.194 CRC patients. In order to obtain CD4 naive and CD4 th1 profiles, we carried out a differential expression analysis on gene expression dataset GSE22886⁴¹. We defined the CD4 naive signature to include genes that are at least 3 fold upregulated both in CD4 naive vs. Th1 and in CD4 naive vs. Th2, and Benjamini and

Hochbergs' False Discovery Rate (FDR)⁴² < 5% in both comparisons. Similarly, the CD4 Th1 signature included genes that were at least 3 fold upregulated 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 (LRT), while 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 by 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 mice tumours according to Consensus Molecular Subtypes

We used the *k*-Top Scoring Pairs algorithm and trained the classifier for a binary outcome (CMS4 vs not-CMS4) in 960 samples. See Supplementary Methods and Extended Data Fig 5 for a detailed description.

Data availability and reproducibility

MTO whole exome sequencing and RNAseq data have been deposited in the ENA archive with accession number PRJEB22559. Expression arrays data are available at GEO with accession number GSE103562. Computer code is available upon request. Source data are available online.

Method references

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. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2016). at <<http://www.r-project.org>>
34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag

New York, 2009). at <<http://ggplot2.org>>

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**, 1077–85 (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. lme4: Linear mixed-effects models using Eigen and S4. (2015). at <<http://arxiv.org/abs/1406.5823>>
40. Kuznetsova, A., Bruun Brockhoff, P. & Christensen, Haubo Bojesen, R. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-32. (2016). at <<https://cran.r-project.org/package=lmerTest>>
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. Ser. B* **57**, (1995).
43. Gentleman, R. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Journal of the American Statistical Association* **102**, (Springer New York, 2005).

Extended Data Figure Legends

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 between brackets; Mantel-Cox test. **b)** Counted tumour numbers (both 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 \pm SEM. **c)** Number and grade of adenocarcinomas per individual mouse (bars), by genotype. P value for pairwise comparisons of tumour number (taking LAK, LAT and LAP as 2x, and LAKT, LAKP and LATP as 3x): LA vs 2x: 0.0198, LA vs 3x: 0.0337, LA vs LAKTP: 0.0001, 2x vs LAKTP: 0.0051, and 3x vs LAKTP: 0.0014; n=9 mice (LA), 8 (LAK, LAT, LAKT), 7 (LAP, LAKP), 10 (LATP), 13 (LAKTP), two-sided MWW tests. **d-i)** Example images of tumour types scored by the TNM criteria, representative of 23 independent experiments. Indicated with arrows are invasive cells or glands, absent in an adenoma (**d**), constricted to within 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 H&E 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 induced LAKTP mice to the mesentery, diaphragm, liver and lung; each representative of 2 experiments. *Scale bars: 500 μ m.*

Figure 2. Analysis of the TME of LAKTP adenocarcinomas in the genetic model. a-a'') Immune infiltration in LAKTP carcinomas. IHC for CD3, indicating infiltrating T cells, in an intestinal adenocarcinoma (CA) of an LAKTP mouse (**a**). Insets for normal mucosa (NM, **a'**) and tumour centre (TC, **a''**) are shown. Arrowheads indicate T cells. **b-d''')** Markers of TGF-beta activation and poor prognosis in LAKTP carcinomas in a representative section of intestine with two aggressive invasive adenocarcinomas (CA). **b-b''')** Staining for phospho-SMAD3. Note nuclear staining in stromal cells of the tumour centre (TC). **c-c''')** Staining for CALD1. Note its expression in the muscle layer (ML) throughout the intestine, and in a subset of cells in the stroma of the normal mucosa (NM). In the tumour centre (TC) of invasive carcinomas, the staining intensity is clearly higher, especially in fibroblast-like cells. **d-d''')** Staining for IGFBP7, which is hardly detectable in the normal mucosa (NM). Expression is strong in the TME of tumour centre (TC). Images are representative of 4 independent experiments *Scale bars: 1mm (overviews in b-d); 500 μ m (a, b', c', d'); 100 μ m (rest: NM and TC)*

Figure 3. Niche factor analysis of Mouse Tumour Organoids (MTO) *in vitro*. Organoid formation assay in different media with representative MTOs from LA and LAKTP genotype. **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 TGFB1. **b)** Results from automated organoid detection for the MTOs depicted in **(a)**, data points are counted organoids from three replicate wells for representative analyses. Shown are mean \pm SEM. **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, both from primary and LiM tumours. Data (individual values and Tukey boxplots) are normalized to the mean in full medium. *Scale bars: 1 mm.*

Figure 4. Histology and TME of primary tumours from orthotopically isografted LAKTP MTOs. a-b'') H&E staining of an LAKTP T4 carcinoma in the genetic model **(a)** and of a recapitulated adenocarcinoma developed from orthotopically isografted LAKTP MTO93 **(b)**. Bottom: Liver and lymph node (LN) metastasis (Met) observed in orthotopic isografted mouse. *Iso: primary isograft; ML: muscle layer; MM: muscularis mucosae; NM: normal mucosa.* **c-f)** Markers of poor prognosis in isografted primary tumours. **c-c')** IHC for T cell marker CD3 **(c, c''-c''')** and H&E stain **(c')** of an orthotopic isograft, with typical glandular differentiation. Note the paucity of infiltrating T cells in the tumour centre (TC, **c''')**, compared to their presence at the periphery and in the normal mucosa (NM, **c''**). **d-f)** Images of an orthotopically isografted, invasive primary tumour, stained for pSMAD2 **(d)**, arrows indicate positive stromal cells), CALD1 **(e)** and IGFBP7 **(f)**. Images represent 3 independent experiments. *Scale bars in c-c': 1 mm; in the rest: 500 μ m.*

Figure 5. Schematic of the CSM4 classifier, using the k-TSP algorithm. a) A metacohort of 4 pooled human CRC datasets with CMS annotation is used to find up and down regulated genes in the CMS4 subtype (434 with FC higher than 1.5 and 300 with FC lower than 0.8). In order to ensure usability in mouse samples, candidate genes are filtered by high-confidence homology between the two species. The *k*-Top Scoring Pairs (*k*-TSP) algorithm selects among the up (UP) and down (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 pair of genes. For each sample, the score is increased by one unit if the expression of the UP gene is larger than that of the DOWN gene. The maximum score for CMS4 classification is 14. **b)** Cross-validation by a leave-one-out approach. In order to assess

performance, we repeatedly trained a classifier on 3 out of the 4 human datasets and tested in the remaining one. **c)** Results for 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 (ROC curve, shown on the right).

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). Below: heatmap and table of the number of indicated type of somatic mutation per MB of sequenced exome. **c)** Mutational signatures in the MTOs and mouse 2D cell lines CT26 and MC38. Like 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 *Ppia* $2^{-\Delta CT}$, and normalized to that of MTO93; shown as three technical replicates with mean \pm SEM. **e)** Number of liver nodules after intrasplenic injection of primary LAKTP tumour and spontaneous (Spont. LiM) or experimental liver metastasis-derived MTOs in syngeneic C57BL/6J animals. Data points represent individual mice; bars are mean \pm SEM. P values [with 95% CI] are MTO34 (n=10 mice) vs MTO48 (n=9): 0.0325 [-2, 0], MTO48 vs MTO138 (n=10): 0.0305 [-108, 0]; MTO34 vs MTO138: 0.0031 [-108, 0]; MTO68 (11) vs MTO129 (n=8): 3.2×10^{-4} [-207, -30]; MTO93 (n=10) vs MTO94 (n=10): 1.1×10^{-4} [-60, -19]; MTO93 vs MTO140 (n=9): 2.2×10^{-5} [-149, -81]; MTO93 vs MTO163 (n=8): 9.1×10^{-5} [-220, -31]; two-sided MWW tests. **f)** Number of liver nodules (mean \pm SEM) 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 intervals are [-223, 0] (MTO129) and [-805, -66] (MTO138); two-sided MWW tests. **g)** Number of liver nodules (mean \pm SEM) of intrasplenically injected 3x and 4x MTOs, n=22 mice (MTO54), 20 (MTO220), 5 (MTO221), 7 (MTO193), 19 (MTO207), 5 (MTO219), 14 (MTO34), 16 (MTO68), 29 (MTO93). Right: the distribution of tumour diameters (right).

Figure 7. Source of tumoral TGF- β and pathway blockade *in vivo*. **a-b)** *TGFB* mRNA expression levels in sorted cell populations from disaggregated human CRCs (**a**) or mouse liver metastasis (**b**). (Epithelial) cancer cells, immune cells, endothelial cells and cancer-associated fibroblasts (CAFs) were sorted with labelled antibodies against Epcam, CD45 (n=14 patients for each), CD31 and FAP (n=6 for each) for

human and PDGFR-beta for mouse, respectively. RNA from sorted cells was analysed by microarray (human) or RT-qPCR (mouse). Shown in **(a)** are standardized expression data, depicted as Z-score: individual values and Tukey boxplots; P values are for *TGFB1*: Epithelial vs Leukocytes 7.5×10^{-6} , Epithelial vs Endothelial 6.2×10^{-4} , Epithelial vs CAFs 6.2×10^{-4} , and the Kruskal-Wallis (KW) P value is 6.3×10^{-6} ; for *TGFB2*: CAFs vs Epithelial 6.2×10^{-4} , CAFs vs Leukocytes 8.4×10^{-4} , CAFs vs Endothelial 0.0051, and the KW P value is 0.0017; for *TGFB3*: Epithelial vs CAFs 6.2×10^{-4} , Epithelial vs Leukocyte 0.0409, Epithelial vs Endothelial 0.0150, Leukocyte vs CAFs 6.2×10^{-4} , Endothelial vs CAFS 0.0051. The KW P value is 8.5×10^{-5} . Direct comparisons by two-sided MWW tests. In **(b)**, data are shown as mRNA expression levels normalized to epithelial cells, mean \pm SEM from technical triplicates, *Tgfb2* was undetectable (ND) in leukocytes. Note that in both settings, *TGFB1* is expressed by all stromal cell types, *TGFB2* and -3 mainly by CAFs. Expression in epithelial cancer cells is comparatively low. **(c-h)** Immunohistochemistry for TGF-beta target gene products in the tumour microenvironment (TME) of MTO138-derived liver metastases, representing 2 independent experiments. (Met, **c-c'**, **e-e'**, **g-g'**), as well as for the inhibition of stromal TGF-beta signalling by Galunisertib (**d**, **f**, **h**). **(c-d)** Micrographs of phospho-SMAD2 stained sections of liver metastases. **(e-f)** Staining reveals TGF-beta target CALD1 expression exclusively in the metastatic TME. Treatment with Galunisertib for 3 days abrogates expression of this cytoplasmic protein. **(g-h)** Staining for TGF-beta target IGFBP7, exclusively expressed in the metastatic TME. Treatment with Galunisertib for 3 days abrogates expression of this secreted protein. *Scale bars: 100 μ m.*

Figure 8. Characterization of MTOs in metastatic experiments. a) Representative examples of GFP positivity measured by flow cytometry indicating genetic recombination in the UbC-CreER^{T2}; *Tgfb2*^{fl/fl} genetic model recombination (representative of 3 independent experiments). **(b)** Liver metastases counted 5 weeks after intrasplenic injection of LAKP MTO54 or MTO220 and treatment from d2-d14 with Galunisertib (Gal, 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 \pm SEM, 95% CI for MTO220 is [-3, 0], from a two-sided MWW test. **(c)** Images of stromal TGF-beta activation markers in LAKP MTO liver tumours, representative of 2 independent experiments. **(d)** Number of nodules counted 5 weeks after injection of LAKP-eg or LAKP+S MTOs and treatment with Gal (red) or control (blue), mean \pm SEM, 95% CI for LAKP+S is [-554, -176], n=5 mice per condition; two-sided MWW test. **(e)** Number of liver nodules (mean \pm SEM) 4 weeks after intrasplenic injection of MTOs (either unlabelled or with mCherry-

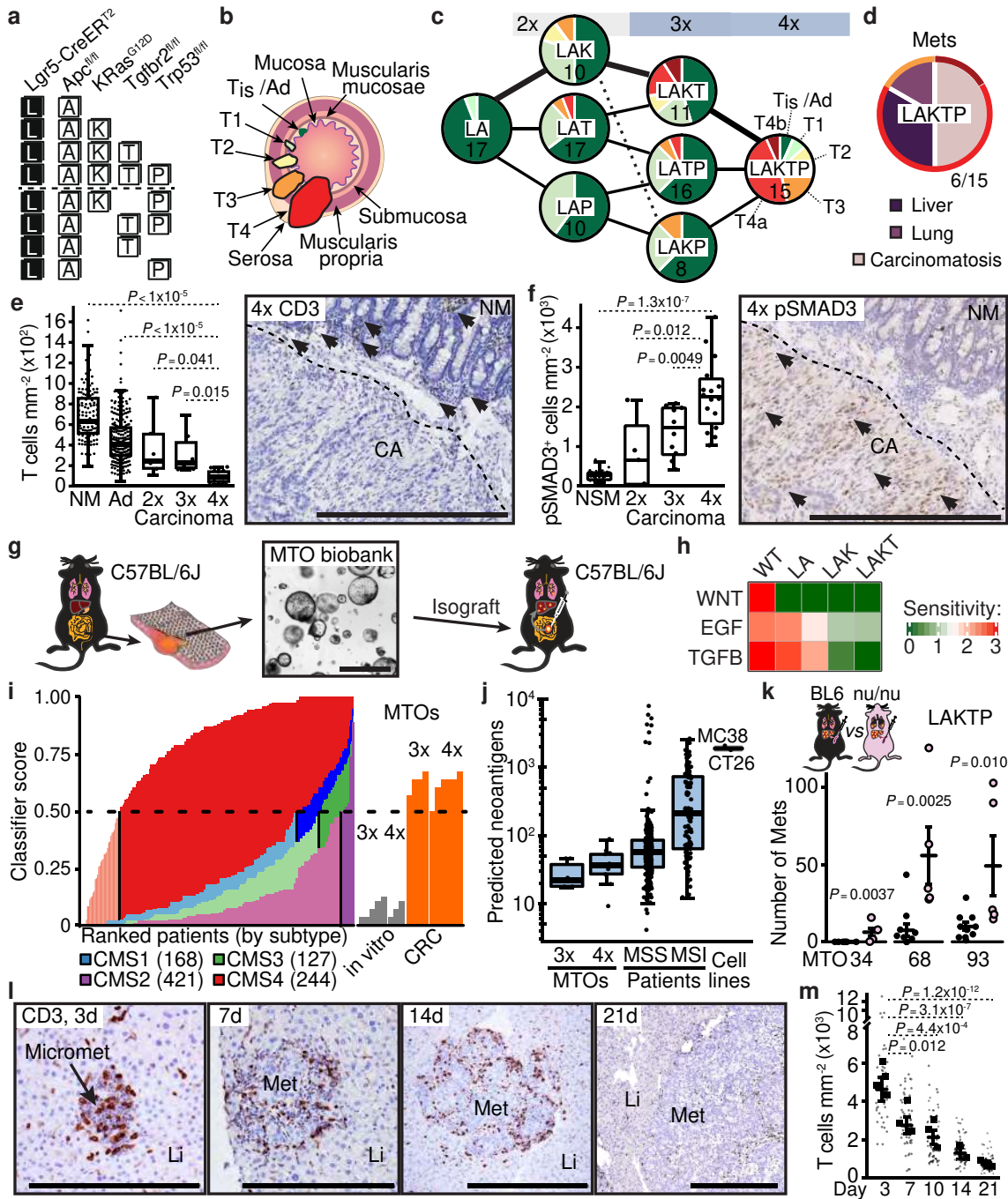
luciferase vector, or after CRISPR-KO of the mCherry-LUC coding region for MTO138), after treatment. 95% CIs are for MTO93: [-113, 0], (n=5 mice per condition); MTO93-LUC: [-104, -4], (n=5 control and n=4 Gal); MTO140: [-297, -21], (n=7 control and n=5 Gal); MTO140-LUC: [-251, -21], (n=5 per condition); MTO138-LUC: [-86, -8], (n= 5 per condition); MTO138-KO-LUC: [-70, -13], (n=7 per condition); two-sided MWW tests. **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 Gal 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% CI for the difference at day 25: [-222.7, -23.9], two-sided MWW test. **g**) BLI (as in **f**) of a metastasis initiation experiment with MTO138 in C57BL/6J mice, n=24 control mice, n=15 mice treated with Gal. **h**) BLI (as in **f**) of Gal 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 ± SEM) counted in mice intrasplenically injected with MTO129 and treated with Gal alone or with antibodies against CD4 or IgG control; 95% CI for Gal vs Control is [-79, -8], n=5 mice per condition; two-sided MWW test. **j**) Number of liver nodules (mean ± SEM) after intrasplenic injection of MTO93 in syngeneic, *Lgr5-eGFP-CreER^{T2}* mice—from the colonies that gave rise to the 3x and 4x genetic models—treated with Gal (red) or control (blue). 95% CI is [-13, -1], n=20 control mice and n=25 for Gal; two-sided MWW test. *Scale bars: 500 μm.*

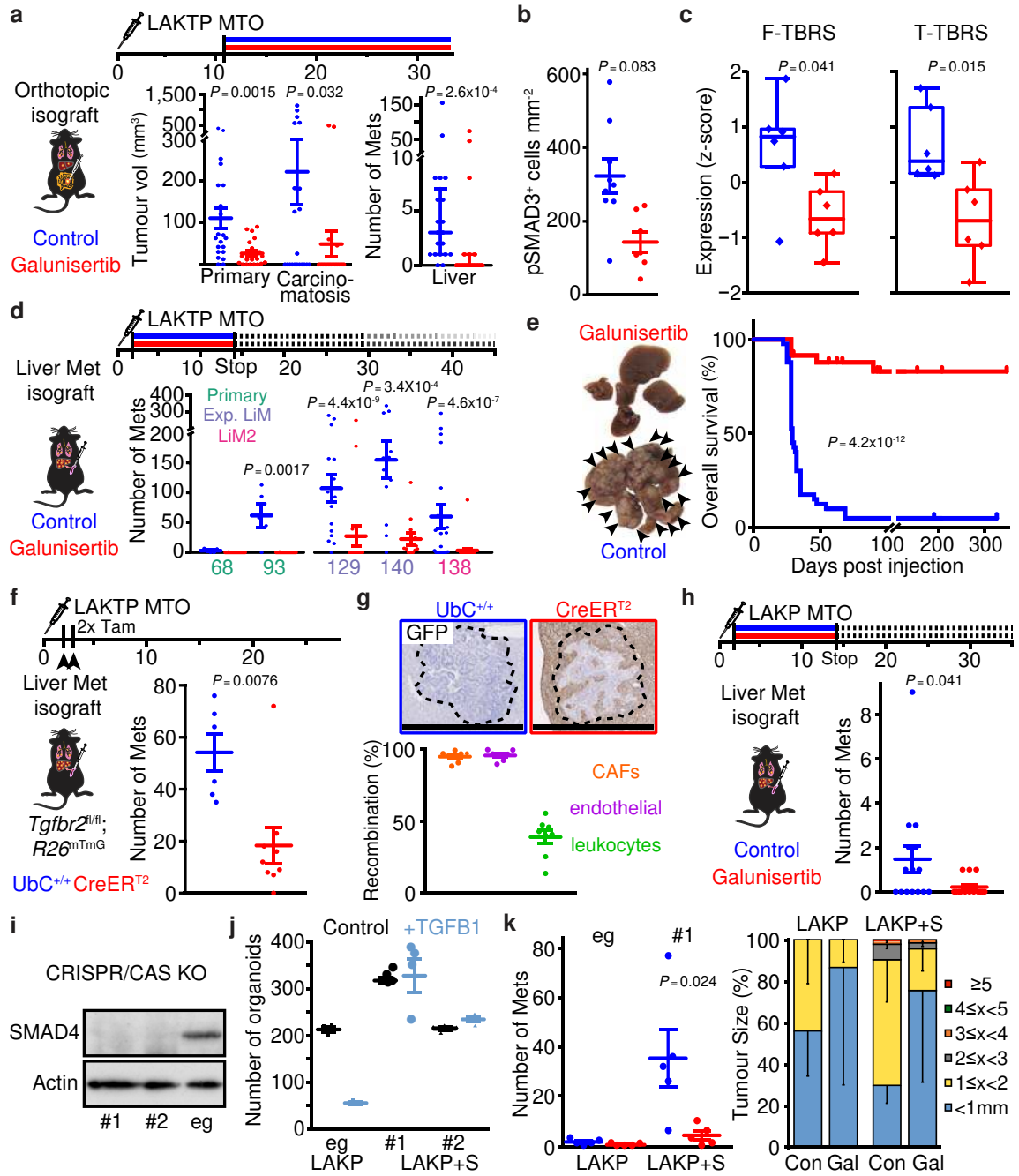
Figure 9. Immune modulation by TGF-beta in CRC. A human meta-cohort was used to analyse the interaction of TGF-beta expression levels and T cell activation state. **a**) Expression signature ratio between active CD4 Th1 and naive CD4 T cells differs significantly between microsatellite-stable (MSS) and -unstable (MSI) patients. Standardized expression for individual patients as well as Tukey boxplots, n=981 samples for MSS patients and n=198 for MSI; two-sided Wald test. **b-c**) Expression signature for naive CD4 T cells (**b**) and Th1-differentiated CD4 T cells (**c**) in patients with low, medium and high *TGFB1+2+3* expression. Standardized expression for individual patients as well as Tukey boxplots. P values are for (**b**): Low vs Medium (n=327): 1.7x10⁻⁴, Low vs High: 2.2x10⁻¹⁶, Medium vs High: 2.2x10⁻¹⁶; and for (**c**): Low vs Medium: 0.0503, Low vs High: 6.8x10⁻¹⁴, Medium vs High: 8.1x10⁻⁸; n=327 patient samples per group; two-sided Wald tests. The Likelihood Ratio Test (LRT) P value is 2.2x10⁻¹⁶ for (**b**) and 2.3x10⁻¹⁴ for (**c**). **(d)** *TGFB1+2+3* mRNA expression levels predict

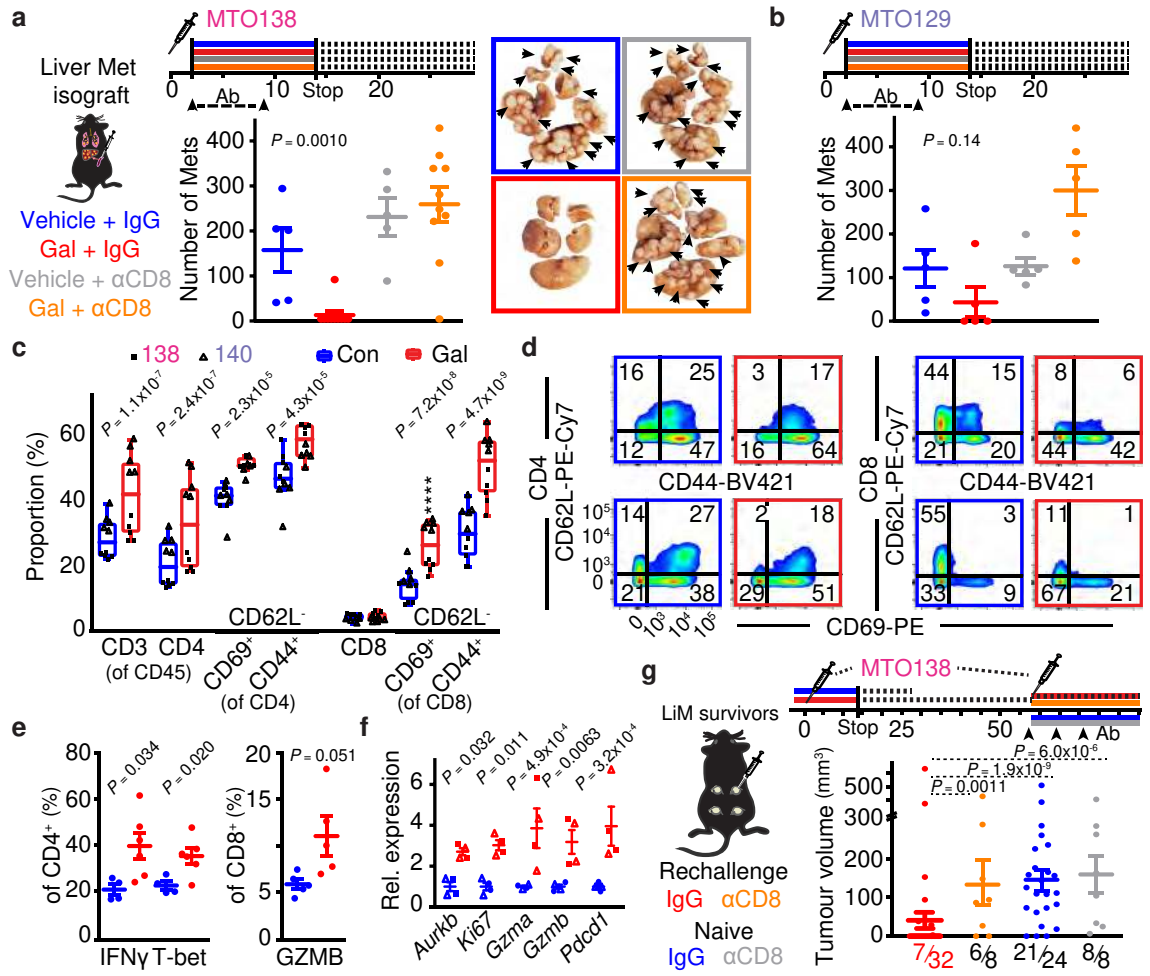
poor prognosis. Hazard risk (HR) 95% CIs are Low vs Medium: [0.65, 1.42]; Low vs High: [1.08, 2.20]; Medium vs High: [1.12, 2.32]; n=255 patient samples (Low), n=254 (Medium) and n=249 (High); two-sided Wald tests. **e-f**) Correlation between ratio of CD4-Th1 vs naïve CD4 signatures and *TGFB1+2+3* 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 tests. **g**) Kaplan-Meier for relapse free survival for patients according CD4 Th1-naïve signatures ratio. HR 95% CIs are: Low vs Medium [0.40, 0.81], Low vs High [0.37, 0.78], Medium vs High [0.64, 1.41]; n=243 (Low), n=259 (Medium) and n=257 (High); two-sided Wald tests. **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 \pm SEM derived from a mixed-effects linear model; n=4 mice per condition (MTO129) and n=6 for MTO140. **i**) Representative images of IHC of MTO140 micrometastases, quantified in (**h**). Mean cell densities (cells mm⁻²) are indicated. Images represent 2 independent experiments. *Scale bars: 100 μ m.*

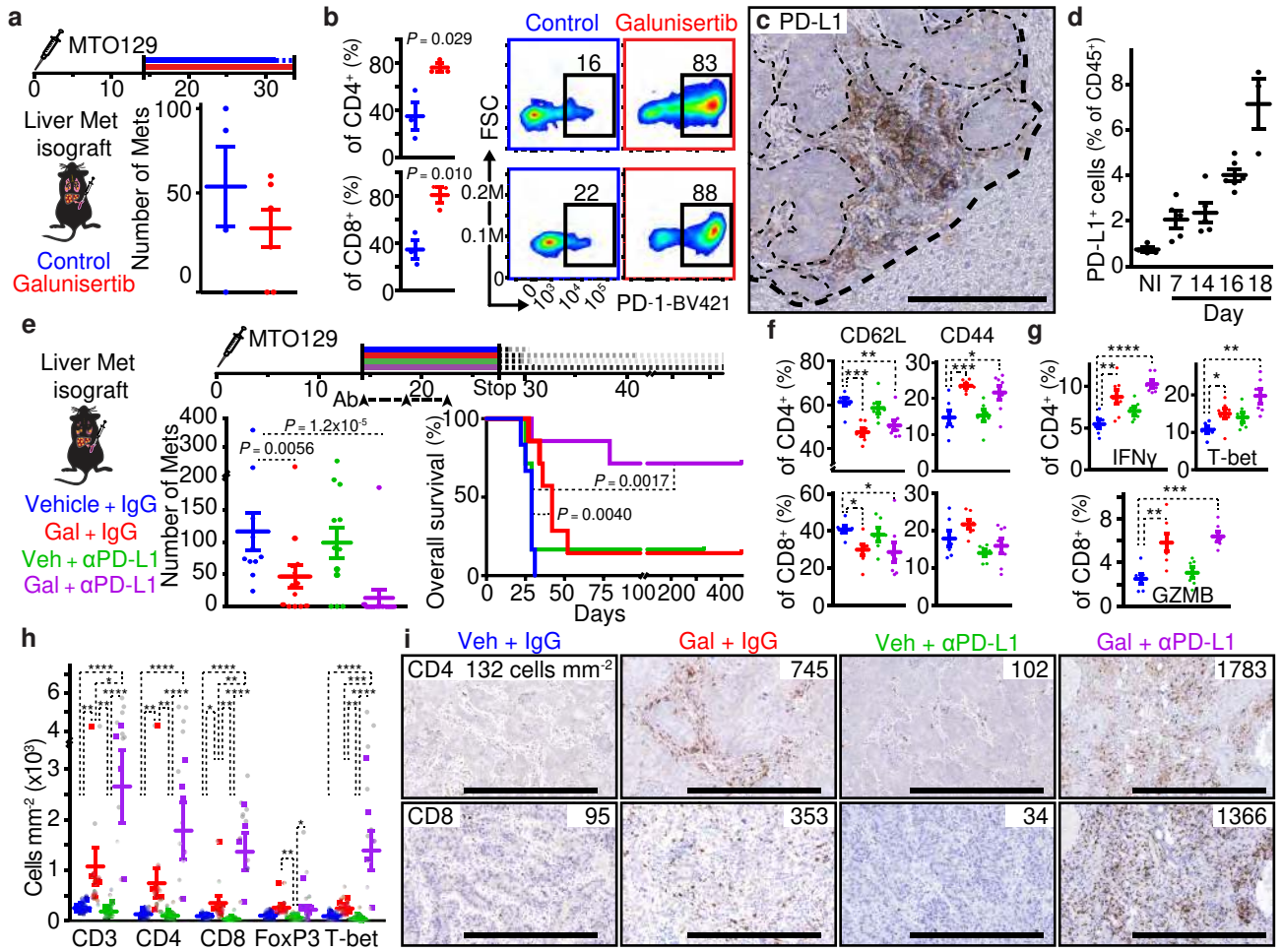
Figure 10. TGF-beta inhibition and the tumour immune microenvironment.

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; Tgfb2^{fl/fl}* background, compared to *Ubc^{+/+}* control (**c**), or 2 days after start of treatment in established liver metastases (which started at day 14), (**d**). Shown are individual tumours (grey circles, **a, c-d**), mouse means (squares, or triangles for MTO93, **a**), and group averages \pm SEM. Data are analysed with a mixed-effects linear model; for (**a**) n=5 mice (Con) or n=4 (Gal) per condition; for (**b**); n=9 mice (Con) or n=8 (Gal) per condition; for (**c**); n=5 mice (*Ubc^{+/+}*) or n=7 (*Ubc-CreERT²*) per condition; and for (**d**); n=6 mice per condition. **e**) Distribution of cell types within the PD-L1+ population (mean \pm SEM) in microdissected tumours (sacrificed at day 18, 4 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 \pm SEM percentage of n=3 mice.

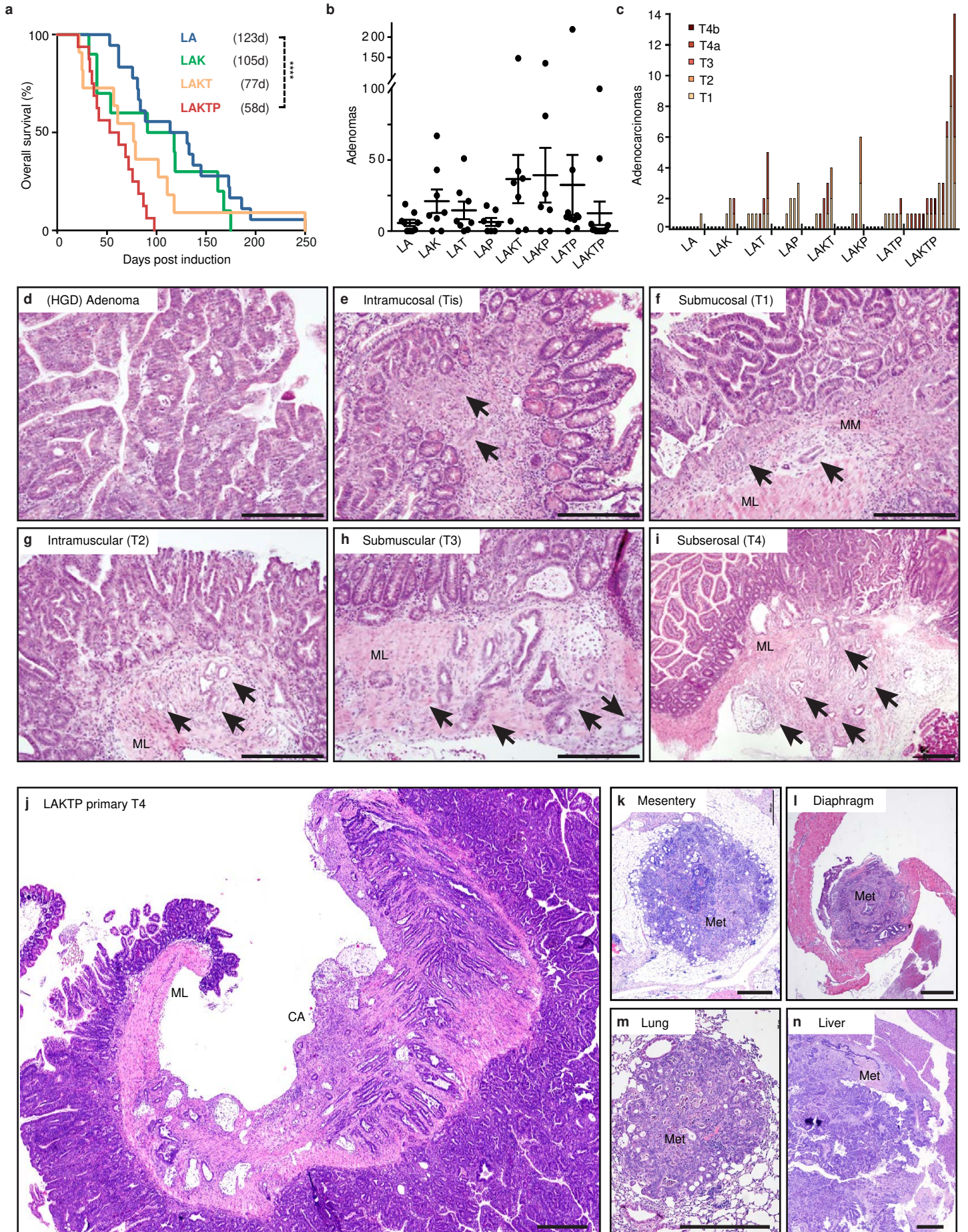




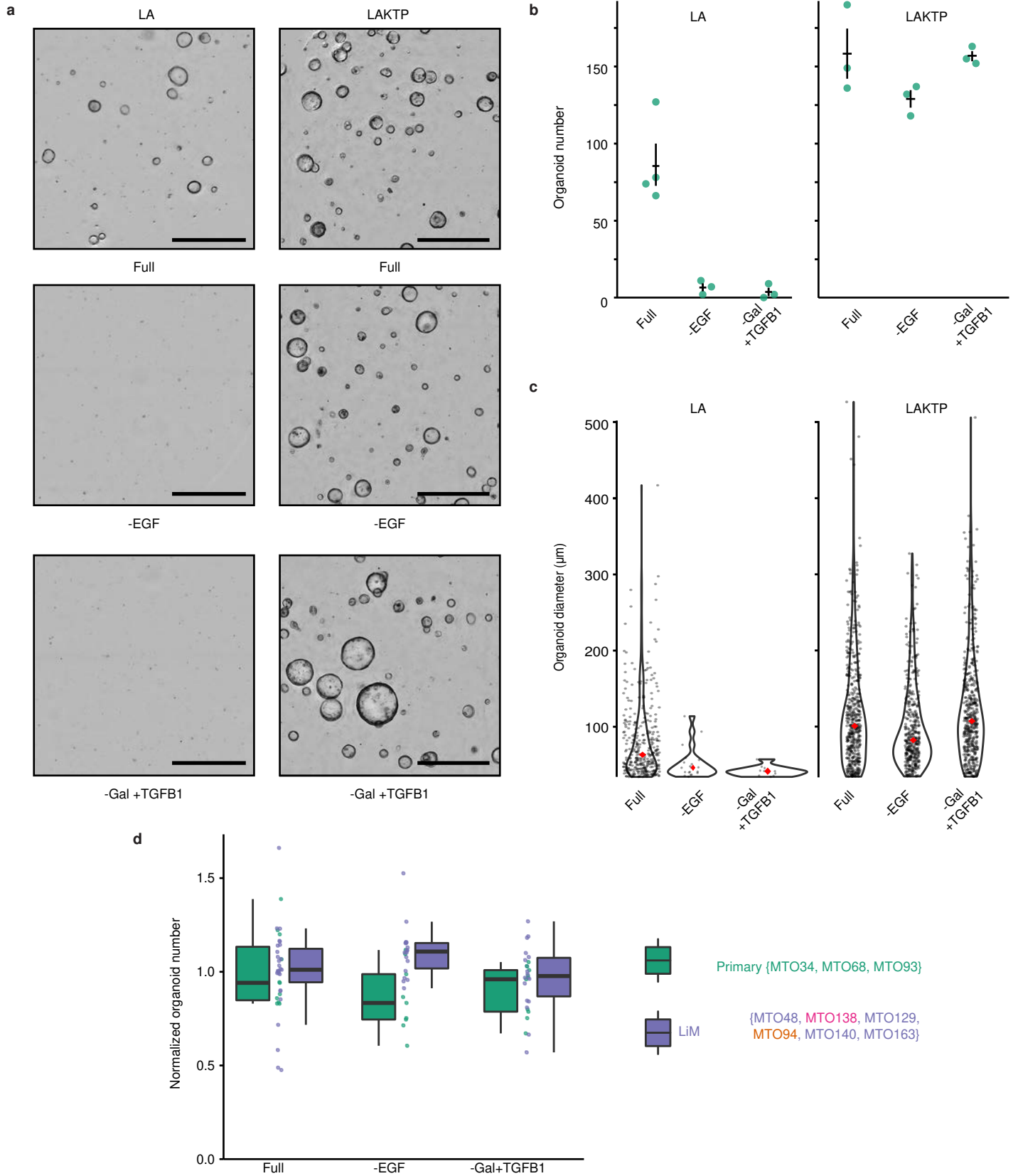




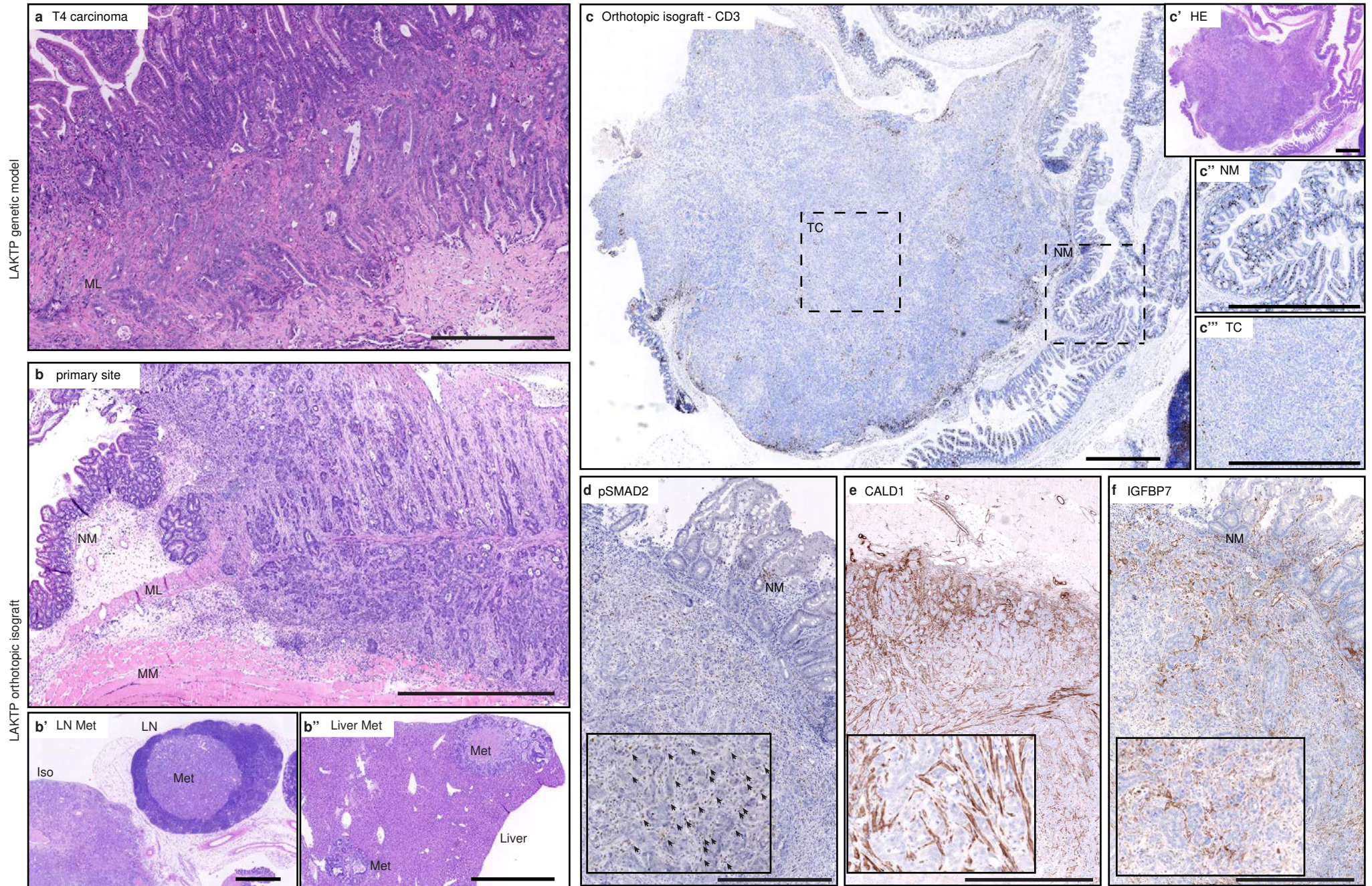
Extended Data Figure 1: Analysis of genetic models



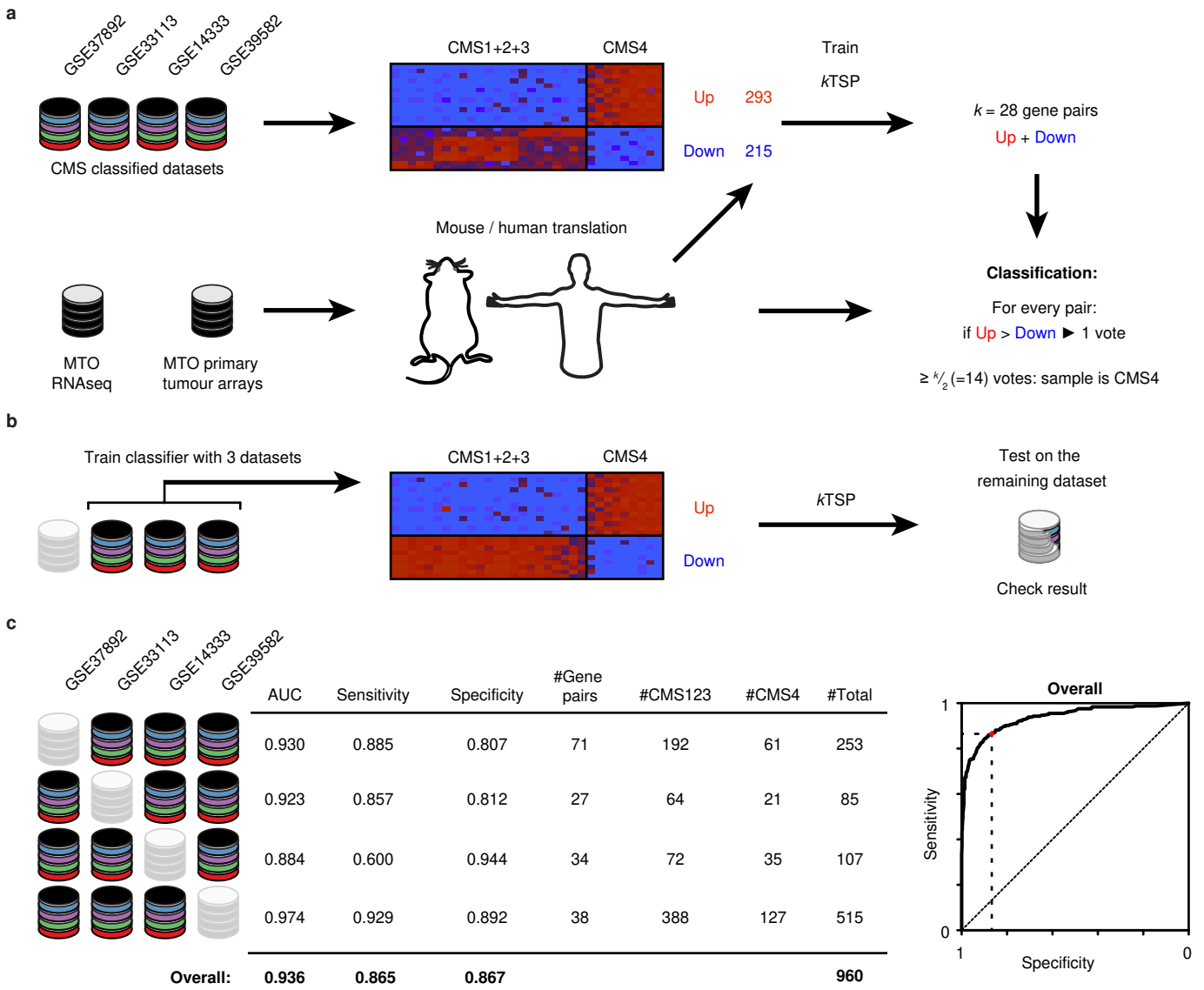
Extended Data Figure 3: Functional analysis of mouse tumour organoids *in vitro*



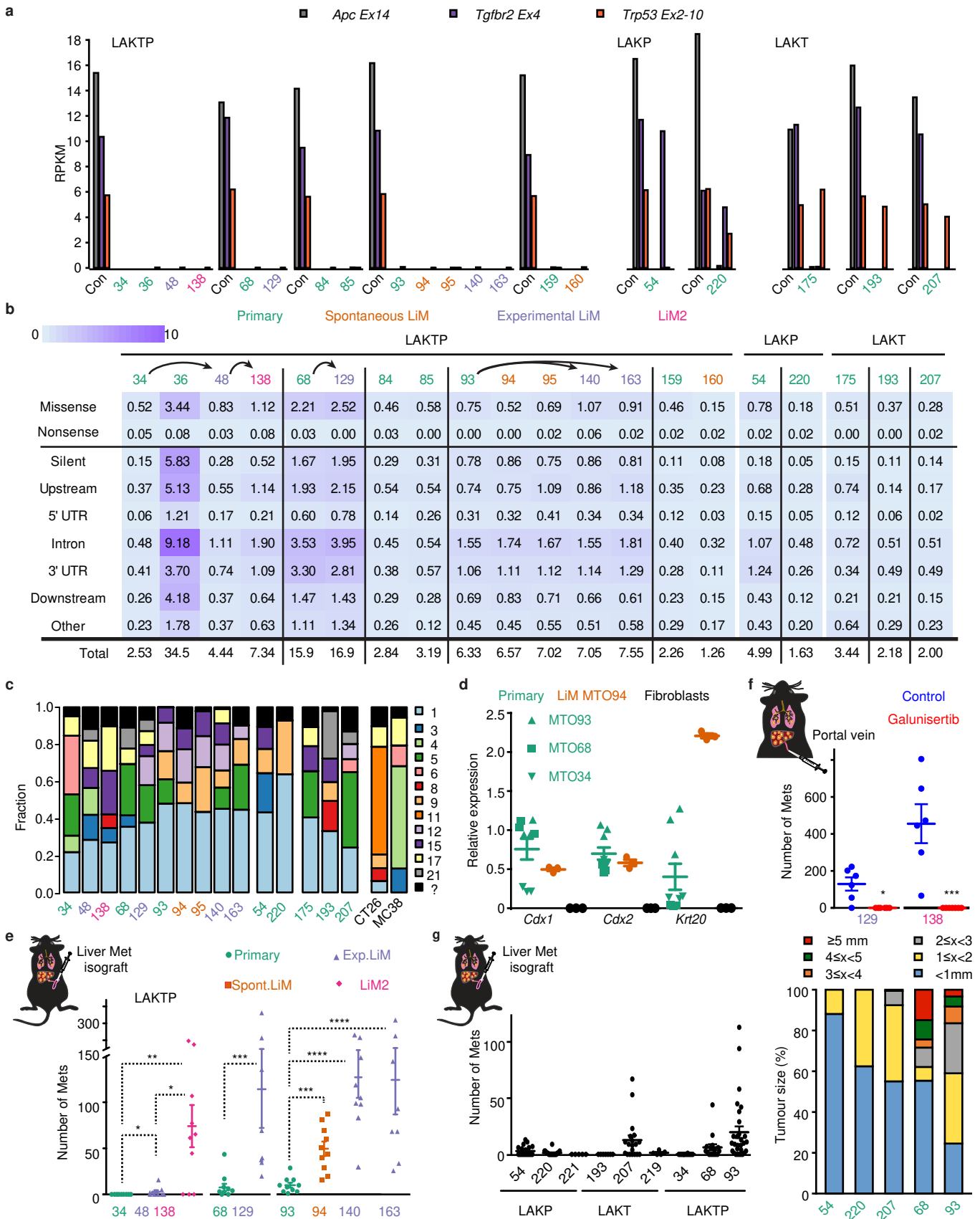
Extended Data Figure 4: Histology and TME of orthotopic isograft-derived primary tumours (LAKTP)



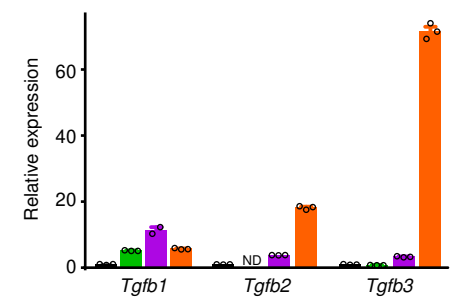
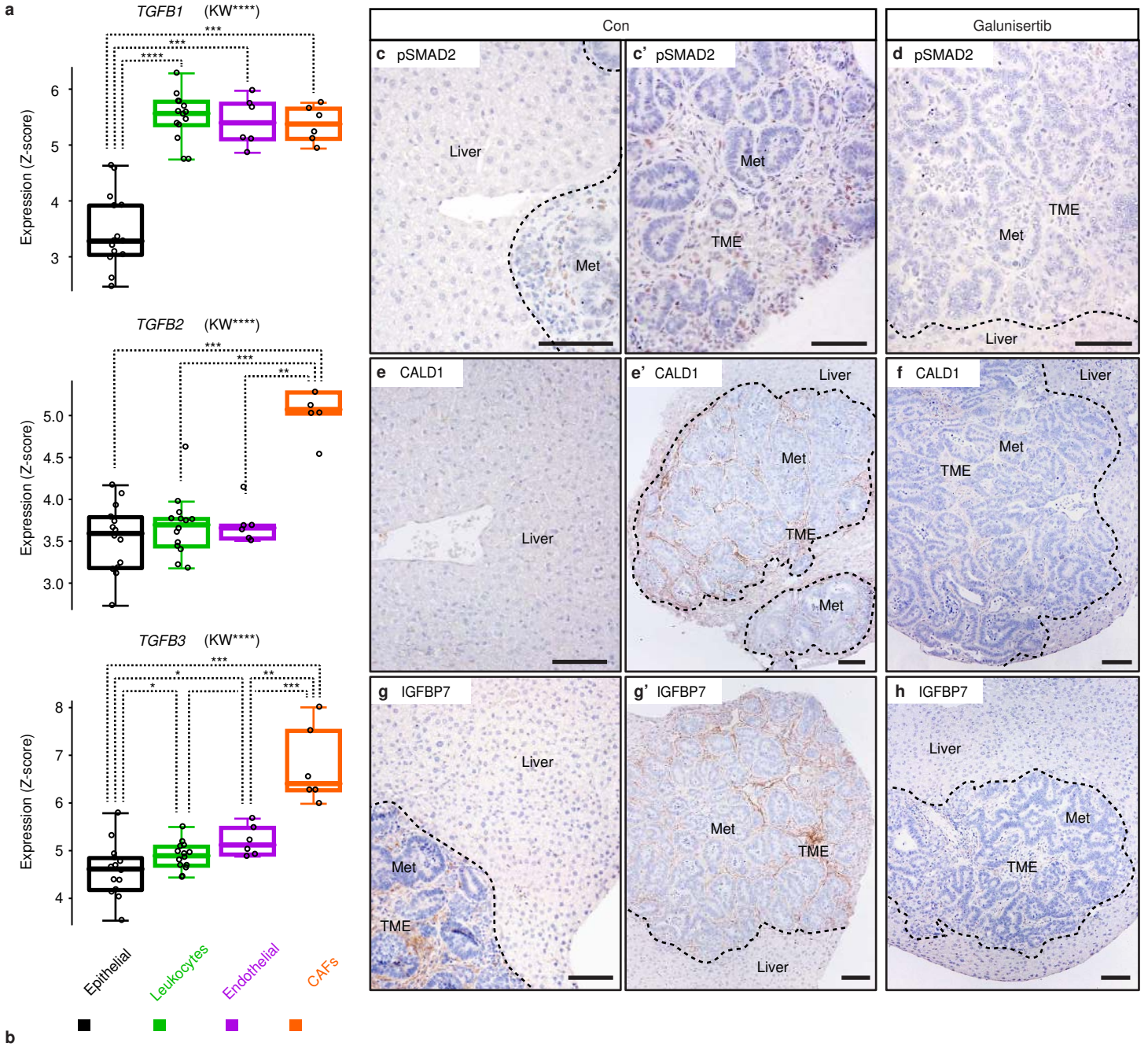
Extended Data Figure 5: CMS4 classifier using *k*-Top Scoring Pairs



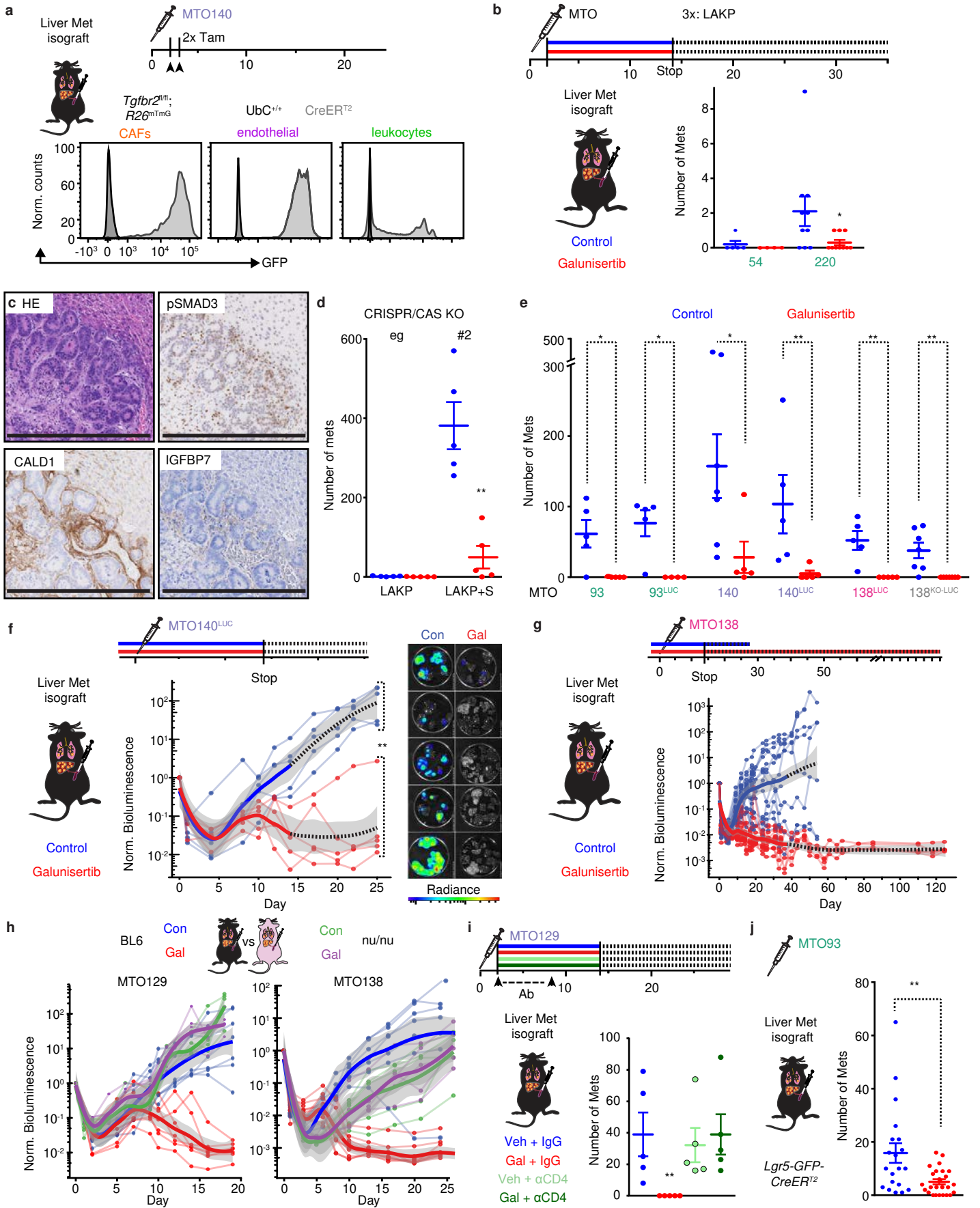
Extended Data Figure 6: Genetic analysis of mouse tumour organoids



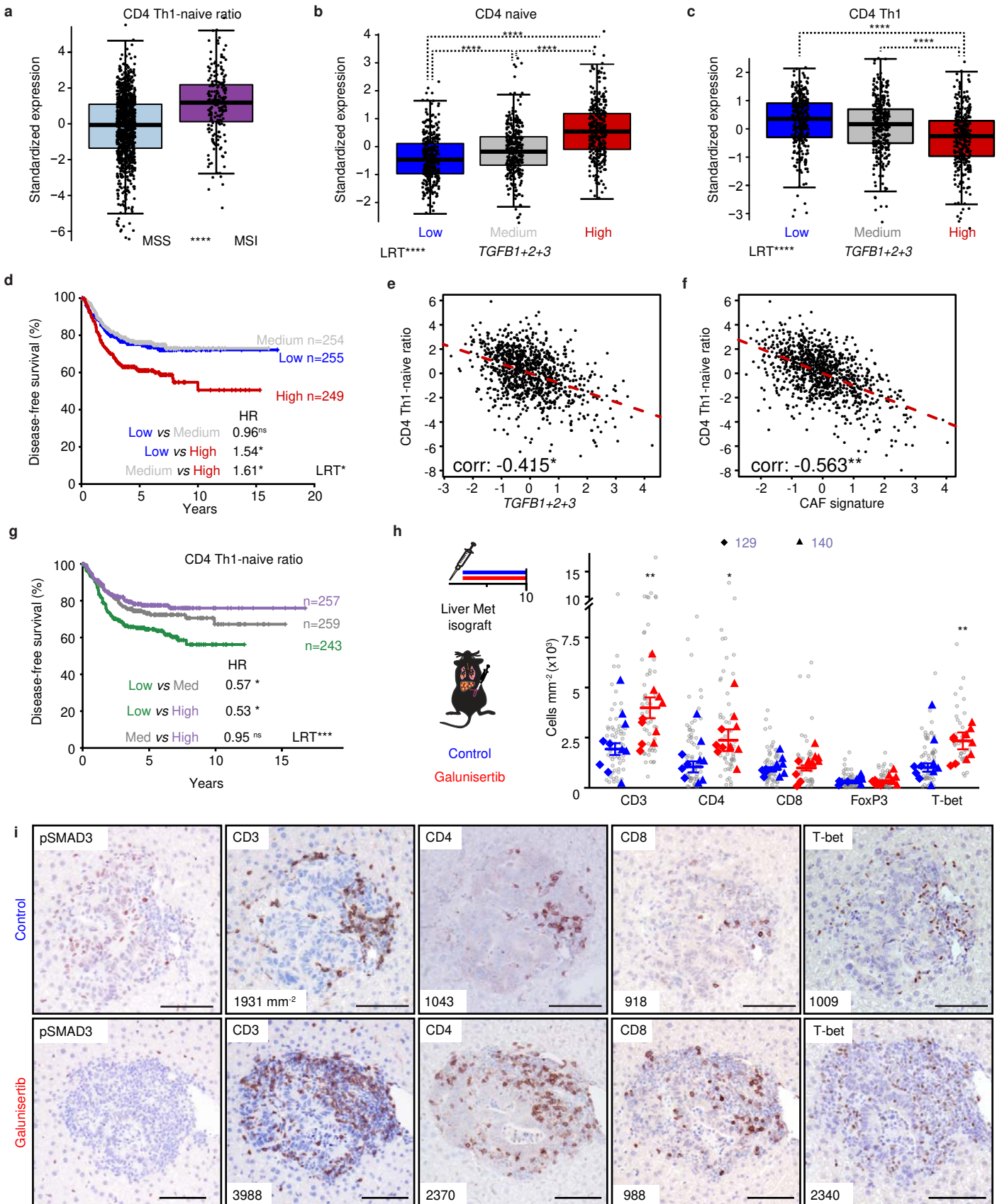
Extended Data Figure 7: Source of tumoral TGF- β and pathway blockade *in vivo*



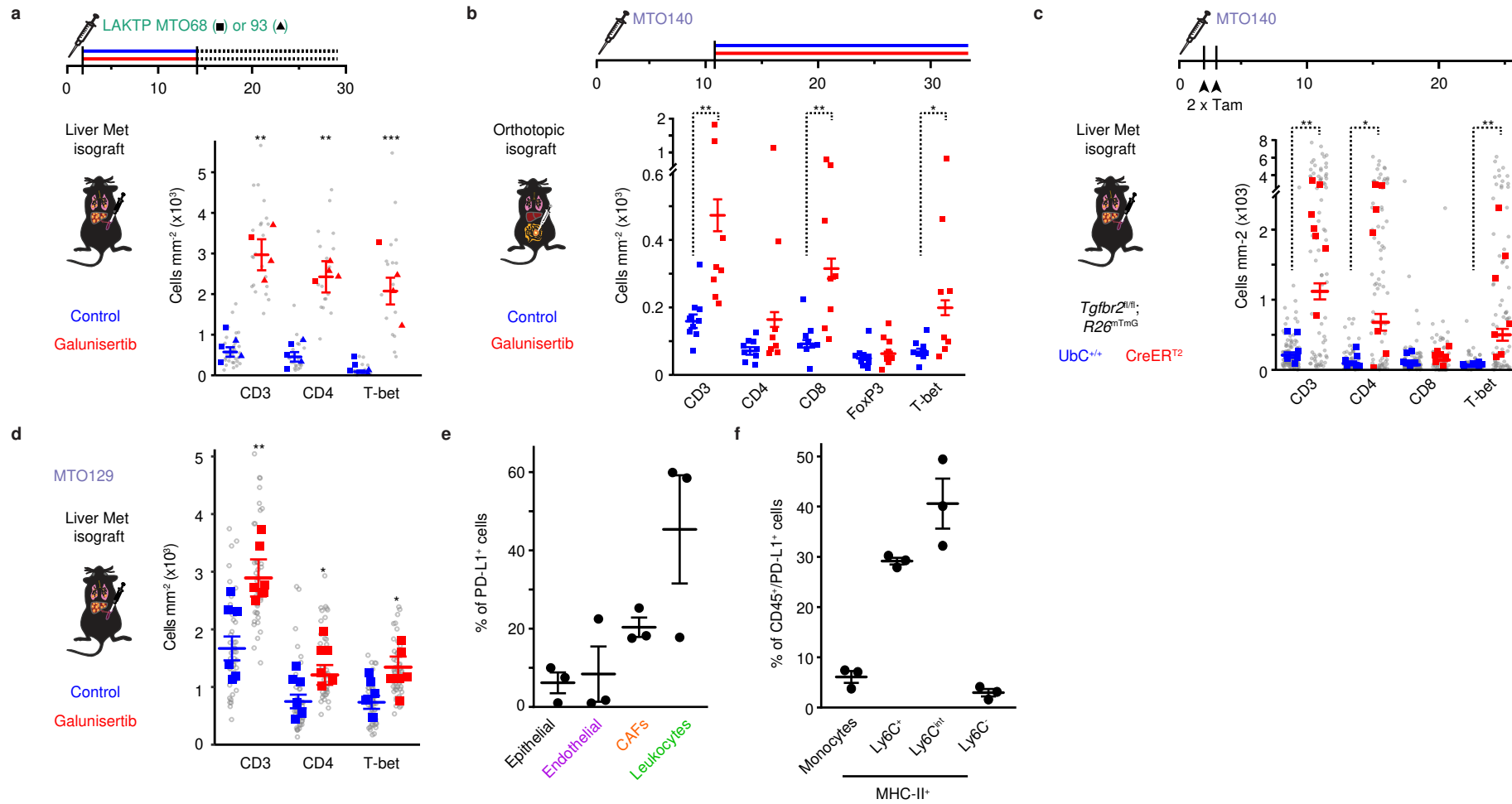
Extended Data Figure 8: Characterization of MTOs in metastatic experiments



Extended Data Figure 9: TGF-beta and immune activation



Extended Data Figure 10: TGF-beta inhibition and the tumour immune microenvironment



Tauriello et al.

TGF-beta drives immune evasion in genetically reconstituted colon cancer metastasis

Supplementary Information

Contents

Source data westernblots

Supplementary Discussion

Supplementary Discussion References

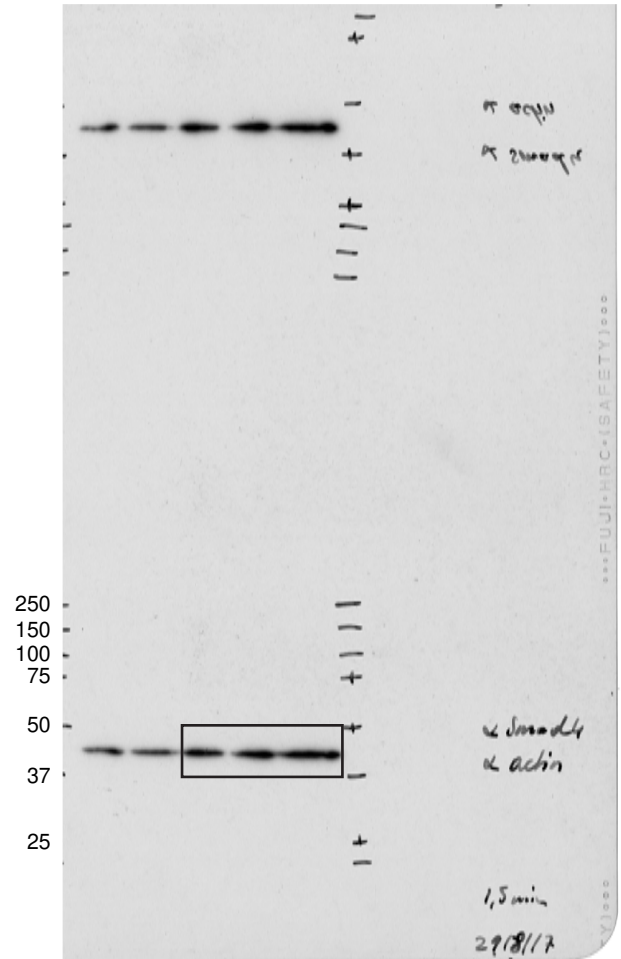
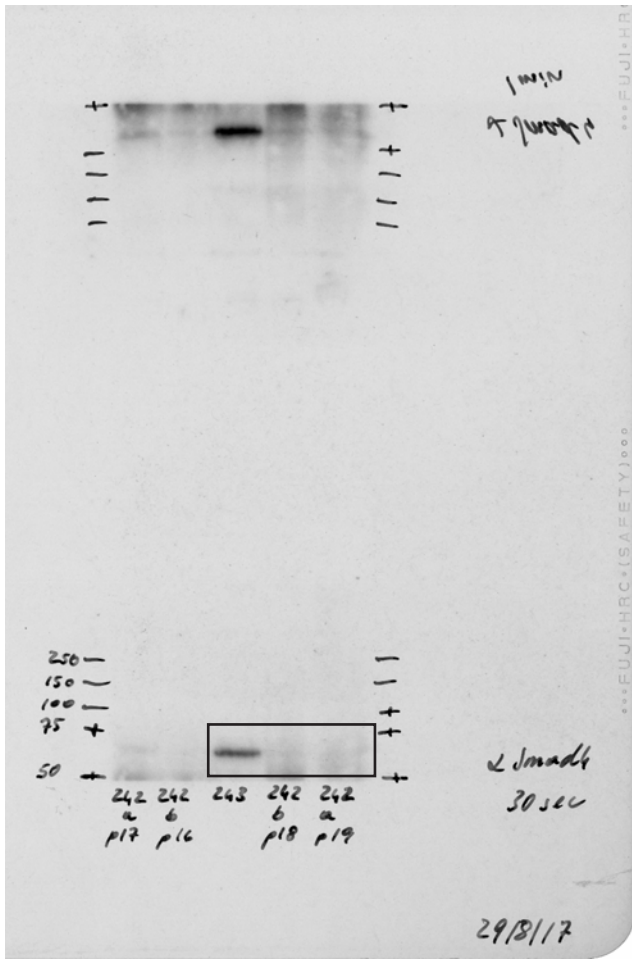
Supplementary Acknowledgements

Supplementary Methods

Supplementary Methods References

Supplementary Fig. 1 - Source data westernblots

Fig. 2i raw images



Supplementary Discussion

The mutant mice and MTOs described herein are models to study poor prognosis human MSS CRC. Mutations in the four driver pathways render tumour cells independent of ISC niche factors, a feature that enables autonomous growth in unfavourable environments such those encountered by disseminated tumour cells in foreign organs as previously observed¹⁻⁴. We found that malignancy co-evolves with changes in the TME reminiscent of those associated to poor prognosis CRCs such as T cell exclusion^{5,6} and prominent recruitment of TGF-beta activated CAFs^{7,8,9}. The finding that the penetrance of metastasis is ultimately modulated by the TME supports previous studies indicating that features of the immune system and the tumour stroma can help prognosticate the evolution of the disease in patients^{5,7,9,10}. Recent consensus molecular classifications of CRC revealed the existence of a poor prognosis patient subset - CMS4 - characterized by prominent desmoplastic reaction, high TGF-beta levels and expression of signatures that indicate immunosuppression^{8,11}. We show triple and quadruple mutant MTOs, when grown orthotopically, can be classified as CMS4 CRCs. It is, however, important to consider that frequency of genetic alterations in the main driver genes does not substantially differ between CMS2, CMS3 and CMS4 subtypes (the three subtypes that encompass most of the MSS patients)⁸, implying that genotype – or at least mutations in the key oncogenes and tumor suppressor that modulate CRC progression - does not completely define molecular subtypes. It remains vital to characterize which other features and processes drive the acquisition of a TGF-beta activated TME.

It has been proposed that MSS CRCs are immunologically cold due to low mutational burden and therefore that they are unlikely to benefit from immune therapies¹². In contrast, we found that metastatic cells from this class of CRCs can be effectively killed by the adaptive immune system through a CTL-dependent process, which CRC cells avert by raising TGF-beta levels. This finding provides a rationale for the overarching association between TGF-beta-driven gene programmes in the TME and poor prognosis in CRC^{5,7,8,10,13}. We described two TGF-beta-driven mechanisms that cooperatively dampen immune responses. First, TGF-beta impacts directly on the capacity of T cells to mount a robust adaptive anti-tumor response. We observed a tight association between lack of T cell differentiation towards a Th1 effector phenotype and TGF-beta levels in CRCs. In line with this finding, mice deficient for *Tgfb2* in T cells develop an early onset autoimmune syndrome characterized by premature differentiation of naïve CD4⁺ T helper cell to a Th1 phenotype^{14,15}. Mice that express a dominant negative *Tgfb2* in T cells showed improved T-cell responses against melanoma and thymoma cell lines¹⁶. The expression of T-bet, the master transcription factor to drive the Th1 effector phenotype, is directly repressed by TGF-beta signalling¹⁷. TGF-beta signalling also inhibits CD8⁺ T responses to T cell receptor stimuli¹⁸⁻²⁰. In addition, the

immunosuppressive TME imposed by TGF-beta is likely the result of a coordinated effect on multiple immune and non-immune cell types¹⁰.

Second, our data also reveal that a TGF-beta-activated TME excludes T cells from tumours, a phenomenon that characterizes MSS CRCs as well as many other cancer types^{5,21,22}. The mechanisms of immune exclusion remains poorly understood but it may occur through direct action of TGF-beta on T cells or/and indirectly through regulation of certain cytokines and extracellular matrix proteins by TME components such as CAFs and endothelial cells^{23,24}. Nevertheless, facilitating T cell infiltration through TGF-beta inhibition appears a pre-requisite for enabling anti-tumour responses by the adaptive immune system. In the particular case of CRC, blockade of immune exclusion by Galunisertib is sufficient to confer susceptibility to PD1/PDL1 checkpoint-based therapies, a strategy that may have broad application for treatment of multiple cancer types that thrive in a TGF-beta rich environment. Our findings can also be interpreted such that TGF-beta-mediated immune evasion operates hierarchically upstream of the PD1/PD-L1 inhibitory checkpoint.

Whereas future studies should be aimed at dissecting the relevance of each of these responses for CRC immune evasion, our results immediately suggest the use of TGF-beta signalling inhibitors as immunotherapy (possibly in combination with other immunotherapies²⁵) for patients with metastatic CRC and in particular with poor prognosis CMS4 tumours. The tumour suppressor role of TGF-beta in epithelial cancer cells has warned against the use of TGF-beta inhibitors for cancer therapy¹⁰. Yet, we show that mice bearing metastases with an intact TGF-beta pathway are eradicated by TGF-beta inhibition implying that benefits of unleashing the immune system prevails in this setting and that TGF-beta based-immunotherapies may be safe in a broad range of CRC patients. Finally, although our experiments demonstrate a dependency on a TGF-beta-activated TME during metastasis formation, it remains to be proven that these results truly reflect therapeutic potential to eliminate residual CRC cells present in AJCC Stage II and Stage III CRC patients after surgical removal of the primary tumour. At present there is lack of CRC models that faithfully reproduce disease relapse after therapy.

Supplementary Discussion References

1. 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).
2. Matano, M. *et al.* Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat. Med.* **21**, 256–62 (2015).
3. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).

4. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc. Natl. Acad. Sci.* **114**, E2357–E2364 (2017).
5. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–4 (2006).
6. Mlecnik, B. *et al.* The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8**, 327ra26 (2016).
7. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320–329 (2015).
8. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
9. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–9 (2015).
10. Tauriello, D. V. F. & Batlle, E. Targeting the Microenvironment in Advanced Colorectal Cancer. *Trends in Cancer* **2**, 495–504 (2016).
11. Becht, E. *et al.* Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin. Cancer Res.* **22**, 4057–4066 (2016).
12. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
13. 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).
14. Gorelik, L. & Flavell, R. A. Abrogation of TGF β signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12**, 171–81 (2000).
15. 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-beta Receptor. *Immunity* **25**, 441–454 (2006).
16. Gorelik, L. & Flavell, R. A. Immune-mediated eradication of tumors through the blockade of transforming growth factor-beta signaling in T cells. *Nat. Med.* **7**, 1118–1122 (2001).
17. Gorelik, L., Constant, S. & Flavell, R. A. Mechanism of transforming growth factor beta-induced inhibition of T helper type 1 differentiation. *J. Exp. Med.* **195**, 1499–505 (2002).
18. Thomas, D. A. & Massague, J. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* **8**, 369–380 (2005).
19. Mehal, W. Z., Sheikh, S. Z., Gorelik, L. & Flavell, R. A. TGF-beta signaling regulates CD8+ T cell responses to high- and low-affinity TCR interactions. *Int. Immunol.* **17**, 531–538 (2005).
20. Zhang, N. & Bevan, M. J. TGF-beta signaling to T cells inhibits autoimmunity during

lymphopenia-driven proliferation. *Nat Immunol* **13**, 667–673 (2012).

21. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).

22. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science (80-.)*. **348**, 74–80 (2015).

23. Kraman, M. *et al.* Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein-alpha. *Science (80-.)*. **330**, 827–830 (2010).

24. Feig, C. *et al.* Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20212–7 (2013).

25. Tauriello, D. V. F., Calon, A., Lonardo, E. & Batlle, E. Determinants of metastatic competency in colorectal cancer. *Molecular Oncology* **11**, 97–119 (2017).

Supplementary Acknowledgements

We thank Natalia Plana, Carolina Sanchez and Edgar Cristóbal Lecina for technical assistance, Francisco Barriga for materials, Raquel Batlle, Gonzalo Fernández Miranda and Raúl Méndez for mouse strains, Elena Sancho for help with the manuscript and important support, and Cristina Mendez for logistic support. From our outstanding facilities, we wish to thank from histopathology: Neus, Mònica, Begoña, Alicia, Anaïs and Antonia; from functional genomics: Nacho and David; from advanced digital microscopy: Sébastien and Anna; from flow cytometry: Jaume, Sonia, Chary and Ricardo; from the animal facilities: Steve, Barbara, Ferran, Conchi, Jesús, Rosa, Javi, Julia, Jordi, Vanessa and Fernando; and from the CRG genomic unit: Jochen, Anna and Maria.

Supplementary Methods

Cell lines

Murine cell lines MC38, CT26 and L-Wnt3a were obtained from ATCC and cultured in DMEM with 10% FBS. For MC38 and CT26, gDNA and mRNA was harvested from trypsinized, cultured cells and exome/RNA sequenced with the same protocols as for MTOs. Cell lines were regularly tested for mycoplasma contamination.

In vitro crypt niche factor dependency assays

Starting with full MTO-medium (above), we used combinations leaving out components, or adding 5 ng/ml recombinant human TGF- β 1 (Peprotech) while removing galunisertib (+TGFB1). Wildtype normal intestinal organoids were obtained from a naïve C57BL/6J mouse, using the crypt shaking method⁵, and cultured in MTO-medium supplemented with 1 μ g/ml RSPO-1 and 50% v/v Wnt3a-CM. MTOs were never cultured with either supplement and WT organoids did not survive in MTO-medium without them. RSPO-1 was produced in-house⁶ and Wnt3a-CM (from L-Wnt3a cells) was produced and tested according to previously described methods⁷.

For Fig. 1h: MTOs were seeded in standard 48^{well} plates in triplicate as single cells and growth/death was manually scored at day 2, 4 and 6. Pictures were taken on day 5. Sensitivity was defined as difference in growth/survival compared to the control. This was assessed both manually, where 3 was the maximum level (0 vs +++), or using automated imaging (Olympus CellR/ScanR multiwell plate scanner; individual wells were scanned in a 5x4x8 (x-y-z) matrix) followed by ImageJ⁸ analysis (Raw images were projected, stitched and analysed for automated organoid counting by ImageJ macros written by Anna LLado and Sébastien Tosi (IRB Barcelona ADM core facility); the result is shown in Extended Data Fig. 3a). Although the analyses were in high concordance, the heatmap is composed of data from the former. For each genotype, between 3 and 7 MTOs were analysed, p-values were calculated using the 0-3 sensitivity scores: for the LA-LAKTP EGF comparison $p=0.0001$, for the LA-LAKTP TGFB1 comparison: $p<0.0001$, and for the LAK-LAKTP TGFB1 comparison: $p=0.0010$ (two-sided MWW tests). Data were analysed and represented using R (v 3.4.2)⁹, RStudio (v 1.1.383) and the *ggplot2* package (v 2.2.1)¹⁰. For diameter calculations, organoid sphericity was assumed.

Whole exome sequencing

Genomic DNA was harvested from cultured MTOs (median passage 5), as well as from mouse tail fragments, and purified using the GenElute kit (Sigma). Library prep was done using NebNext Ultra and the exome selection was done using SeqCapEZ developer (4 reaction, 110624_MM9_exome_L2R_D02_EZ_Hx1; Cat. No. 06740278001, NimbleGen-Roche). Exome enrichment reaction was done in pools of 5 or 6 libraries. Sequencing was done HiSeq2500, Paired End, 125nts (2x125, v4); one lane per pool.

Reads preprocessing: Paired end reads were aligned to the mm10 version of the mouse genome using the mem algorithm implemented in the bwa software¹¹ (bwa-0.7.4) with default parameters. SAM files were sorted and indexed using Sambamba (v0.5.9; <http://lomereiter.github.io/sambamba/>) and duplicated reads removed with the Picard software (v1.128; <https://github.com/broadinstitute/picard>). Read qualities were recalibrated using the Genome Analysis Tool Kit (GATK, v3.5)¹².

Somatic SNP and INDEL calling: For each control-sample pair processed BAMs were compared using MuTect2 from GATK with default parameters. No extra filtering was applied other than the PASS/REJECT call from the Mutect2 algorithm. We decided to include

mutations marked as "clustered_events" since some positive controls (e.g. Kras (G12V) mutation in CT26) were otherwise excluded. Resulting VCF files were annotated with *snpEff* (v4.1)¹³. Somatic mutation load was computed with respect to the total length of the non-overlapping regions captured in the exome capture kit.

We constructed mutational signatures using the somatic mutations found in each of our samples. We normalized by the overall trinucleotide frequency in the mouse exome and compared against the human mutational signatures¹⁴ using the algorithm implemented in the deconstructSigs software (v1.8.0)¹⁵. We removed signature "1B" from the initial set since it has been reported to be a linear combination of signatures 1 and 5¹⁶.

RNA sequencing

mRNA was harvested from cultured MTOs and purified using the Ambion PureLink RNA mini kit (Life Technologies). Library prep was done using the Illumina kit Truseq Stranded mRNA Sample Prep kit. Sequencing was done on the HiSeq 2500, Paired End, 50nts (2x50, v4) with pools of 8-10 samples (one pool per lane). Paired end reads were aligned to the mm10 version of the mouse genome using the Star software (v2.3.0e)¹⁷. Expression was estimated using the R package *casper*¹⁸. The full expression matrix was normalized by quantiles.

Neoantigen prediction

Affinity prediction in mouse samples: SNVs annotated as non-synonymous or nonsense and inframe insertions or deletions were selected for each sample. Protein sequences were downloaded from Biomart using the *biomaRt* package¹⁹ and the "mmusculus_gene_ensembl" dataset available on July 2017. A peptide including 10 bases up and downstream from the alteration was extracted from the canonical protein associated with the corresponding gene. Resulting peptides matching any other annotated protein were discarded from the dataset. MHC haplotypes for the C57BL (haplotype b) and BalbC (haplotype d) mouse strains were obtained from the Affymetrix eBioscience mouse haplotype table [http://tools.thermofisher.com/content/sfs/brochures/Mouse_Haplotype_Table.pdf]. All samples in the mouse cohort belong to the C57BL/6J strain except for the CT26 cell line, which has been reported as BalbC²⁰. Peptides and the corresponding haplotypes for each sample were submitted to the netMHCpan affinity prediction software (v2.8)²¹.

Human colorectal cancer mutations and neoantigens database: Total number of mutations and predicted neoantigens for human colorectal cancer (CRC) samples were downloaded from [<https://tcia.at/neoantigens>]. Samples were classified as MSS or MSI according to the TCGA consortium²². As well as for mouse, neoantigens in the CRC database were predicted using the netMHCpan software.

Affinity thresholds in mouse and human predictions: In order to find comparable thresholds between human and mouse affinity values, we generated 600,000 predictions from 12 human haplotypes and random peptides. We found good equivalence between a cutoff of 500nM and a 2% rank threshold, which has been used to filter the mouse dataset.

Neoantigen filtering: The CRC predicted neoantigens were filtered by a maximum affinity of 500nM and minimum expression of 0 (log₂) RPKM²³. Neoantigens in mouse samples were filtered with the following criteria: 2% Rank affinity as reported by netMHCpan and 0 RPKM of the mean expression of all available replicates for the corresponding gene.

Lentiviral infection of MTOs

For bioluminescent tracking, MTOs were infected with a lentivirus encoding an mCherry-firefly luciferase fusion reporter construct, followed by an IRES-Zeocin or IRES-Puro resistance cassette³, cloned under control of the Ubiquitin promoter in a FUW vector²⁴. Virus was produced using packaging constructs in HEK293T cells in DMEM 10% FBS medium, and filtered. Trypsinized organoids (single cells) were suspended in ultra-low attachment plates (Corning) in MTO-medium and treated with successive rounds of infection (adding virus-containing medium 50% v/v) in the presence of 8 µg/ml polybrene. After a week, MTOs were put back in BME drops and selected with 50 µg/ml zeocin or 0.5-1 µg/ml puromycin (InvivoGen). After 2-3 weeks, selected cells were sorted for mCherry expression.

Tumour dynamics with Bioluminescence imaging

Growth kinetics of luciferase-expressing MTOs were tracked with in vivo bioluminescence, using an IVIS-Spectrum (Perkin Elmer) imager. Animals were anesthetized with 2.5% isoflurane and received a retro orbital injection with 50 µl D-luciferin at 15 mg/ml (Resem BV). Mice were shaved before every measurement using electrical trimmers. For quantification, two or more images per mouse (typically in the 0.5-60s exposure range, bin 4-16) were averaged, quantifying the total flux (photons s⁻¹) of a ROI spanning the abdomen (IC) or lower thorax and upper abdomen (IS). Values were normalized per mouse on the value on day 0, measured 5 min after injection. Data were processed and visualized with R/RStudio⁹ and ggplot2¹⁰. Depicted are longitudinal curves, connecting measurements of individual mice, and the group smooth (LOESS, span = 0.5) with 95% confidence interval. LOESS stands for locally weighted regression and is a non-parametric regression method analogous to a moving average in time²⁵.

Patient data analysis: Transcriptomic datasets

In this study, we used five Affymetrix microarray datasets publicly available in the NCBI GEO repository, which include gene expression and clinical information from a total of 1.194 CRC patients: briefly, GSE14333²⁶ contains a pool of 290 patients with CRC treated at 2 different hospitals: the Peter MacCallum Cancer Centre (Australia) and the H. Lee Moffitt Cancer Center (United States); the GSE33113²⁷ data set includes 90 patients with AJCC stage II disease collected at the Academic Medical Center in Amsterdam (the Netherlands); GSE39582²⁸ includes expression and clinical data for 566 patients with CRC collected for the Cartes d'Identité des Tumeurs (CIT) program, from the French Ligue Nationale Contre le Cancer; GSE3782 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37892>, accessed 20/10/2016) is a series of stage II and III CRC patients collected at five different hospitals from France (Marseille La Timone, Nice Lacassagne, Marseille Institut Paoli-Calmettes, Paris Lariboisiere, Nancy Brabois and Paris Saint-Antoine); and GSE38832²⁹ contains transcriptomic and clinical information from 127 patients treated at the H. Lee Moffitt Cancer Center (United States). In order to facilitate the integration of the datasets, samples from centres contributing with less than 10 samples (GSE37892: Marseille La Timone and Nancy Brabois; GSE38832: Nashville Veterans Affairs Medical Center) were excluded, leaving a total of 1.179 samples for downstream analyses.

Patient data analysis: Microarray processing

Processing of microarray samples was carried out separately for tumour samples of each dataset using packages *affy*³⁰ and *affyPLM*³¹ from Bioconductor³². Raw CEL files were normalized using RMA background correction and summarization³³. Standard quality controls were performed in order to identify abnormal samples³⁴ regarding: a) spatial artefacts in the hybridization process (scan images and pseudo-images from probe level

models); b) intensity dependences of differences between chips (MvA plots); c) RNA quality (RNA digest plot); d) global intensity levels (boxplot of perfect match log-intensity distributions before and after normalization and RLE plots); e) anomalous intensity profile compared to the rest of samples (NUSE plots, Principal Component Analyses). Technical information concerning samples processing and hybridization was retrieved from the original CEL files: date of scanning were collected in order to define scan batches in each dataset separately; technical metrics described by Eklund AC and Szallasi Z³⁵ were computed and recorded as additional features for each sample. Probeset annotation was performed using the information available in Affymetrix web page (Affymetrix Analysis Center. Netaffx. <https://www.affymetrix.com/analysis/index.affx>, accessed 07/27/2016). Each dataset was a-priori corrected by potential sources of bias due to technical variability. For doing so, a linear model was fitted to each gene and dataset separately including centre, three Eklunds' metrics (PM IQR, RMA IQR and RNA DEG), scanning day and the interaction between centre and Eklunds' metrics. This correction was carried out using a mixed-effect model in which gender, age at diagnosis, tumour location and microsatellite instability (MSI) status were also included as covariates, when available. Scanning day was modelled as a random effect in these models. The five transcriptomic datasets were then merged in a unique expression matrix after applying quantile normalization³⁶. Finally, expression data from each dataset was standardized at probeset level using GSE39582 as a reference; for each dataset and probeset, we selected randomly a subset of samples from GSE39582 matching the same clinical characteristics regarding gender, age, tumour location and MSI, and then centred and scaled the dataset according to these parameters.

Patient data analysis: MSI imputation

MSI status was imputed in each dataset separately, based on the expression of genes included in a published transcriptomic signature³⁷. For doing so, clustering analysis was performed on this signature via non-parametric density estimation^{38,39}. Accuracy of this imputation was evaluated in dataset GSE39582, which included annotation of microsatellite-stable (MSS) and -unstable (MSI) samples in their clinical information (97% and 80% accuracy for MSS and MSI samples, respectively).

Patient data analysis: Signatures summarization

To summarize the signatures analysed in this work, Z-scores were computed for each gene and each sample in the unique corrected expression matrix, which were then averaged across all genes included in the profile. The resulting score was centred and scaled across samples.

Patient data analysis: CD4 Th1 and Naive signatures

In order to obtain CD4 naive and CD4 th1 profiles, we carried out a differential expression analysis on gene expression dataset GSE22886⁴⁰. Analyses were performed separately for HG-133A and HG-133B platforms using a linear model with empirical shrinkage⁴¹ as implemented in *limma* R package³¹. We defined the CD4 naive signature to include genes that are at least 3 fold upregulated both in CD4 naive vs. Th1 and in CD4 naive vs. Th2, and Benjamini and Hochbergs' False Discovery Rate (FDR)⁴² < 5% in both comparisons. Similarly, the CD4 Th1 signature included genes that were at least 3 fold upregulated compared to CD4 naive, and *limma* FDR < 5%. These signatures were summarized as scores as described in the previous sections. Relative expression of Th1 and Naive cells were then measured by means of the difference of these scores, which corresponds to a log-ratio in the original scale of the microarray intensities.

Patient data analysis: Association analyses

To test the association between expression profiles, a mixed-effect model was fitted to each gene/signature independently in which scan day was included as a random effect, and centre, Eklunds' metrics and the interaction between both were introduced as covariates. Association was assessed using the corresponding Wald and Likelihood Ratio Test (LRT) provided by the linear model. As measure of association and for continuous variables, we computed the Pearson's Correlation Coefficient after correction by technical effects using the mixed-effect model described in the previous section; for categorical variables, adjusted group means and corresponding intervals at 95% confidence were calculated. Association with Cancer Associated Fibroblasts (CAFs) was assessed using a FAP signature derived from human CRC sorted cells (see section *Gene expression in tumour cell populations*). To assess association with overall TGFB levels, a signature was created using microarray probesets *203085_s_at*, *220406_at* and *209747_at*, respectively, as these probesets correlated better with measurements of *TGB1*, *TGFB2* and *TGFB3* obtained by qRT-PCR data³. When assessing correlation between TGF-beta and any signature, probesets mapping to *TGFB1*, *TGFB2* or *TGFB3* were previously excluded from the later. For visualization purposes, scatter plots and boxplots were yielded using the technical adjusted values for expression intensities.

The estimates of the association between Th1 and Naive cells ratios and the *TGF-beta* and the FAP signatures were compared with the overall trend observed in the whole genome. For doing so, we generated a null distribution for the correlation coefficients between *TGF-beta* and the scores derived for random signatures. We randomly selected the same number of genes as those corresponding to the original naïve and Th1 signatures and computed the score's ratio. *TGF-beta* probes were excluded for the *TGF-beta* correlation analysis. Bilateral p-values were computed from 10,000 repetitions. The threshold for statistical significance was set at 5%. All these analyses were carried out using the R^{9,31} packages *lme4*⁴³ and *lmerTest*⁴⁴.

Patient data analysis: Survival analyses

Association with metastasis was evaluated using a frailty Cox proportional hazards model as previously described⁴⁵ and implemented⁴⁶. Statistical significance was assessed by means of a Log-likelihood Ratio Test (LRT), while 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 by technical effects, as described in the previous section. When evaluating expression intensities, technical effects were included in the models as covariates. In an analogous way to the correlation analyses, scan batch was included as a random effect in the frailty Cox models. Hazard Ratios (HR) and their corresponding 95% confidence intervals were computed as a measure of association. 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^{9,31}.

To compare the risk estimations with the general trend observed in the dataset, we generated a null distribution in the same way described in the previous section. As with the original signature, a Cox model was used to compute the statistic of the three pairwise comparisons for random scores and a bilateral p-value was computed from 10,000 repetitions.

MTO-derived whole tumour expression data

cDNA Library preparation and amplification were performed with the WTA2 kit (Sigma-Aldrich) from 25 ng of starting material. The cDNA was amplified for 17 cycles and purified using PureLink Quick PCR Purification Kit (Invitrogen). Quantification of amplified cDNA was done on a Nanodrop ND-1000 spectrophotometer (Thermo-Fisher Scientific, Waltham, MA, USA). 8.5 ug of the cDNA from each sample was fragmented and labelled with the GeneChip Mapping 250 K Nsp assay kit (Affymetrix). Hybridization was performed using the GeneAtlas Hyb, Wash and Stain Kit for 3' IVT arrays. Samples were denatured at 96°C for 10 min prior to incubation in the Affymetrix GeneAtlas Mouse MG-430 PM Array Strip. Hybridization was performed for 16 h at 45 °C in the GeneAtlas Hybridization Oven. Washing and Stain steps after hybridization were performed in the GeneAtlas Fluidics Station, following the specific script for Mouse MG-430 PM Arrays. Arrays were scanned with GeneAtlas Scanner using default parameters, and the generation of Cel files for bioinformatics analysis was done with GeneAtlas software (all Affymetrix).

As for human transcriptomic datasets, CEL files from organoids-derived mice whole tumours were normalized with RMA³³ using the Bioconductor R packages *affy*³⁰ and *affyPLM*³¹. Standard quality controls (see section Patient data analysis: Microarray processing) discarded the presence of abnormal samples. For probeset annotation we used the information available in the Affymetrix web page (Affymetrix Analysis Center. Netaffx. <https://www.affymetrix.com/analysis/index.affx>, accessed 08/01/2017).

TBRS comparisons:

Previous to the summarization of TBRS signatures, the expression matrix was corrected a-priori by Eklund's metrics³⁵ RMA.IQR and RNA.DEG, as they were observed to be an important source of technical variability. For such correction, a linear model was used with no other variables included as covariates. Signatures scores were computed independently for samples with three or four mutations. Scores between treatment conditions were compared using a two-sided Mann-Whitney-Wilcoxon test.

Classification of mice tumours according to CMS (extended)

To assess the similarity to the poor-prognosis CMS4 phenotype⁴⁷, we first attempted to use the Single-Sample-Predictor (SSP) provided in their original work. Nevertheless, this analysis did not succeed in classifying any of the samples being tested. For this reason, we hypothesized that the differences of platform and/or species between the training (human) and test data (mice) could be harming the performance of this classifier.

As an alternative method, we used *k*-Top Scoring Pairs (*k*TSP)⁴⁸ as implemented in the switchBox R package⁴⁹. *k*TSP is a ranked-based method that relies in the ordering in small number of features within each sample and, thus, is very suitable in multi-platforms settings in comparison with other top-performer methods⁵⁰. This *k*TSP classifier was trained for a binary outcome (CMS4 vs not-CMS4) in 960 samples with CMS annotation provided by the original work (168 CMS1, 421 CMS2, 127 CMS3 and 244 CMS4)⁴⁷. The concordance between the SSP and *k*TSP methods was assessed in two human datasets that did not participate in the definition of the CMS classification: GSE38832 and GSE44076⁵¹; phi coefficients were 0.83 and 0.80 for these series, respectively:

GSE38832		kTSP	
		CMS1/2/3	CMS4
SSP	CMS1/2/3	67	7
	CMS4	0	23
	Not classified	16	9

GSE44076		kTSP	
		CMS1/2/3	CMS4
SSP	CMS1/2/3	61	4
	CMS4	0	8
	Not classified	13	12

Overall, these results suggested that, as *kTSP* could be more suitable for handling platforms and species differences between training and test datasets, it still retained the good classification performance attributable to SSP.

Previous to any attempt of classification, both mice and human CRC data were summarized from probeset to gene level (Entrez). For doing so, we computed the first principal component from all probesets mapping to the each gene. This component was then centred and scaled to the weighted mean of the means and standard deviations of the probesets using the corresponding contribution to the component as weight. The sign of the component was eventually changed to be congruent to the sign of the probeset contributing the most to the component. Next, mice genes were translated to their corresponding human homologous using the Mouse Genome Informatics Database⁵². Only high confidence homologous genes according to Biomart^{53,54} were kept for downstream analyses (Mouse Orthology Confidence=1 and Mouse Gene-order conservation score=100; 10,015 genes). Genes to be included in the model were pre-selected from those over-expressed (1.5 fold-change minimum difference; 293 genes) or under-expressed (1.25 fold-change minimum difference; 215 genes) in CMS4 samples compared with any other subtype in the patients transcriptomic dataset. The classifier was internally tested in the training set of human samples by cross-validation leaving one dataset out for test at each iteration. A classifier trained with all the human samples available was then applied to the mice tumours using a simple rule of majority of votes. The decision cut-off was set at the midpoint of the maximum number of votes, which corresponds to the number of gene pairs used for classification. For visualization purposes, the number of votes for each classifier (cross-validation and training) were re-scaled to a score ranging between zero and one in order to represent all the classification results at the same scale.

Gene pairs:

UP	DOWN
<i>MSRB3</i>	<i>EPT1</i>
<i>FERMT2</i>	<i>RMI1</i>
<i>EFEMP2</i>	<i>ASF1B</i>
<i>SPOCK1</i>	<i>STIL</i>
<i>DDR2</i>	<i>CCNA2</i>
<i>TAGLN</i>	<i>HMGB2</i>
<i>CCDC80</i>	<i>CDC45</i>
<i>GLI3</i>	<i>KIF18A</i>
<i>TNSI</i>	<i>UNG</i>
<i>STON1</i>	<i>WHSC1</i>
<i>PTRF</i>	<i>KIF18B</i>
<i>SFRP2</i>	<i>HK2</i>
<i>MGP</i>	<i>RBM47</i>
<i>GAS1</i>	<i>DONSON</i>
<i>BNC2</i>	<i>PLK4</i>

<i>SLIT2</i>	<i>CCDC134</i>
<i>DPYSL3</i>	<i>EIF4E</i>
<i>AEBP1</i>	<i>RHPN2</i>
<i>PCDH7</i>	<i>FAM83F</i>
<i>MAP1B</i>	<i>FANCD2</i>
<i>CRYAB</i>	<i>HOOK1</i>
<i>PRRX1</i>	<i>CDCA2</i>
<i>FBN1</i>	<i>EZH2</i>
<i>MXRA8</i>	<i>GMCL1</i>
<i>PTGIS</i>	<i>CENPA</i>
<i>ZFPM2</i>	<i>ORC1</i>
<i>MLLT11</i>	<i>SPAG5</i>
<i>MYL9</i>	<i>TMEM54</i>

In agreement with the reasoning of a single-sample classifier, SSP and *k*TSP were both applied to a version of the mice data previous to any correction by technical effects. In addition, *k*TSP was also trained before technical corrections were applied to the human transcriptomic dataset. For the concordance analyses between SSP and *k*TSP, GSE44076 cell files were downloaded from GEO and processed with RMA using packages *affy*³⁰ and *affyPLM*³¹ from Bioconductor³².

Flow cytometry: single cell suspension 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 37°C 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 μ m 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 4°C. 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. After filtration through a 70 μ m 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).

Flow cytometry: immunophenotyping analysis and sorting

Cells were incubated 10 min on ice in FACS buffer in the presence of anti-CD16/CD32 (clone 93, eBioscience) to block Fc receptor. Mix of conjugated antibodies was added in the presence of anti-CD16/CD32 and the cells were stained during 20 min on ice. After staining, cells were washed and labelled for cell viability with the LIVE/DEAD Fixable Cell Dead stain kit (Life Technologies) during 7 min. For cytokine analysis, cell suspensions were incubated 4 h in PMA/ionomycin (Sigma-Aldrich) and brefeldin A (eBioscience) at 37 °C. Intracellular staining was performed using IC fixation/permeabilization kit (eBioscience). Cells were stained using PBS, 1% FBS, 1% HEPES and 0.6% EDTA (Gibco).

Gating strategies: Figure 1g (MTO biobank generation): To sort Lgr5-GFP+ tumour cells, we gated: Cells/Single cells/Living cells/GFP+. Figures 2d, S9b and S11a: To assess genetic recombination in the *Ubc-CreER^{T2}; Tgfb2^{fl/fl}; R26^{mTmG}* 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 3h: T cell activation markers were assessed inside: Living cells/Cells/Single cells, then CD45+/CD3+, then CD4+/CD8- or CD4-/CD8+. Figure 3i and 4j: T cell cytokine levels were assessed inside: Living cells/Cells/Single cells/CD45+/CD3+/NK1.1-, then CD4+/CD8- or CD8+/CD4-. Figure 3j: CD8+ T cells were sorted using gates: Living cells/Cells/Single cells/CD45+/CD19-/MHCII-/CD3+/CD4-/CD8+. Figure 4d: PD-L1 expression was assessed inside: Living cells/Cells/Single cells/CD45+/CD3+, then CD4+/CD8- or CD8+/CD4-. Figure 4e: 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-). Figure 4g: PD-L1 was measured inside Living cells/Cells/Single cells/CD45+. Figure 4i: For T cell activation markers we gated: Living cells/Cells/Single cells/CD45+/CD3+/NK1.1-, then CD4+/CD8- or CD8+/CD4-. Figure S13d: Myeloid populations were assessed inside: Living cells/Cells/Single cells/CD45+/PD-L1+/Ly6G-/Ly6C+/CD11b+/CD11c+/-.

Flow cytometry analysis and cell separation were performed in a FACSAriaFusion flow cytometer (Beckton Dickinson). Data were analyzed using FlowJo software (v 10.4).

The following antibodies were used for the 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-IFN γ (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.

Gene expression in tumour cell populations

RNA from sorted cells was processed and amplified as previously described⁵⁵. For mouse liver metastases: To assess *Tgfb1*, -2 and -3 mRNA levels, we performed RT-qPCR with triplicate reactions (each 5 ng of cDNA) in a StepOne instrument (Thermo Fisher) with Taqman probes (*Tgfb1* Mm01178819_m1; *Tgfb2* Mm00436955_m1; *Tgfb3* Mm00436960_m1; endogenous control *Actb* Mm00607939_s1). Other probes used are: *Aurkb* Mm_01718140_m1, *mKi67* Mm01278617_m1, *Gzma* Mm01304452_m1, *Gzmb* Mm00442837_m1, *Pdcd1* Mm01285676_m1, and *Gapdh* Mm99999915_g1.

Sorted CRC populations from human samples were described previously^{1,3}. Briefly, two GEO data sets were used to characterize gene profiles according to specific gene expression in tumour cell subpopulations: GSE39395 and GSE39396³. In these datasets and as described previously³, FACS was used to separate the following populations from 14 fresh CRC samples: CD45+/EpCAM-/CD31-/FAP-, CD45-/EpCAM+/CD31-/FAP-, CD45-/EpCAM-/CD31+/FAP- and CD45-/EpCAM-/CD31-/FAP+. To homogenize these two expression matrices, dataset GSE39396 was centred and scaled pairwise to the mean and standard deviation of GSE39395. A signature of Cancer Associated Fibroblasts (CAFs) was derived from the resulting expression matrix. For doing so, we selected probesets

overexpressed in the CD45-/EpCAM-/CD31-/FAP+ samples with a three minimum fold and raw p-value < 0.05 compared to any other cell population. These analyses were performed using a linear model with empirical shrinkage⁴¹ as implemented in *limma* R package³¹. Also TGF- β levels by cell populations were assessed in this dataset using Kruskal Wallis (KW) and two-sided Mann-Whitney tests.

Supplementary Methods References

1. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320–329 (2015).
2. 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).
3. 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).
4. 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**, 1077–85 (2007).
5. Merlos-Suárez, A. *et al.* The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. *Cell Stem Cell* 511–524 (2011). doi:10.1016/j.stem.2011.02.020
6. Jung, P. *et al.* Isolation and in vitro expansion of human colonic stem cells. *Nat Med* **17**, 1225–1227 (2011).
7. Tauriello, D. V. F. *et al.* Loss of the tumor suppressor CYLD enhances Wnt/ β -catenin signaling through K63-linked ubiquitination of Dvl. *Mol. Cell* **37**, 607–19 (2010).
8. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
9. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2016). at <<http://www.r-project.org>>
10. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2009). at <<http://ggplot2.org>>
11. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
12. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–8 (2011).
13. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w 1118; iso-2; iso-3. *Fly (Austin)*. **6**, 80–92 (2012).
14. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
15. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

16. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
17. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
18. Rossell, D., Stephan-Otto Attolini, C., Kroiss, M. & Stöcker, A. Quantifying alternative splicing from paired-end RNA-seq data. *Ann. Appl. Stat.* **8**, 309 (2014).
19. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–91 (2009).
20. Castle, J. C. *et al.* Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190 (2014).
21. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
22. Cancer Genome Atlas Network, T. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–7 (2012).
23. Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **18**, 248–262 (2017).
24. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–72 (2002).
25. Cleveland, W. S. & Devlin, S. J. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J. Am. Stat. Assoc.* (2012).
26. Jorissen, R. N. *et al.* Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer. *Clin. Cancer Res.* **15**, 7642–7651 (2009).
27. De Sousa E Melo, F. *et al.* Methylation of cancer-stem-cell-associated wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* **9**, 476–485 (2011).
28. Marisa, L. *et al.* Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLoS Med.* **10**, e1001453 (2013).
29. Tripathi, M. K. *et al.* Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res.* **74**, 6947–6957 (2014).
30. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
31. Gentleman, R. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Journal of the American Statistical Association* **102**, (Springer New York, 2005).
32. Gentleman, R. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
33. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

34. Gentleman, R., Huber, W., Carey, V., Irizarry, R. & Dudoit, S. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Book* (Springer Science+Business Media, 2005).
35. Eklund, A. C. & Szallasi, Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biol.* **9**, R26 (2008).
36. Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
37. Jorissen, R. N. *et al.* DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers. *Clin. Cancer Res.* **14**, 8061–8069 (2008).
38. Azzalini, A. & Torelli, N. Clustering via nonparametric density estimation. *Stat. Comput.* **17**, 71–80 (2007).
39. Azzalini, A. & Menardi, G. Clustering via Nonparametric Density Estimation: The R Package pdfCluster. *J. Stat. Softw.* **57**, 1–26 (2014).
40. 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).
41. Smyth, G. K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–26 (2004).
42. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, (1995).
43. Bates, D., Maechler, M., Bolker, B. M. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. (2015). at <<http://arxiv.org/abs/1406.5823>>
44. Kuznetsova, A., Bruun Brockhoff, P. & Christensen, Haubo Bojesen, R. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-32. (2016). at <<https://cran.r-project.org/package=lmerTest>>
45. Therneau, T. M., Grambsch, P. M. & Pankratz, V. S. Penalized Survival Models and Frailty. *J. Comput. Graph. Stat.* **12**, 156–175 (2003).
46. Therneau, T. M. coxme: Mixed Effects Cox Models. R package version 2.2-5. (2015). at <<https://cran.r-project.org/package=coxme>>
47. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
48. Tan, A. C., Naiman, D. Q., Xu, L., Winslow, R. L. & Geman, D. Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics* **21**, 3896–3904 (2005).
49. Afsari, B., Fertig, E. J., Geman, D. & Marchionni, L. switchBox: an R package for k-Top Scoring Pairs classifier development. *Bioinformatics* **31**, 273–274 (2015).
50. Marchionni, L., Afsari, B., Geman, D. & Leek, J. T. A simple and reproducible breast cancer prognostic test. *BMC Genomics* **14**, 336 (2013).
51. Sanz-Pamplona, R. *et al.* Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol. Cancer* **13**, 46 (2014).

52. Blake, J. A. *et al.* Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Res.* **45**, D723–D729 (2017).
53. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
54. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589-98 (2015).
55. Gonzalez-Roca, E. *et al.* Accurate Expression Profiling of Very Small Cell Populations. *PLoS One* **5**, e14418 (2010).