Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth

Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth.

Germano G, Lamba S, Rospo G, Barault L, Magrì A, Maione F, Russo M, Crisafulli G, Bartolini A, Lerda G, Siravegna G, Mussolin B, Frapolli R, Montone M, Morano F, de Braud F, Amirouchene-Angelozzi N, Marsoni S, D'Incalci M, Orlandi A, Giraudo E, Sartore-Bianchi A, Siena S, Pietrantonio F, Di Nicolantonio F, Bardelli A.

Nature. 2017 Dec 7;552(7683):116-120. doi: 10.1038/nature24673. Epub 2017 Nov 29.

Inactivation of DNA repair triggers neoantigen generation and impairs tumor growth

Giovanni Germano[1,2,3], Simona Lamba[1], Giuseppe Rospo[1], Ludovic Barault[1,2,], Alessandro Magrì[1,2], Federica Maione[1], Mariangela Russo[1,2], Giovanni Crisafulli[1,2], Alice Bartolini [1], Giulia Lerda[1], Giulia Siravegna[1,2,3], Benedetta Mussolin[1], Roberta Frapolli [4], Monica Montone[1], Federica Morano[5], Filippo de Braud[5,6], Nabil Amirouchene-Angelozzi[1], Silvia Marsoni[1], Maurizio D'Incalci[4], Armando Orlandi[7], Enrico Giraudo[1,8], Andrea Sartore-Bianchi[9], Salvatore Siena[6,9], Filippo Pietrantonio[5], Federica Di Nicolantonio[1,2], and Alberto Bardelli[1,2,#]

[1]Candiolo Cancer Institute – FPO, IRCCS, Candiolo, Turin, Italy; [2]University of Turin, Department of Oncology, Candiolo, Turin, Italy; [3] FIRC Institute of Molecular Oncology (IFOM), Milan, Italy; [4] IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy; [5] Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [6] Department of Oncology and Hemat-Oncology Università degli Studi di Milano, Milan, Italy; [7]Policlinico Universitario A. Gemelli, Roma, Italy; [8]University of Torino, Department of Science and Drug Technology, Turin, Italy; [9]Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy.

Address correspondence to Alberto Bardelli: alberto.bardelli@unito.it

Abstract

Molecular alterations in genes involved in DNA mismatch repair (MMR) promote cancer initiation and foster tumor progression [1]. MMR deficient cancers frequently show favorable prognosis and indolent progression [2]. The functional basis of the clinical outcome of patients with MMR deficient tumors is not clear. To address this, we genetically inactivated MutL homolog 1 (MLH1) in colorectal, breast and pancreatic mouse cancer cells. The growth of MMR deficient cells was comparable to their proficient counterparts in vitro and upon transplantation in immunocompromised mice. In contrast MMR deficient cancer cells grew poorly when transplanted in syngeneic mice. MMR inactivation increased the mutational burden and led to dynamic mutational profiles, resulting in persistent renewal of neoantigens *in vitro* and *in vivo*, while MMR proficient cells exhibited stable mutational loads and neoantigen profiles over time. Immune surveillance improved when cancer cells in which MLH1 had been inactivated accumulated neoantigens for several generations. MMR-driven dynamic generation of neoantigens, when restricted to a clonal population, further increased immune detection. Inactivation of MMR repair, driven by acquired resistance to the clinical agent temozolomide (TMZ), increased mutational loads, promoted continuous renewal of neoantigens in human colorectal cancers and triggered immune surveillance in mouse models. These results suggest that targeting DNA repair processes can increase the burden of neoantigens in tumor cells and could be exploited for therapeutic approaches.

Main text

Several tumor suppressor genes are involved in amending DNA replication errors that occur during cell division [3,4]. Post replicative DNA mismatch repair is performed by protein complexes, consisting of MutL homolog 1 (MLH1), MutS protein homolog 2 (MSH2), MutS homolog 6 (MSH6) and PMS1 homolog 2 (PMS2)[5,6]. When the MMR machinery is defective, cancer cells display characteristic microsatellite instability (MSI) [7]. MMR deficient colorectal tumors have peculiar clinical features, which include early onset and rapid progression, but remarkably a favorable prognosis [2]. The molecular bases of these clinical features are poorly understood. Recent evidence that MSI tumors respond prominently to anti-immune checkpoint blockade led to the hypothesis that elevated mutation load (high mutation burdens) is required for immunotherapy to be effective [8-10]. Indeed tumors with high environmental exposure-related mutational burdens (such as melanoma, bladder and lung cancers) also preferentially respond to immunotherapy [11-13]. However, the association between number of mutations and response to immuno-checkpoint blockade is not complete, as a large fraction of hyper-mutated tumors do not respond to immune checkpoint modulators [14].

To functionally define the role of mismatch repair in tumor formation and response to immunotherapy we studied MMR proficient murine colorectal (CT26, MC38), breast (TS/A) and pancreatic (PDAC) cancer cells (Extended Data Figure 1a). Genome editing with the CRISPR-Cas9 system was employed to inactivate *Mlh1* in each of these cell models. Independent sgRNA guides targeting distinct *Mlh1* exonic regions were used and multiple clones were isolated. Clones derived from cells treated with non-targeting gRNA served as controls (Ctrl clones). Inactivation of DNA mismatch repair was confirmed at the genomic and protein level (Figure 1 and Extended Data Figure 1b, 2a, 2b). Functional inactivation was also evaluated by analyzing repetitive mouse DNA elements (Extended Data Figure 3).

In vitro, the proliferative rates of MMR deficient cells were comparable to that of control clones (Extended Data Figure 4a). MMR deficient cells rapidly developed tumors when injected subcutaneously in immunocompromised mice and within a few

weeks all animals had to be sacrificed according to ethical guidelines (Figure 1a, 1e and Extended Data Figure 2).

When CT26 cells were injected in immunocompetent syngeneic mouse (BalbC), tumors grew rapidly and after 30 days the animals had to be sacrificed. On the contrary MMR deficient CT26 cells (KO-1 and KO-2 clones) grew poorly in the same conditions (Figure 1b). Similar data were obtained in MC38 cells (Extended Data Figure 2a) although the phenotype was less marked since parental MC38 are already partially immunogenic as previously reported [15].

Since expression of an exogenous Cas9 protein from *S.Pyogenes* might affect cell growth in syngeneic mouse models by triggering immune-mediated responses [16], we also generated *Mlh1* KO cell lines where Cas9 was transiently expressed (Extended Data Figure 4c). The latest also grew less than controls in immunocompetent syngeneic mouse, however the phenotype was less evident compared to the first set of *Mlh1* KO 'constitutively Cas9- expressing' clones (Figure 1c). Nevertheless, we found that the two models gave comparable results when transiently Cas9-expressing cells were expanded *in vitro* for several generations before *in vivo* injection (Figure 1c). Notably, tumor formation was also severely impaired if *Mlh1* KO cells were sub-cloned (by single-cell dilution) before injection in immunocompetent animals (Figure 1c).

To assess the impact of the anatomic location on the tumorigenic potential of MMR proficient and deficient models, *Mlh1* KO CT26 (with transient expression of Cas9) cells were injected orthotopically in the intestine of BalbC mice. Tumors developed rapidly in mice injected with control cells, while *Mlh1* KO cells either did not form tumors or developed small lesions (Figure 1d).

To assess whether these findings might extend to another cancer type we generated mouse pancreatic ductal adenocarcinoma (PDAC) [17] cells lacking *Mlh1*, using both constitutive and transient Cas9-expressing systems (Figure 1e and Extended Data 1c). Proliferation of MMR proficient and deficient PDAC cells was comparable (Extended Data Figure 4a and 4b) and *Mlh1* KO cells rapidly developed tumors when injected ortothopically in immunocompromised mice (Figure 1e). While large tumors were evident in syngeneic (FVB/N) mice three weeks after transplantation of control cells,

Mlh1 KO cells either did not form tumors or developed very small lesions (Figure 1f and 1g).

We then deleted *Mlh1* in a breast adenocarcinoma cell line (TS/A). Ctrl and *Mlh1* KO TS/A clones efficiently formed tumors in immunocompromised mice (Extended Data Figure 2b). After 120 days Ctrl and *Mlh1* KO cells were injected subcutaneously and orthotopically in the mammary fat pad. Control cells were tumorigenic, while MMR deficient breast cells failed to grow or formed only small masses (Extended Data Figure 2b).

We next assessed the impact of DNA repair inactivation also on fully established tumors. Since MMR deficient do not grow efficiently in syngeneic animals, we first inoculated *Mlh1* knockout CT26 cells in multiple immunocompromised mice until tumors reached 2000 mm^3 in size, at which point tumor fragments were transplanted in syngeneic BalbC. Under these conditions, when re-introduced into immunocompetent mice *Mlh1* knockout, CRC cells continued to grow. We reasoned that the transplantation system might allow us to recapitulate the clinical setting of patients with fully established MSI tumors treated with checkpoint inhibitors. When transplanted tumors were treated with antibodies targeting programmed death-1 (PD-1) and cytotoxic T lymphocyte-associated antigen 4 (CTLA-4), the growth of MMR deficient cells was markedly impaired, while MMR proficient cells grew despite treatment with checkpoint inhibitors (Figure 2a). Analogous results were obtained with *Mlh1* KO clones generated with the transiently Cas9-expressing editing system (Figure 2b). Increased levels of cytotoxic CD8+ T cells were found in MMR deficient tumors (Extended Data Figure 5).

These data suggested that functionally reactive T cells might be responsible for the impaired tumorigenesis of MMR deficient cells. To test this, we injected MMR deficient cells in the presence of depleting CD8 antibody; isotype matched antibodies served as controls. MMR deficient cells readily formed tumors in syngeneic mice only when CD8 T cells were suppressed (Figure 2c).

It has been shown that the mutational burden correlates with response to immunotherapy in human malignancies[10,11,18].We performed exome sequencing of CT26 obtain-

ing that the mutational load of parental cells (152 mutations /megabase) was comparable with previous analyses[19].

To study the relationships between the number of mutations and the phenotype of MMR deficient cells we performed exome sequencing of cells harvested at distinct time-points. RNA sequencing analysis indicated that a large proportion of the mutant genes are transcribed and can act as potential neoantigens when analyzed with HLA matching algorithms (Supplementary Table 1). In MMR proficient cells the number of neoantigens remained essentially constant over time. On the contrary, *Mlh1* KO cells, showed increased mutational loads, which dynamically evolved along with the number of predicted neoantigens (Figure 3a). Analogous results were obtained in *Mlh1* KO clones expressing Cas9 constitutively (Extended Data Figure 6).

We next studied the impact of increasing the fraction of clonal neoantigens by single cell cloning MMR deficient cells (Figure 1c). We found that the sub-cloning process increased the mutational load and the fraction of predicted clonal neoantigens in *Mlh1* KO but not in their control cells (Extended Data Figure 7a and 7b). These results suggest that the increase of mutational loads and clonal neoantigens burden, which occurred during the sub-cloning procedure, likely contributed to the growth impairment observed *in vivo*.

We further found that in MMR deficient cells (but not in matched Ctrl) new arrays of antigens dynamically evolve in tumors growing in syngeneic animals, indicating that DNA repair inactivation promotes continuous emergence of neoantigens also *in vivo* (Extended Data Figure 7c).

To evaluate the impact of MMR deficient cells on the host adaptive immunity we analyzed the repertoire of TCR rearrangements in mice that received either MMR proficient or deficient tumor cells. To determine TCR productive rearrangements (TCRB CDR3), DNA from peripheral blood was amplified with TCR specific primers and subjected to NGS. Bioinformatics analyses revealed the expansion of the 20 most represented TCR rearrangements in animals injected with MMR deficient cells (Figure 3b). Analogous findings were obtained in sub-cloned cells (Extended Data Figure 7d).

We reasoned that ectopically increasing the mutational burden, in cancer cells, may be–paradoxically- beneficial for therapeutic purposes. We therefore considered how to foster the mutational loads in cancer cells using pharmacological agents. We and others previously reported that treatment with genotoxic agents can drive inactivation of MMR in human cancer cells [20].

A pharmacological screen showed that MMR proficient cells display preferential sensitivity to temozolomide (TMZ) (Extended Data Figure 8). Temozolomide is approved for treatment of several tumor types and triggers DNA damage [21]. It has been previously shown that TMZ exposure affects DNA repair and treatment with this agent can result in the selection of cells with MMR inactivation [22,23]. We treated CT26 and MC38 MMR proficient cells with TMZ until resistant populations emerged. Upon injection in syngeneic mice, drug resistant CT26 cells readily formed tumors and grew at rates comparable to their parental counterparts, while TMZ resistant MC38 cells did not form tumors (Extended Data Fig. 9a). We assessed MMR status, mutational loads and number of predicted neoantigens in CT26 and MC38 cells resistant to TMZ (Extended Data Fig 9b). Upon exposure to TMZ MC38 but not CT26 cells displayed microsatellite instability (Extended Data Figure 9c). We further found that MLH1 expression was noticeably reduced in MC38 (but not in CT26) cells (Extended Data Fig. 9d) and that TMZ resistant MC38 cells carried an *Mlh1* deletion, which was not present in (sensitive) parental cells (Extended Data Figure 9e).

Mouse cancer models have several limitations, for example they do not entirely reflect the extent of tumor heterogeneity and the burden of disease observed in human tumors. To assess whether results gathered in mouse cancer models might translate to human disease we exploited a large database of CRC lines [24]. We characterized a total of 47 CRC cell models for O^6-methylguanine-DNA-methyltransferase (MGMT) expression, methylation, MSI status and sensitivity to TMZ. A subset of cells highly sensitive to TMZ was identified (Extended Data Figure 10a). Consistent with previous work, most of the cells with high sensitivity to TMZ showed low MGMT expression and high methylation levels, in line with the notion that MGMT is the main enzyme responsible for repairing the DNA adducts induced by TMZ (Extended Data Figure 10b)[25]. Several of the sensitive lines were exposed to TMZ until drug resistant popu-

lations emerged (Extended Data Figure 10c). Consistent with earlier clinical observations [26] [27], in some cell models, resistance was associated with increased expression (or de-novo expression) of MGMT (Extended Data Figure 10b). However, in multiple TMZ resistant cells, in which MGMT levels remained low, exome data revealed frequent molecular alterations in MMR genes (Extended Data Figure 10d). In analogy with data obtained in mouse models, human cells in which resistance was associated with alteration of MMR showed higher numbers of predicted neoantigens (Extended Data Figure 10d), which increased over time.

We sought to extend these findings in clinical settings. We previously found that metastatic CRC patients whose tumors display high levels of *MGMT* promoter methylation and low MGMT protein expression are more likely to respond to TMZ containing regimens but unfortunately relapse within 4-11 months of initiating therapy [25]. We retrieved tissues from patients enrolled in previous [28] or ongoing clinical trials (NCT02414009) with TMZ based chemotherapy in CRC. Post-treatment tissue biopsies were available for molecular analyses in five individuals who achieved partial responses or prolonged stabilization (Extended Data Figure 10e). When MGMT protein status was assessed, three cases showed increased expression in the post treatment biopsy (Extended Data Figure 10f). Exome sequencing of DNA from tissue obtained at progression and matched PBMC, followed by bioinformatics analyses, showed low-average mutational burdens (<10Mut/Mb) in three out of five samples but high mutational levels (> 60 Mut/Mb) in the other two. The latter corresponded to the samples with low MGMT protein expression after TMZ treatment (Extended Data Figure 10f). When pre-treatment tissue from these two cases was subjected to exome sequencing, we found that specimens obtained before TMZ had low mutational burdens (< 10 Mut/Mb). These findings prompted us to study the status of MMR genes before and after treatment in all five cases. The three tumors in which resistance was associated with MGMT protein expression did not show alterations in MMR genes (Extended Data Figure 10f). On the contrary, the cases with high mutational loads displayed a mutation in the MSH6 gene (pT1219I) predicted to negatively alter the protein function, which was absent in pre-treatment tissue (Extended Data Figure 10d and 10f).

Extensive efforts have been placed at developing drugs capable of restoring the function of tumor suppressor proteins in the hope they could act as anticancer agents [29].

However our data indicate that permanent inactivation (rather than reactivation) of certain tumor suppressors involved in DNA-repair could instead be pursued for therapeutic purposes. The rationale for this unconventional approach is based on the concept that dynamic increases of the number of mutations in cancer cells can result in immune responses. Inactivation of mismatch repair leads to increased number of SNV and indels, some of the latter result in frameshifts generating neoepitopes that are qualitatively very different from self[30]. Accordingly, MMR deficiency leading to indels-frameshifts might be key to the striking immune-surveillance we observed upon inactivation of MLH1.

Results obtained with TMZ suggest that drugs leading to inactivation of DNA repair in cancer cells could be systemically tolerable. Importantly, these results should not be interpreted to infer that increasing tumor heterogeneity with chemotherapy is an effective way to promote immune-surveillance, rather that it is possible to inactivate DNA repair *in vivo* to improve immune-surveillance and response to immune-checkpoint blockade. Increasing the number of mutations and neoantigens *per se* might not be sufficient to trigger tumor rejection. For such an approach to be effective, cancer cells must accumulate high number of de novo mutations/neoantigens.

Our results also suggest that MMR-driven dynamic generation of neoantigens, when restricted to a clonal population, further increases immune detection. Accordingly, tumor debulking and concomitant induction of MMR deficiency may be coupled to obtain effective results. Therefore our data support the exploitation of neoantigen generation prior to or at the time of a profound bottleneck such as surgery in the adjuvant setting rather than in advanced disease.

Besides *Mlh1,* other genes involved in DNA repair (such as the POLE and POLD polymerase) could also be targeted to increase mutational loads. Mismatch repair proteins and DNA polymerases are endowed with well-characterized enzymatic activity (ATPases and exonucleases respectively) which are amenable to pharmacological blockade.

In conclusion our data suggest that inactivation of DNA mismatch repair causes a hyper-mutation status that increases tumor neoantigens, which in turn, trigger long-lasting immune surveillance that can be further enhanced by immune-modulators.

Material and Methods

Cell models

CT26 is a mouse undifferentiated colon carcinoma derived from BalbC . MC38, is a murine colon adenocarcinoma derived from C57/Bl6 [31] . MC38 and CT26 cells were provided by Maria Rescigno (European Institute of Oncology). The TS/A breast cancer cell line was established from a moderately differentiated mammary adenocarcinoma that arose spontaneously in a BALB/c mouse [32]. TS/A cells were provided by Federica Cavallo (Molecular Biotechnology Center, University of Torino, Italy). Murine pancreatic ductal adenocarcinoma cells (PDAC) were isolated, as previously described [33] , from transgenic mice bearing pancreatic cancers with the following genotype: $p48^{cre}$, $Kras^{LSL\_G12D}$, $p53^{R172H/+}$, $Ink4a/Arf^{flox/+}$. PDAC cells were kindly provided by Doug Hanahan (ISREC, EPFL, Lausanne, Switzerland)[33]. CT26 and MC38 were cultured in RPMI1640 10% FBS, plus glutamine, penicillin and streptomycin (Sigma Aldrich). TS/A and PDAC were cultured in DMEM 10% FBS plus glutamine, penicillin and streptomycin (Sigma Aldrich). All the cells have been exome sequenced. For CT26 and MC38 the data have been compared with published data. For PDAC and TSA we compared the exome analysis with the relative background (BalbC and FVB/N) that were confirmed. All cell lines were tested for mycoplasma regularly. Cells were also tested for human (using the human Comprehensive Clear Panel, Charles River) and murine pathogens (using the Mouse/rat Comprehensive Clear Panel, Charles River). To ensure that the parental cell models were tumorigenic, before starting the genome editing experiments, all the lines were injected in matched syngeneic mice. Upon tumors formation we reestablish *in vitro* cell cultures.

Animal studies

All animal procedures were approved by the Ethical Commission of the University of Turin and by the Italian Ministry of Health, and they were performed in accordance

with institutional guidelines (4D.L.N.116, G.U., suppl. 40, 18-2-1992) and international law and policies (EEC Council Directive 86/609, OJ L 358, 1, 12-12-1987; *NIH Guide for the Care and Use of Laboratory Animals*, US National Research Council, 1996). The number of mice included in the experiments and the inclusion/exclusion criteria were based on the institutional guidelines (above). We observed the tumor size limits in accordance with institutional guidelines. Our protocol limits six- to eight-weeks old female and male C57/BL6J, BalbC, FVB/N and Nod/Scid. Mice were obtained from Charles River (Calco, Como Italy). All experiments involved a minimum of five mice per group (with the exception of the experiment in Figure 1e where we enrolled 4 mice). Tumor size was measured every 4 days and calculated using the formula: $V = ((d/2)^2 \times (D/2))/2$ (d = minor tumor axis; D = major tumor axis) and reported as tumor mass volume (mm^3, mean ± SEM of individual tumor volume). The same calculations were applied to measure PDAC and CT26 orthotopic tumors after explant. The investigators were not blinded and the measures were acquired before the identification of the cages. No statistical methods were used to predetermine sample size. For raw data of mouse experiment see Supplementary Table 2.

Gene editing

To knockout the *Mlh1* we used the genome editing one vector system (lentiCRISPR-v2) (Addgene #52961). sgRNAs were designed using the CRISPR tool http://crispr.mit.edu to minimize potential off-target effects. The following sgRNA sequences were used: sgRNA2: TCACCGTGATCAGGGTGCCC; sgRNA3: CAACCAGGGCACCCTGATCA; sgRNA6: ATTGGCAAGCATAAGCCATG. Annealed sgRNA oligonucleotides targeting the murine *Mlh1* were cloned into Bsmbl **lentiCRISPR-v2** plasmid as described previously [34] . Lentiviral particles were generated by co-transfection of HEK293T cells with the viral vector and packaging plasmids **pCMV-VSV-G** (Addgene #8454), **psPAX2** (Addgene #12260) [34]. Supernatant from transfected HEK293T was harvested, passed through a 0.22 µm filter to remove cell debris, and frozen as 1 mL aliquots at -80° C. Cells were infected with lentivirus at approximately 60 % confluence in the presence of 8 µg/mL polybrene (Millipore). Puromycin (Sigma Aldrich) was used to select CRISPR/Cas9 infected cells. To identify *Mlh1* Knockout clones, infected populations were single-cell cloned in 96 well plates. Approximately thirty clones for each cell models were selected from 96-well plates and the presence of *MLH1* was verified by western blot analysis. For transient

expression of CRISPR-Cas9 system, we transfected cells with **lentiCRISPR-v2** vector plasmid (same guides as previously described). Transfection was carried out using Lipofectamine 3000 (Life technologies) and Opti-MEM (Invitrogen), according to the manufacturer's instructions. After 48h cells were incubated with puromycin (Sigma Aldrich) for 4 days and subsequently single cell diluted in 96-well plates. We selected clones lacking *Mlh1* and confirmed absence of Cas9 based on western blot.

Microsatellite instability analysis

Microsatellite instability in mouse cells was determined using a panel of three microsatellite markers as previously described [7]. Amplification was performed with following labeled primers. Fluorescein mBat64 Fw: GCCCACACTCCTGAAAACAGTCAT; Rev: CCCTGGTGTGGCAACTTTAAGC, AC096777 JOE Fw: TCCCTGTATAACCCTGGCTGACT; Rev: GCAACCAGTTGTCCTGGCGTGGA; AA003063 Tamra FW: ACGTCAAAAATCAATGTTAGG; Rev: CAGCAAGGGTCCCTGTCTTA; U12235 JOE FW GCTCATCTTCGTTCCCTGTC and Rev: CATTCGGTGGAAAGCTCTGA; L24372 Fluorescein FW GGGAAGACTGCTTAGGGAAGA and Rev ATTTGGCTTTCAAGCATCCATA. The PCR reaction was performed in 20ul of PCR reaction using Platinum Taq Polymerase Kit from Invitrogen and 20 ng of DNA. The cycling profile: 1 cycle 94^0C for 4 min, then 35 cycles of 94^0C for 30 sec, 56^0C for 45 sec and 72^0C for 30 sec. A final extension at 72^0C for 6 min completed the amplification. PCR fragments were separated on a 3730 DNA analyzer (Applied Biosystems) and raw data analyzed with GeneMapper software.

Drug screening and resistance protocols

Bendamustine, lomustine, pemetrexed, cisplatin, gemcitabine were purchased from Selleckchem. Carmustine, SN38, chlorambucil were obtained from Sigma Aldrich; temozolomide from Carbosynth. Oxaliplatin and 5-Fluorouracil were obtained from the pharmacy of Candiolo Cancer Institute. MLH1 proficient and deficient cell lines were seeded at different densities (1-1.5 × 10^3 cells/well) in 500 µl complete growth medium in 24-well plastic culture plates at day 0. The following day, serial drug dilutions were added to the cells in serum-free medium; DMSO-only treated cells were included as controls. Plates were incubated at 37°C in 5% CO2 for 5 or 7 days, after

which cells were fixed with 3% paraformaldehyde and stained with 1% Crystal Violet-Methanol solution (Sigma-Aldrich). Crystal Violet was then solubilized with 10% acetic acid and absorbance was quantified at 595nm. For TMZ treated human cell lines plates were incubated between 10 and 21 days until confluency of untreated cells was reached. All assays were performed independently at least two times. CT26 and MC38 temozolomide-resistant derivatives were obtained treating parental (sensitive) cells with 100 μM of temozolomide for the first three months. The concentration of temozolomide was subsequently raised to 500 μM until resistant derivatives emerged. SKCO1, RCM1 and SW620 were cultured with 100 μM of TMZ for three months until they acquired resistance.

Mice treatments

The anti-mouse PD-1 (clone RMP1-14), anti-mouse CTLA-4 (clone 9H10), anti-mouse CD8a (YTS169.4), Rat IgG2a, polyclonal Syrian Hamster IgG and rat IgG2b were purchased from BioXell (USA). Randomization was used for the experiments where therapeutic effects had to be evaluated (e.g. anti-PD1 and anti-CTLA4). Mice were treated i.p. with 250 ug/mouse of anti PD-1 and 200 ug/mouse of anti CTLA-4. Treatments were administrated at days 3, 6 and 9 after initial cell injection or they started when the size was between 700 mm^3 and 1000 mm^3 for the experiments in Figure 2. Anti PD-1 was given continuously every three days. Isotype controls were injected according to the same schedule. Anti-mouse CD8a was used for depleting cytotoxic T cells in immune-competent mice. Anti-mouse CD8a antibodies (200ug/mouse) were injected i.p. the same day of tumor inoculation. Two and three days post tumor injection mice were treated with 100ug/mouse of the CD8a antibody. Flow cytometry analysis was performed, every three days, to assess the level of CD8+ cells in the bloodstream of mice. The fraction of CD8+ cells relative to CD3+ cells was 20% before the depletion and 0,5 % after the administration of depleting antibody. The level was maintained throughout the entire experiment.

Western Blot analysis

For biochemical analysis, cells were grown in media supplemented with 10% FBS. Total cellular proteins were extracted by solubilizing the cells in boiling SDS buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, and 1% SDS). Samples were boiled for 10 minutes and sonicated for 30 seconds. Extracts were clarified by centrifugation and

amounts of proteins were normalized with the BCA Protein Assay Reagent Kit (Thermo Scientific). Western blot detection was performed with the enhanced chemiluminescence system (GE Healthcare) and peroxidase-conjugated secondary antibodies (Amersham). The following primary antibodies were used for Western blotting: anti mMLH1 (epr3894 from AbCam), anti Actin (I-19) from Santa Cruz Biotechnology, anti Cas9 (7A9) from GeneTex, anti-MGMT MT3.1 from Millipore. For western blot source data see Supplementary Figure 1.

Immuno-phenotypic cell analysis

Mouse tumors were cut into small pieces, disaggregated with collagenase (1.5 mg/ml) and DNAse (100ug/ml), and filtered through strainers. Cells (10^6) were stained with specific antibodies and Zombie Violet Fixable Viability Kit (Biolegend). Flow cytometry was performed using the FACS Dako instrument and FlowJo software. Phenotype analysis was performed with the following antibodies purchased from Biolegend: PerCp- Rat CD45 (30F11), Rat APC CD11b (M1/70), Rat PE/Cy7 CD3 (17A2), FITC Rat CD4 (RM4-5) and PE Rat CD8 (YTS156.7.7). Detection of cytotoxic T cells by Immunofluorescence was performed as described previously [35]. Briefly, tumor samples were included in killik (Bio-optica), serially cut (10 μm) and fixed using cold acetone/methanol (1:1). Samples were incubated for one hour in blocking buffer (1% BSA and 2% of goat serum in PBS with 0,05% of Tween and 0,1% of Triton) and incubated over-night with 1:100 of anti-CD8 (clone YTS169 from Thermoshientific). For detection 1:500 of anti-rat Alexa Fluor 488 was used (ThermoFisher). For Real Time PCR we negative selected CD3 T cells from spleen of tumor bearing mice by PAN T cells Isolation Kit II (Miltenyi Biotec). RNA was extracted by using miRNeasy Mini Kit (Qiagen), according to the manufacturer's protocol. Synthesis of cDNA was performed using the Reverse Transcription System (Promega). All samples within an experiment were reverse transcribed at the same time. Real-time PCR was performed using the GoTaq qPCR Master Mix, an ABI Prism 7700 thermocycler with fluorescence detection (Applied Biosystems) and the following primers as previously reported[36] :

GranzymeB: Fw CTGCTAAAGCTGAAGAGTAAGG and Rev TTTAAAGTAG-GACTCACACTCCC.

IFNg Fw: AAGTTTGAGGTCAACAACCCAC and Rev: GCTGG-CAGAATTATTCTTATTGGG.

Perforin1 Fw: GGGACTTCAGCTTTCCAGAG
and Rev: GTAGTCACATCCATGCCTTCC.
TBX21 Fw: CAGAACGCAGAGATCACTCAG
and Rev: AGGATACTGGTTGGATAGAAGAG.
IL21 Fw: ACATAGCTAAATGCCCTTCCT
and Rev: ATTCGTGAGCGTCTATAGTGTC.

Mutational loads and neoantigen prediction analysis in cell lines

Genomic DNA was extracted using ReliaPrep™ gDNA KIT (Promega). Whole exome sequencing was performed at Integragen (Evry, France). Libraries were sequenced using Illumina HiSeq 4000. The bioinformatics analysis was performed at our institution (IRCC) on exome sequencing data provided by Integragen. Raw data, in FastQ format, were initially de-multiplexed using CASAVA 1.8 software as paired-end 75-bp reads. On average we observed a median depth of 70x (84x for human), with more than 97% of targeted-region covered by at least one read. Before further analysis, pair-end reads were aligned to the mouse and human references, assembly mm10 and hg38, respectively, using BWA-mem algorithm [37]. Next, PCR duplicates were removed from the alignment files using the "rmdup" samtools command [38]. An NGS pipeline previously developed [39] by our laboratory was used to identify Single Nucleotide Variants (SNVs) and Indels. For methylation of human cell lines we used the β-value of cg12434587 (MGMT probe assessed by the methylation specific PCR used in patients) from the InfiniumHumanMethylation450 (HM450). For murine samples, somatic variations were called subtracting germline variations present in BalbC and C57Bl6 (downloaded from "Mouse Genome Project" http://www.sanger.ac.uk/science/data/mouse-genomes-project). For human TMZ resistant cell lines, somatic variations were called subtracting alterations found in the parental (sensitive) counterpart. Only positions present with a minimum depth of 5x (9x for human) and supported by at least 1% allelic frequency were considered. To calculate the significance of the allele's frequency a Fischer test was performed for each variant. The mutational burden (number of variants/Mb) was calculated considering only coding variants normalized on the targeted region for each data point. Predicted neoantigens were calculated starting from the file of coding variations, annotated and filtered for gene expression values (expected count > 10) using RNAseq data

of the same sample (see below). For each variation, the mutant peptide sequence was obtained: for SNV we introduced the new altered amino acid in the candidate peptide, for frameshift we took the newly generated frame. Mutated peptide sequences were trimmed and then fed to NetMHC 4.0 using kmer of 8-11 length [40]. Haplotypes for murine samples were set to H2-Kd and H2-Dd for BalbC and H2-Db and H2-Kb for C57Bl6 background. For human cell lines the haplotypes were HLA-A0101, HLA-A0301 and HLA-B0702 for SKCO1. HLA-A0201 and HLA-B0801 for SW620. HLA-A0201, HLA-A0301, HLA-B3503, HLA-B4403, HLA-C0401, HLA-C0501 for RCM1 cell line. Predicted neoantigens were filtered by a rank threshold of 0.5. Alterations that produced more than one predicted neoantigen were clustered through a custom script (exploiting the Levenshtein distance) to create a consensus family. For each family the peptide with the best rank was considered.

RNAseq analysis

To extrapolate expressed neo antigens, we performed RNA sequencing of the MMR proficient and MMR deficient clones. The RNA concentration and integrity was evaluated with the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Kit. Total RNA (800ng) with RIN (RNA integrity number) score from 8 to 10 was used as input to the Illumina TruSeq RNA Sample Prep Kit v2-Set B (48Rxn) according to manufacturer protocol. The standard RNA fragmentation profile as recommended by Illumina was used (94 °C for 8 min for the TruSeq RNA Sample Prep Kit). PCR amplified RNA-Seq library quality was assessed using the Agilent DNA 1000 kit on the Agilent 2100 BioAnalyzer and quantified using Qubit 3.0 Fluorometer (LifeTechnologies). Libraries were diluted to 10nM using TrisHCl 10mM PH=8.5 and then pooled together. Diluted pools were denatured according to standard Illumina protocol and 1.8pM were run on NextSeq500 using High output Reagent cartridge V2 150 cycles. A single read 150 cycles run was performed. The transcriptome profile of RNAseq data were calculated using MapSplice v2.2.0 [41] and RSEM [42] software package. Genes, with at least 10 expected counts in the output file, were considered expressed.

T cell Receptor Sequencing

Murine TCR-β complementary determining regions 3 (CDR3) were amplified using the survey ImmunoSeq platform in a multiplex PCR method using 45 forward primers specific to TCR Vβ gene segments and 13 reverse primers specific to TCR Jβ gene

segments (Adaptive Biotechnologies®, Seattle, WA). Each sample was run in duplicate [43]. Samples plus controls were pooled in different libraries and sequenced on MiSeq or NextSeq500 (Illumina). Raw data were uploaded to the ImmunoSeq pipeline through the immunoSEQ Data Assistant to be analyzed by Adaptive Biotechnologies. For each sample the results indicate rearrangements with relative details, including amino acid sequences and number of templates. Amino acidic sequences, resulting from different rearrangements, were merged adding the relative number of productive templates.

Patients

Five metastatic colorectal cancer patients were considered for analysis. Biospecimens were collected in accordance with protocols approved by the review Boards of Fondazione IRCCS Istituto Nazionale dei Tumori and Policlinico Universitario Agostino Gemelli, to which the patient provided written informed consent, and all studies were conducted in accordance with the Declaration of Helsinki. All patients had *MGMT* promoter methylation determined by methylation specific-PCR and mismatch repair-proficient status assessed by both IHC and multiplex PCR. The analyses were performed on archival tumor tissue samples obtained prior to any treatment.

Three patients were treated within the CAPTEM trial (EudraCT number 2014-002417-36; *NCT02414009*) and received second-line therapy with capecitabine 750 mg/m^2/day b.i.d. on days 1-14 plus temozolomide 75 mg/m^2/day b.i.d. on days 10-14, every 28 days. The remaining two patients were treated in the setting of refractory disease. One received out-trial/off-label treatment with TEMIRI schedule (temozolomide 150 mg/m^2/day on days 1-5 and irinotecan 100 mg/m^2 on days 1,15, every 28 days), whilst the other was treated with temozolomide monotherapy at 150 mg/m2/day on days 1-5, every 28 days (EudraCT number 2012-002766-13).

Next Generation Sequencing in tumor samples

Libraries from peripheral blood mononucleated cells (PBMC) and fresh tissues were prepared using Nextera Rapid Capture Exome kit (Illumina, Inc., San Diego, CA), according to manufacturer's protocol. In particular, 100 ng of gDNA were fragmented using transposons, adding simultaneously adapter sequences. Purified gDNA after the tagmentation step was amplified with the indexing PCR, during which unique sample barcodes are inserted. DNA fragments' size distribution was assessed by means of

2100 Bioanalyzer with High Sensitivity DNA assay Kit (Agilent Technologies, Santa Clara, CA) and equal amount of DNA libraries were pooled for the subsequent step of targeted hybridization capture. DNA extracted from formalin-fixed paraffin-embedded (FFPE) biopsies was treated with TruSeq Exome kit (Illumina, Inc., San Diego, CA). Briefly, up to 300 ng of FFPE-derived DNA were sheared with S220 sonicator (Covaris®, Inc., Woburn, Massachusetts, USA.) by means of specific settings (duty factor 10%, peak power 140, cycles/burst 200, temperature 12°C, seconds 120) in order to obtain the optimal fragments' size distribution. Clean-up steps after end-repair, A-tailing and adapter ligation were optimized to increase the quantity and the quality of the libraries before pooling them for targeted hybridization capture. Libraries were sequenced on the Illumina NextSeq500 sequencer (Illumina, Inc.) and 150 cycles paired end reads were generated.

Mutational Load in Patient Samples

Mutational discovery analysis was performed with an updated version of a NGS pipeline we previously developed[39]. The approach involves identification of somatic SNV and indel by subtracting germline variants found in matched normal samples. FastQ were pre-processed to remove adapter sequences and were mapped to assembly hg19 of human reference genome using BWA-mem algorithm [37]. The "rmdup" command of SAMtools package was used to remove PCR duplicates [38]. To delete sequencing artifacts [44] , that could alter the mutational burden in FFPE and fresh tissue, high depth WES was performed. Moreover sequencing data were further filtered obtaining a final median depth of 236X, with more than 96% of targeted-region covered by at least 10 read. Nucleotides with a Phred Score < 30 and reads carrying more than 3 mismatches were discarded. We further discarded alterations supported by reads carrying a strand bias or displaying mismatches in head/tail of the read and genes with more than 10 variations. Furthermore, only mutations with 5% significance level obtained with a Fisher test and at least 1% allelic frequency were considered for the analysis. The mutational burden was calculated normalizing on the covered target region and measured considering somatic mutations supported by a minimum of 9 mutated reads in regions with a minimum depth of 10X.

Statistical Analyses and Reproducibility

Statistical analyses were performed using the GraphPad Prizm software. During exome analysis the Fischer test was performed to calculate the significance of allele's frequency. For tissue culture experiments, statistical differences were calculated using paired Student's *t*-test. To determine statistical significance in tumor growth, an unpaired, two tailed Student's *t*-test was used. Lifespan of mice were analyzed with a Kaplan-Meier survival curve and P values were obtained by a log-rank (Mantel-Cox) test. *P* value of <0.05 was considered significant. All data were presented as mean ± s.e.m. with the exception of Extended Data Figure 4 and 9 where we showed mean ± s.d. We did not perform assumptions or corrections during the interpretation of results.

The *in vitro* assays were performed at least twice. Sample sizes were chosen with adequate power, based on our previous studies and literature information for in vitro experiments. The number of replicates and sample size for in vivo experiments were limited according to requirements from the Italian Ministry of Health. Animal studies were performed in accordance with institutional guidelines and international law and policies. When therapy was applied we performed randomization. In this case tumor free mice or mice with a tumor larger than 50% of the average were excluded from the experiment. The investigators were not blinded. The measures were acquired before identification of the cages.

Data Availability

Data generated during our study are available in the European Nucleotide Archive (ENA) with the following accession code PRJEB22901. DNA sequencing data corresponding to patient's samples and human cell lines are available in the European Genome-phenome Archive (EGA) with the following accession code EGAS00001002694.

References of M&M Section

31 Corbett, T. H., Griswold, D. P., Roberts, B. J., Peckham, J. C. & Schabel, F. M. Tumor induction relationships in development of transplantable cancers of the colon in mice for chemotherapy assays, with a note on carcinogen structure. *Cancer Res* **35**, 2434-2439 (1975).

32 Nanni, P., de Giovanni, C., Lollini, P. L., Nicoletti, G. & Prodi, G. TS/A: a new metastasizing cell line from a BALB/c spontaneous mammary adenocarcinoma. *Clin Exp Metastasis* **1**, 373-380 (1983).

33 Gilles, M. E. *et al.* Nucleolin Targeting Impairs the Progression of Pancreatic Cancer and Promotes the Normalization of Tumor Vasculature. *Cancer Res* **76**, 7181-7193, doi:10.1158/0008-5472.CAN-16-0300 (2016).

34 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

35 Maione, F. *et al.* Semaphorin 3A overcomes cancer hypoxia and metastatic dissemination induced by antiangiogenic treatment in mice. *J Clin Invest* **122**, 1832-1848, doi:10.1172/JCI58976 (2012).

36 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, doi:10.1158/2159-8290.CD-14-0863 (2015).

37 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

38 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

39 Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med*, doi:10.1038/nm.3870 (2015).

40 Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511-517, doi:10.1093/bioinformatics/btv639 (2016).

41 Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* **38**, e178, doi:10.1093/nar/gkq622 (2010).

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

43 Gaide, O. *et al.* Common clonal origin of central and resident memory T cells following skin immunization. *Nat Med* **21**, 647-653, doi:10.1038/nm.3860 (2015).

44 Chen, L., Liu, P., Evans, T. C. & Ettwiller, L. M. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* **355**, 752-756, doi:10.1126/science.aai8690 (201

Figure Legends

Fig. 1: *Effect of MLH1-inactivation in colorectal and pancreatic cancer cells*. CRISPR/Cas9 mediated knock-out of *Mlh1* in CT26 and PDAC cell lines. Two independent guides (blue and orange lines) were used to exclude off-target effects. A non-target vector was used to generate control cells. **(a)** The indicated CT26 clones were injected, 132 days after genome editing, subcutaneously in Nod/Scid mice (n=7). **(b)** The same CT26 clones in (a) were injected in syngeneic BalbC mice (n=14). **(c)** CT26 clones were generated using a Cas9 transient system and injected (n=7 left panel, n=7 not sub-cloned, n=7 sub-cloned) in syngeneic BalbC mice at the indicated time points. . **(d)** CT26 lines generated with Cas9 transient system were injected 157 days post genome editing (n=5) in the caecum of BalbC animals and after four weeks mice were sacrificed and orthotopic tumors were measured. **(e)** *Mlh1* was inactivated in PDAC cells using CRISPR/Cas9 stable system. Two independent *Mlh1 KO* clones (blue and orange bars) were injected (n=4) in the pancreas of Nod/Scid mice 120 days later the genome editing. Animals were sacrificed after three weeks.. **(f)** The indicated PDAC clones were injected (n=6) in the pancreas of FVB/N syngeneic mice the same days of the experiment in (e). After three weeks mice were sacrificed. **(g)** The same experiment in (f) was performed using pancreatic cells generated with the Cas9 transient system, cells were sub-cloned and injected 150 days post genome editing (red and green bars, n=7). For CT26 we injected 5 X 10^5 cells per mouse, for PDAC 10^3 cells per mouse. For western blot source data see Supplementary Figure 1. Data are mean ± s.e.m., (independent samples per group). We performed each experiment at least two times with the exception of the pancreatic orthotopic models showed in (d) and (f). Statistical analysis: Two-tailed Student's t-test (n.s.: not significantly different).

Fig. 2 *Impact of MMR inactivation on treatment with immuno-modulatory antibodies*. **(a)** The indicated CT26 clones were initially established in Nod/Scid mice until they formed tumors of 1500 mm^3 in size. At this point tumors were explanted, and fragments of 5mm^3 were implanted subcutaneously in BalbC animals (n=7). When the size of tumor was between 700 and 900 mm^3, anti-PD1 (250 µg per mouse) and anti-CTLA-4 (200 µg per mouse) were administrated i.p. for four times every three days. After that, therapy continued with anti PD-1 every three days. Dotted lines indicate days of combinatorial treatment. **(b)** The experiment presented in figure (a) was per-

formed with clones lacking Cas9 expression (n=6). **(c)** The indicated CT26 clones were injected (5 X 10^5 cells per mouse, n=7) and the same day mice were treated with anti CD8 depleting antibody (400 µg per mouse at day 0, 100 µg per mouse at day 1 and 100 µg at the day 2). Isotype antibodies served as control. Mean ± s.e.m., six independent samples per group. (a) and (b) representative of two independent experiments. The anti CD8 depletion was carried out in one experiment. For all the experiments the statistical analysis was a two-tailed Student's t-test.

Fig. 3 *Measurements of neoantigen loads and TCR profiles.*
(a) Exome data of the indicated cell models were compared. Coding variants identified by exome sequencing were used for calculating mutants and predicted neoantigens as described in details in the method sections. Private events were defined as predicted neoantigens present only at one specific time point. Shared neoantigens were present in two time points. Common neoantigens were present in all time points. The numbers of predicted private and common neoantigens are indicated. The first time point corresponds to level of mutations after the establishment of *Mlh1 KO* clones (30 days). The second time point corresponds to approximately 130 days. The third time point has been sequenced after 246 days post *Mlh1 KO.*
(b) Distribution of the twenty most frequent TCR rearrangements identified in peripheral blood from mice (n=4) injected with the indicated CT26 clones. The width of the violins is proportional to the number of TCR templates in each y-level; the bars inside the violins show the quartiles of the 20 templates; the white bullet shows the median value. TCR analysis was performed on blood samples obtained 13 days after injection of the tumor cells as described in the methods.

Extended Data

Extended Data Fig. 1 *Microsatellite profiles of mouse cancer cell lines and Mlh1 alterations in genome-edited CT26 cells*. **(a)** The microsatellite profile of the indicated cell lines was compared to the germline DNA of the corresponding mouse strain as described in the methods. CT26 and TS/A were compared to BalbC. MC38 was compared to C57Bl6 and PDAC was compared to FVB/N. **(b)** Exome sequencing data

were used to identify molecular alterations of the *Mlh1* sequence in the indicated cell models. Alignment of sequence reads showed deletions of *Mlh1* gene upon gene editing using sgRNA2 and sgRNA3 guides. The clone generated by editing with guide 2 (KO-1) carried a deletion of 8 bps that induced multiple frame-shifts. Guide 3 (KO-2) produced a deletion resulting in a frame-shift of 9 codons. The two clones obtained with the guide 3 (KO-3 and KO-4) showed also 2 insertions and 2 deletions for KO-3 and 2 insertions for KO-4. The effects of those changes are frameshifts. The upper sequence corresponds to the mouse reference assembly mm10. **(c)** Western-blot analysis of the indicated CT26 and PDAC cell lines. For western blot source data see Supplementary Figure 1. We showed a representative experiment of routinely western blots performed for validating our gene editing.

Extended Data Fig. 2 *Effect of Mlh1-inactivation in MC38 and (TS/A) a breast adenocarcinoma cell line.* **(a)** Expression of the MLH1 protein in MC38 control and the indicated *Mlh1* knockout clones. The indicated MC38 cell models were injected s.c. (10^5 cells) in Nod/Scid (n=5, left panel) and in syngeneic C57/Bl6 mice (n=5, right panel). Mean ± s.e.m., independent samples. Statistical analysis: two-tailed Student's t-test. **(b)** TS/A cells were *Mlh1* gene edited with the indicated guides. Cells were injected, 120 days after the establishment of the KO subcutaneously in Nod/Scid mice (5 X 10^5 cells per mouse, n=7) and the growth was monitored until sacrifice (left panel). The same TS/A were injected in syngeneic BalbC mice subcutaneously (middle panel, n=5) and orthotopically (right panel, n=7). For western blot source data see Supplementary Figure 1. The growth was monitored until sacrifice. Mean ± s.e.m.. Samples were independent. Statistical analysis: Two-tailed Student's t-test. Western blots in (a) and (c) are representative of at least two experiments performed to confirm the gene editing.

Extended Data Fig. 3 *Impact of Mlh1 inactivation on microsatellite instability in mouse tumor cell lines*. The MSI status was evaluated comparing mononucleotide repeats of CT26 (upper left part), MC38 (lower left part), PDAC (upper right part) and TS/A (lower right part). The mononucleotide regions, Bat64, L24372-A27 and U12235-A24, were used to evaluate microsatellite instability.

Extended Data Fig. 4 *Growth of CT26, MC38, TS/A and PDAC cell lines*.
The growth of the indicated cell models was measured with Cell Titer Glo at the indicated time points. Arbitrary Units represent the ratio between the absorbance at every time point and at time zero. All data are presented as mean (six technical replicate in the same experiment) ± s.d. CT26, MC38, TS/A were plated 1000 cells per well. PDAC were plated at 5000 cells per well. **(a)** The growth of constitutively Cas9-expressing CT26, MC38, TS/A and PDAC was measured as described. **(b)** The same was done for transiently Cas9-expressing CT26 and PDAC. The MSI status was evaluated as in the Extended Data Fig.3 for CT26, and PDAC generated with the Cas9 transient expression system. The mononucleotide regions, Bat64 and U12235-A24, were used to evaluate microsatellite instability. **(c)** The indicated CT26 clones, generated with the constitutive and transient Cas9 systems, were tested for Cas9 expression by western blot analysis. To verify Cas9-loss we performed one experiment. For western blot source data see Supplementary Figure 1.

Extended Data Fig. 5 *T cell infiltration in tumors generated with constitutively and transiently Cas9-expressing CT26 cells*
(a) Immunofluorescence of CD8+ cells in control and *Mlh1 KO* clones showed in (Figure 2a). Staining was performed on *Mlh1* proficient and deficient tumor to assess CD8 and IFNγ levels. (**b**) Immune-infiltrates (CD45+ CD4+ ad CD8+) were measured by FACS analyses in the indicated tumor samples (n=5). The percentage of CD45+ cells relative to total events of live cells (one million of events were acquired per sample) is shown. The percentage of CD4 and CD8 positive cells was relative to CD45+ cells. **(c)** Immune-infiltrates (CD4+ ad CD8+) were measured by FACS analyses in the indicated samples (3 mice per group). The percentage of CD8+ and CD4+ T cells is calculated relative to total live CD45 positive cells. **(d)** Immunofluorescence of DAPI and CD8 T cells in the indicated tumor samples. In (a) and (d) the images are representative sections from one mouse. The staining was performed on four independent mice. **(e)** Percentage of CD45+ (relative to total live cells), CD8+ and CD4+ (relative to CD45+ live cells) is shown for all mice included in the experimental arm (5 mice per group). For (b) and (e) we analyzed the data with the 2 way ANOVA. All data are presented as mean ± s.d. P value represents the significance.

Extended Data Fig. 6 *In vitro neoantigen evolution of constitutively Cas9-expressing CT26.* Exome data of the indicated CT26 cells were analyzed over time. Coding variants identified by exome sequencing were calculated as in Fig. 3a and described in the methods section. Private, shared and common neoantigens are defined also as in Fig. 3a. The first time point corresponds to 30 days after KO establishment. Time point 2 and 3, 120 and 210 days post *Mlh1* KO.

Extended Data Fig. 7 *Clonal and sub-clonal mutational and neoantigen profiles in CT26 cells*.
(a) Mutational load of the not sub-cloned and sub-cloned CT26 calculated 246 and 133 days post *Mlh1* KO, respectively. The number of mutations was obtained considering SNV and Indels. Each bar shows the mutations/Mb and the expressed private neoantigens obtained from a single exome sequencing of the represented clones. **(b)** Allele frequency distribution of single nucleotide variants and frameshifts of the indicated clones. Each violin represents a clone of CT26 cell line. **(c)** Number of private and shared neoantigens (SNV and Indels) before injection of CT26 Ctrl and *Mlh1* KO (pre-injection) in syngeneic animals and 20 days later (post-injection). (**d**) Distribution of the twenty most frequent TCR rearrangements identified in peripheral blood from four mice injected with the indicated CT26 clones. The TCR analysis was performed on blood samples obtained 13 days after injection of the tumor cells as described in the methods. The interpretation of the violin plot is described in the legend of the figure 3b.

Extended Data Fig. 8 *Effects of pharmacological agents on CT26 and MC38 clones*. CT26 **(a)** and MC38 **(b)** were plated in complete media in 24 well plates at 1000 cells per well. The day after, drugs were added in serum free media. After 7 days cells were fixed and stained with crystal violet**.** Crystal violet was then dissolved and quantified by spectrophotometer. All data are presented as mean (four technical replicates in the same experiment) ± s.d. The figure is representative of at least two independent experiments.

Extended Data Fig. 9 *Temozolomide-dependent effects on murine cell lines.*
(a) Tumor forming ability of CT26 (left panel) and MC38 (right panel) cells treated or not with temozolomide. Cells were injected s.c. (5×10^5 cell per mouse, n=5 upper

panel and n=4 lower panel) in syngeneic mice. Mean ± s.e.m., (independent samples, a representative experiment of two performed). The statistical analysis applied was a two-tailed Student's t-test. **(b)** Exome data of the indicated cell models were compared to assess mutational loads (SNV and Indels) and predicted neoantigens. Mutations per Mb (left panel) and predicted private neoantigens (right panel) are listed. **(c)** CT26 and MC38 were treated for four months with 100uM of temozolomide until resistant populations emerged; at this point DNA was extracted and cells were analyzed for microsatellite status. The corresponding cell lines before TMZ treatment were used as comparison. The mononucleotide regions Bat64, AA003063-A23 and U12235-A24 were used to evaluate microsatellite instability. **(d)** MLH1 protein expression in CT26 and MC38 cells before and after temozolomide exposure (representative experiment of two performed). **(e)** Exome sequencing data were used to identify molecular alterations of the *Mlh1* sequence. Alignment of sequence reads showed a deletion of 5 and 19 bps, respectively, that generated a frameshift and a premature termination codon (p.D64fs*39 c.187delGACAA, p.I59fs*78 c.174delAATTCAGATCCAAGACAAT)

Extended Data Fig. 10 *Temozolomide-dependent effects on human cancer cells.*
(a) 47 CRC cell lines were tested with TMZ in long-term colony forming assays, from which IC50 values were obtained. *MGMT* promoter methylation status (from microarray probe number cg12434587), gene expression (normalized Z-score) and MSI status of each cell line are also annotated. ** IC50 obtained through dissolution of clonogenic assay cristal violet staining and assessment of the absorbance. The dot-line corresponds to 12.5 μM, which is the plasmatic concentration reported in patients. **(b)** The indicated CRC cell lines (before and after TMZ resistance) were tested for MGMT level by western blot. The MGMT expression was representative of at least two experiments. For western blot source data see Supplementary Figure 1. **(c)** Sensitivity to temozolomide treatment of six cell lines before and after the acquisition of drug resistance. **(d)** Mutations per Mb in parental and TMZ resistant cells at the indicated time point (left panel). Predicted private neoantigens found only in cells harvested at 20 days (right panel). The table lists the variations of MMR genes found only in TMZ resistant cells. **(e)** Clinical characteristics of mCRC patients with acquired resistance to TMZ treatment are indicated. All patients had histologically confirmed mCRC with *MGMT* promoter methylation assessed by methylation specific-PCR and

mismatch repair-proficient (pMMR) status assessed by both immunohistochemistry and multiplex PCR (all analyses were carried out as per standard practice using archival tumor samples obtained prior to any treatment). List of abbreviations: CAPTEM, capecitabine plus temozolomide (*NCT02414009*); TEMIRI, temozolomide plus irinotecan; SD: stable disease; PR: partial response. **(f)** MGMT immunohistochemical expression and mutational load (mutations per Mb) in tissue biopsies of metastatic CRC patients before and after TMZ-based therapy. The tables list alterations in MMR genes present only in tissue biopsies obtained at progression post TMZ treatment. (n.a.: data not available). Cell lines names, patients' number, genes, nucleotide and amino acidic changes, impact of variations and allele frequencies are listed. Fisher test was performed to calculate the significance ($p<0.05$).

Acknowledgments

We thank B. Van Emburgh, A. Sogari, C. Falcomatà, V. Amodio and C. Cancelliere for helpful assistance with in vitro experiments, F. Sassi and S. Giove for help with the immunohistochemistry, M. Buscarino for suggestions on MSI analysis, D. Cantarella for performing RNASeq, G. Corti and L. Novara for helpful bioinformatic analysis and comments, D. Sangiolo and M. Rescigno for helpful discussion. D. Hananhan and K. Shchors for providing the PDAC cell model. C. Torrance and J. Roix for scientific support, suggestions and critically reading the manuscript. This study was supported by European Community's Seventh Framework Programme under grant agreement no. 602901 MErCuRIC (A.B. e G.G.); no. 635342-2 MoTriColor (A.B. and S.S.); IMI contract n. 115749 CANCER-ID (A.B.); AIRC 2010 Special Program Molecular Clinical Oncology 5 per mille, Project n. 9970 (A.B. and S.S.); AIRC IG n. 16788 (A.B.); AIRC IG n. 17707 (F.D.N.); Fondazione Piemontese per la Ricerca sul Cancro-ONLUS 5 per mille 2011 Ministero della Salute (A.B. and F.D.N.); AIRC-IG Grant n:15645 and Swiss National Science Foundation (SNSF), Sinergia Grant (CRSII3 160742/1) (E.G.); AIRC and the EU under a Marie Curie COFUND (G.G.); and grant Terapia Molecolare Tumori by Fondazione Oncologia Niguarda Onlus (A.S.B. and S.S.). Merck grant for Oncology Innovation -GOI- 2016 (A.B.).

Author Contributions

G.G. and A.B., conceived the study. G.G., A.M., F.M., R.F. and M.M. performed animal experiments. A.Bartolini, B.M., G.S. and M.M. performed DNA analyses experiments. G.R., and G.C. conducted bioinformatics data analyses. M.R. performed cell based drug screens. L.B. designed and performed the temozolomide experiments on human cells. S.L. performed gene-editing experiments. G.G. and G.L., E.G. and N.A., conducted studies on mouse cells. F.M., F.D., A.O., F.P., S.S., A.S.B designed and performed clinical studies, tissue samples accrual and patient data analysis. S.M. and M.D., were involved in data analysis. G.G. and A.B. wrote the manuscript. A.B. and F.D.N. supervised the study. All authors read and approved the final submitted manuscript

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. (alberto.bardelli@unito.it)

References

1 Lord, C. J. & Ashworth, A. The DNA damage response and cancer therapy. *Nature* **481**, 287-294, doi:10.1038/nature10760 (2012).

2 Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol* **7**, 153-162, doi:10.1038/nrclinonc.2009.237 (2010).

3 Orthwein, A. *et al.* A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422-426, doi:10.1038/nature16142 (2015).

4 Rayner, E. *et al.* A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat Rev Cancer* **16**, 71-81, doi:10.1038/nrc.2015.12 (2016).

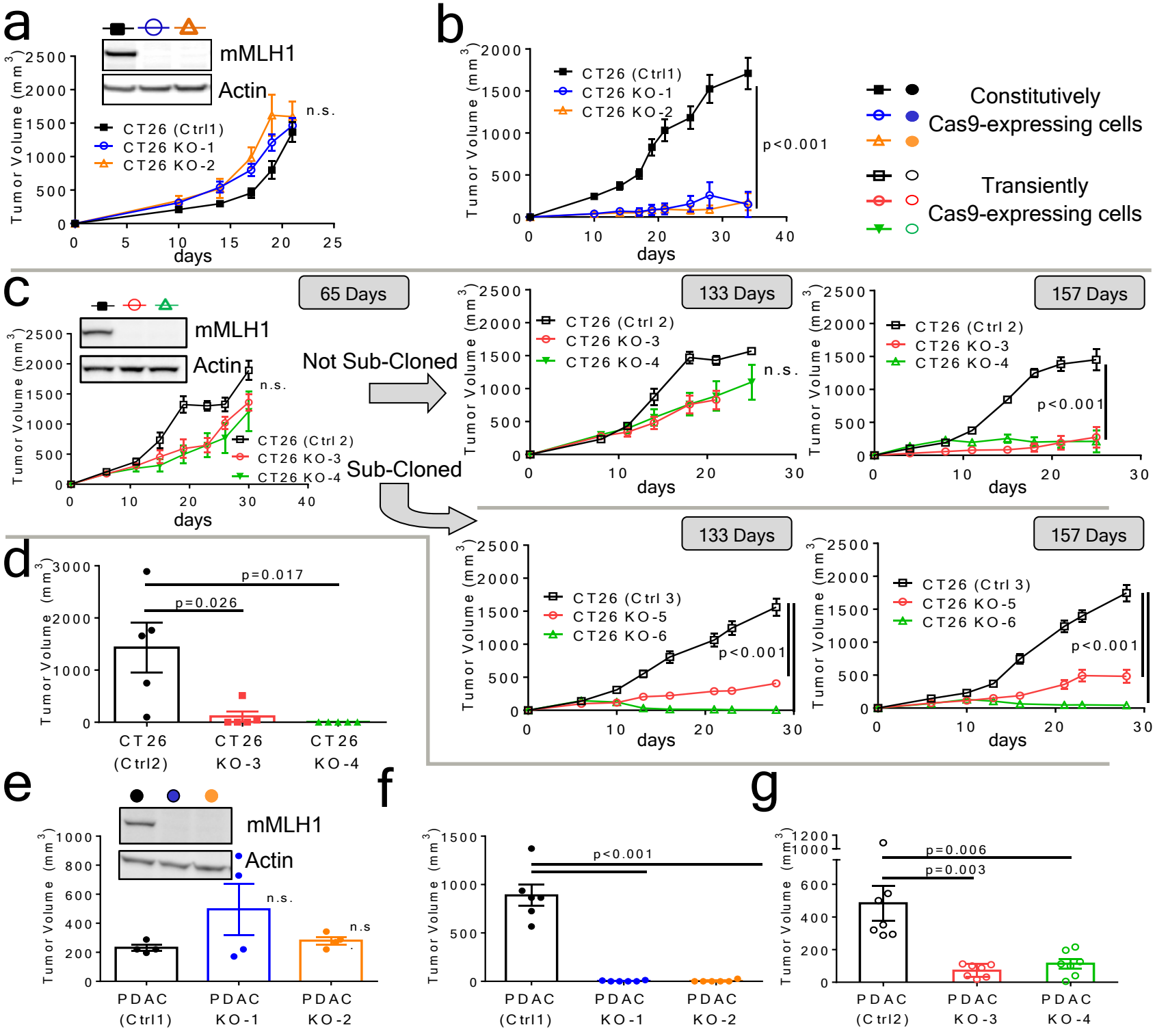
5 Ponti, G., Castellsagué, E., Ruini, C., Percesepe, A. & Tomasi, A. Mismatch repair genes founder mutations and cancer susceptibility in Lynch syndrome. *Clin Genet* **87**, 507-516, doi:10.1111/cge.12529 (2015).

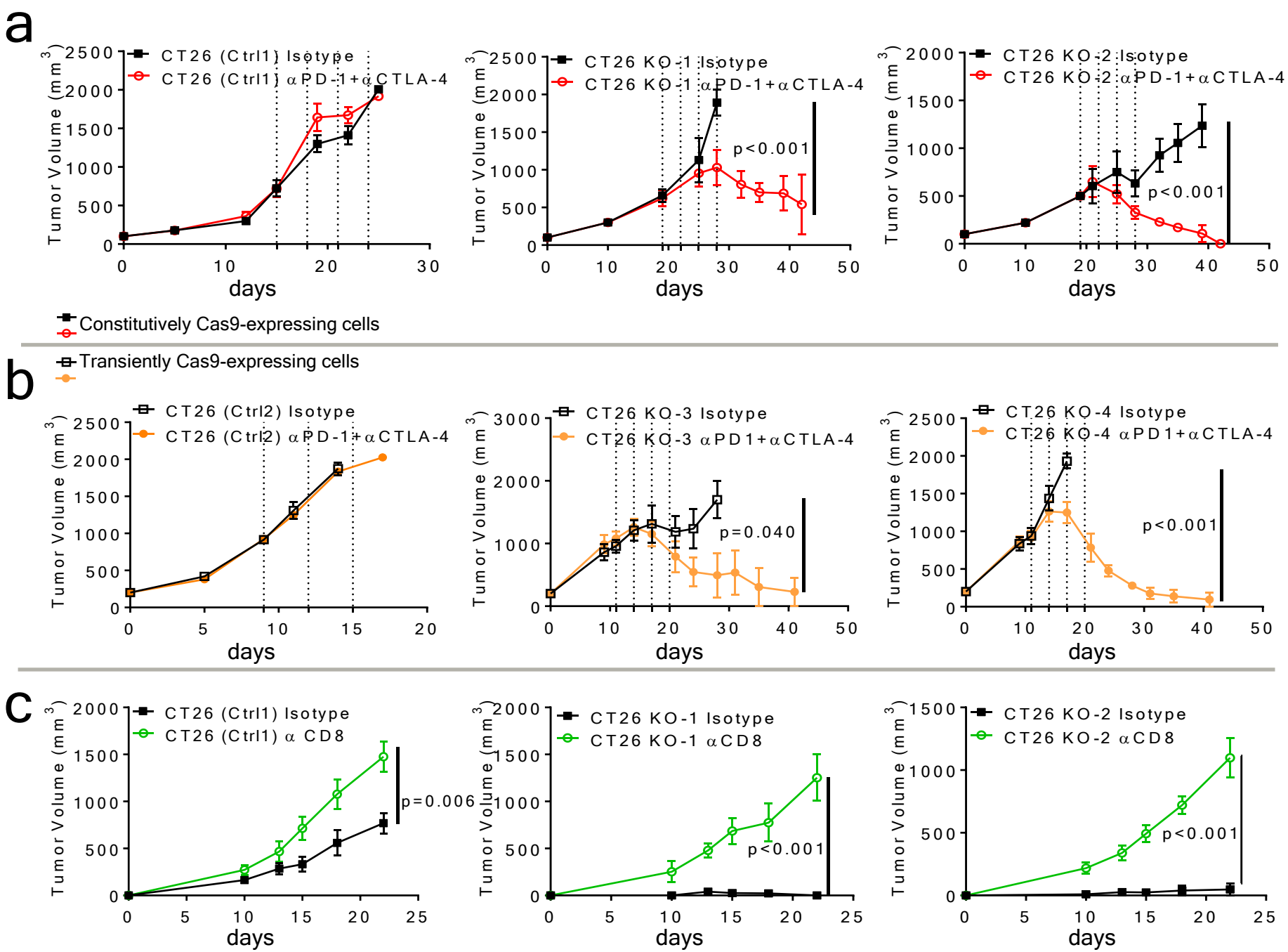
6 Jiricny, J. Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol* **5**, a012633, doi:10.1101/cshperspect.a012633 (2013).

7 Woerner, S. M. *et al.* Detection of coding microsatellite frameshift mutations in DNA mismatch repair-deficient mouse intestinal tumors. *Mol Carcinog* **54**, 1376-1386, doi:10.1002/mc.22213 (2015).

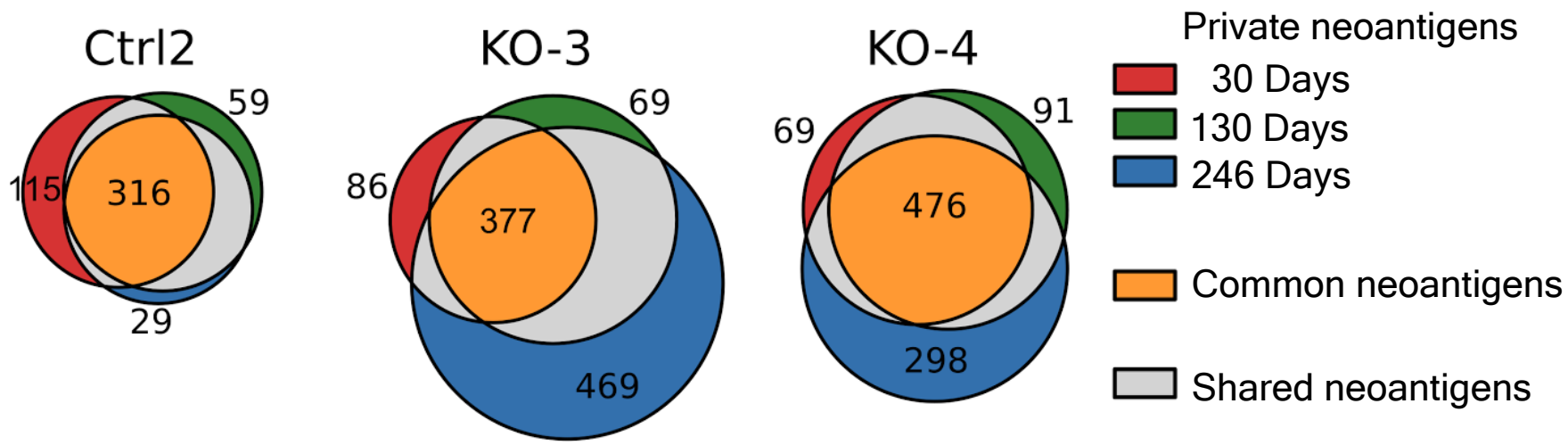
8 Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128, doi:science.aaa1348 [pii] 10.1126/science.aaa1348 (2015).
9 Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577-581, doi:10.1038/nature13988 (2014).
10 Le, D. T. *et al.* Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*, doi:10.1126/science.aan6733 (2017).
11 Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69-74, doi:10.1126/science.aaa4971 (2015).
12 Gros, A. *et al.* Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med* **22**, 433-438, doi:10.1038/nm.4051 (2016).
13 Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science* **348**, 56-61, doi:10.1126/science.aaa8172 (2015).
14 McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, doi:10.1126/science.aaf1490 (2016).
15 Juneja, V. R. *et al.* PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med* **214**, 895-904, doi:10.1084/jem.20160801 (2017).
16 Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* **13**, 868-874, doi:10.1038/nmeth.3993 (2016).
17 Bailey, J. M. *et al.* p53 mutations cooperate with oncogenic Kras to promote adenocarcinoma from pancreatic ductal cells. *Oncogene* **35**, 4282-4288, doi:10.1038/onc.2015.441 (2016).
18 Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520, doi:10.1056/NEJMoa1500596 (2015).
19 Castle, J. C. *et al.* Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190, doi:10.1186/1471-2164-15-190 (2014).
20 Bardelli, A. *et al.* Carcinogen-specific induction of genetic instability. *Proc Natl Acad Sci U S A* **98**, 5770-5775, doi:10.1073/pnas.081082898 (2001).
21 Fink, D., Aebi, S. & Howell, S. B. The role of DNA mismatch repair in drug resistance. *Clin Cancer Res* **4**, 1-6 (1998).
22 Cahill, D. P. *et al.* Loss of the mismatch repair protein MSH6 in human glioblastomas is associated with tumor progression during temozolomide treatment. *Clin Cancer Res* **13**, 2038-2045, doi:10.1158/1078-0432.CCR-06-2149 (2007).
23 Yip, S. *et al.* MSH6 mutations arise in glioblastomas during temozolomide therapy and mediate temozolomide resistance. *Clin Cancer Res* **15**, 4622-4629, doi:10.1158/1078-0432.CCR-08-3012 (2009).
24 Medico, E. *et al.* The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nat Commun* **6**, 7002, doi:10.1038/ncomms8002 (2015).
25 Sartore-Bianchi, A. *et al.* Digital PCR assessment of MGMT promoter methylation coupled with reduced protein expression optimises

prediction of response to alkylating agents in metastatic colorectal cancer patients. *Eur J Cancer* **71**, 43-50, doi:10.1016/j.ejca.2016.10.032 (2017).
26 Lee, S. Y. Vol. Volume 3, Issue 3 , September 2016, Pages 198–210 198-210 (Genes & Diseases, September 2016).
27 Cahill, D. P., Codd, P. J., Batchelor, T. T., Curry, W. T. & Louis, D. N. MSH6 inactivation and emergent temozolomide resistance in human glioblastomas. *Clin Neurosurg* **55**, 165-171 (2008).
28 Calegari, M. A. *et al.* A phase 2 study of temozolomide in pretreated metastatic colorectal cancer with MGMT promoter methylation. *Br J Cancer* **116**, 1279-1286, doi:10.1038/bjc.2017.109 (2017).
29 Bykov, V. J. *et al.* Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nat Med* **8**, 282-288, doi:10.1038/nm0302-282 (2002).
30 Maletzki, C., Schmidt, F., Dirks, W. G., Schmitt, M. & Linnebacher, M. Frameshift-derived neoantigens constitute immunotherapeutic targets for patients with microsatellite-instable haematological malignancies: frameshift peptides for treating MSI+ blood cancers. *Eur J Cancer* **49**, 2587-2595, doi:10.1016/j.ejca.2013.02.035 (2013).

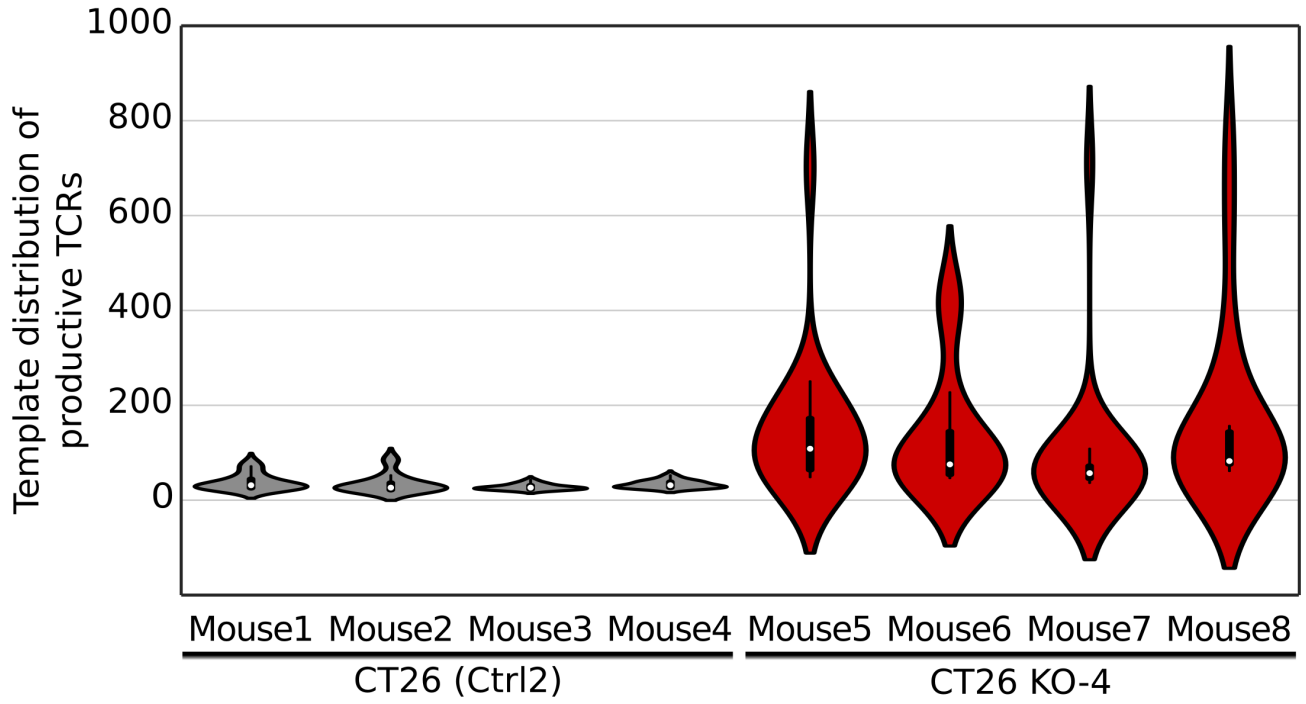

a
mMLH1
Actin
CT26 (Ctrl1)
CT26 KO-1
CT26 KO-2
n.s.
Tumor Volume (mm3)
days
b
CT26 (Ctrl1)
CT26 KO-1
CT26 KO-2
p<0.001
Constitutively Cas9-expressing cells
Transiently Cas9-expressing cells
c
65 Days
mMLH1
Actin
n.s.
CT26 (Ctrl 2)
CT26 KO-3
CT26 KO-4
Not Sub-Cloned
Sub-Cloned
133 Days
157 Days
p<0.001
CT26 (Ctrl 3)
CT26 KO-5
CT26 KO-6
d
p=0.017
p=0.026
CT26 (Ctrl2)
CT26 KO-3
CT26 KO-4
e
mMLH1
Actin
n.s.
n.s
PDAC (Ctrl1)
PDAC KO-1
PDAC KO-2
f
p<0.001
g
p=0.006
p=0.003
PDAC (Ctrl2)
PDAC KO-3
PDAC KO-4

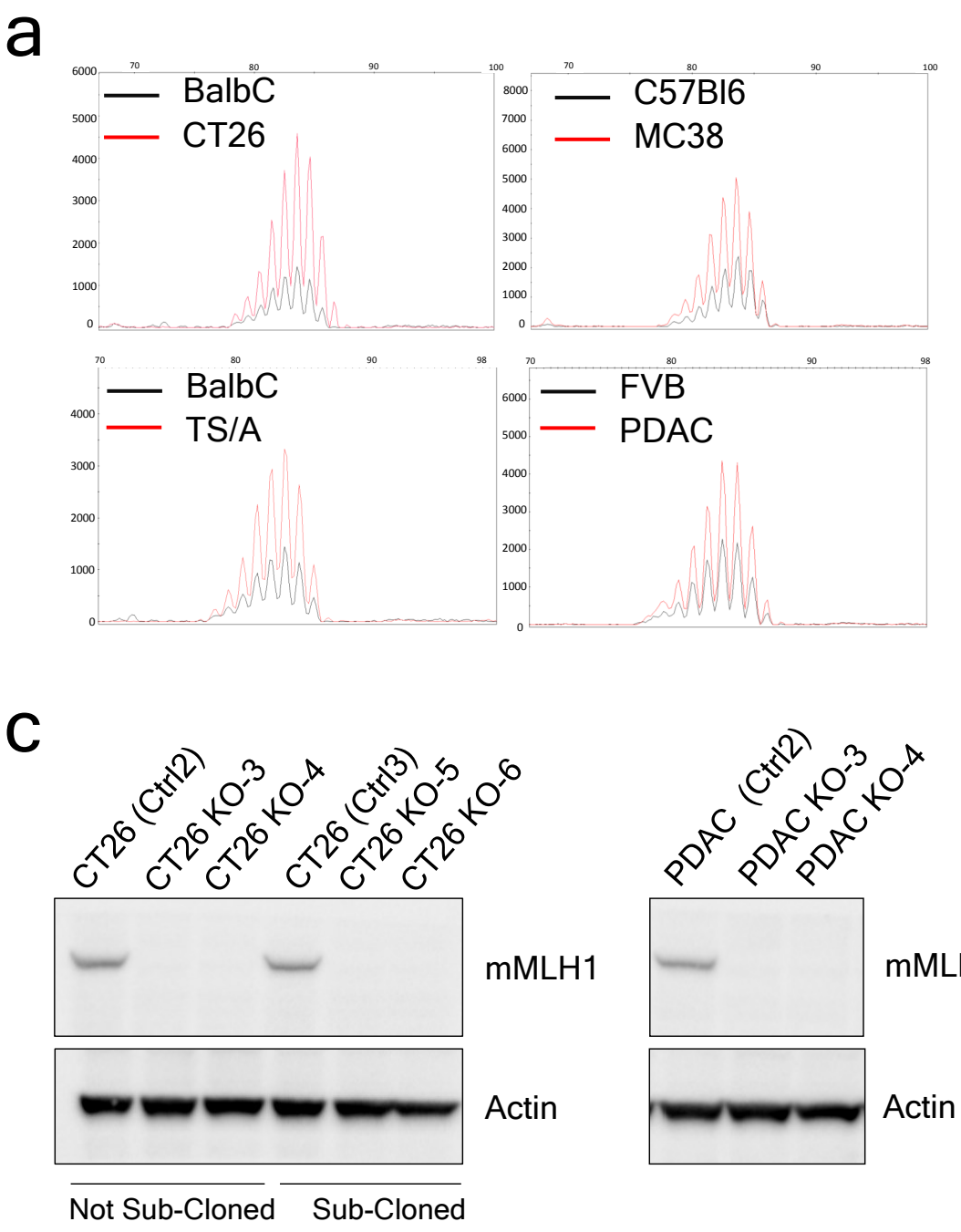

a
CT26 (Ctrl1) Isotype
CT26 (Ctrl1) αPD-1+αCTLA-4
CT26 KO-1 Isotype
CT26 KO-1 αPD-1+αCTLA-4
p<0.001
CT26 KO-2 Isotype
CT26 KO-2 αPD-1+αCTLA-4
p<0.001
Tumor Volume (mm3)
days
Constitutively Cas9-expressing cells
Transiently Cas9-expressing cells
b
CT26 (Ctrl2) Isotype
CT26 (Ctrl2) αPD-1+αCTLA-4
CT26 KO-3 Isotype
CT26 KO-3 αPD1+αCTLA-4
p=0.040
CT26 KO-4 Isotype
CT26 KO-4 αPD1+αCTLA-4
p<0.001
c
CT26 (Ctrl1) Isotype
CT26 (Ctrl1) α CD8
p=0.006
CT26 KO-1 Isotype
CT26 KO-1 αCD8
p<0.001
CT26 KO-2 Isotype
CT26 KO-2 αCD8
p<0.001

a

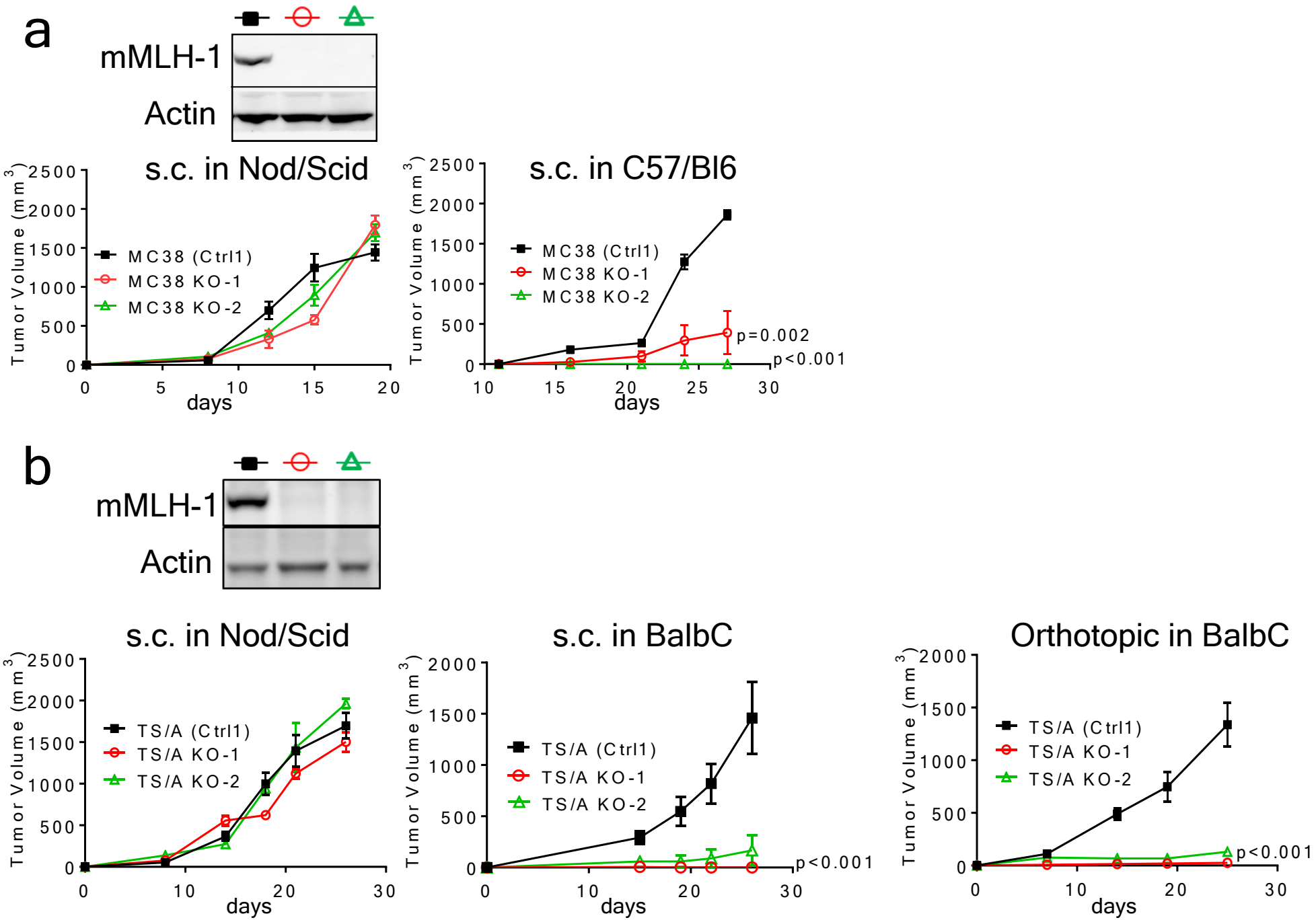

Ctrl2
59
115
316
29
KO-3
69
86
377
469
KO-4
69
91
476
298
Private neoantigens
30 Days
130 Days
246 Days
Common neoantigens
Shared neoantigens

b

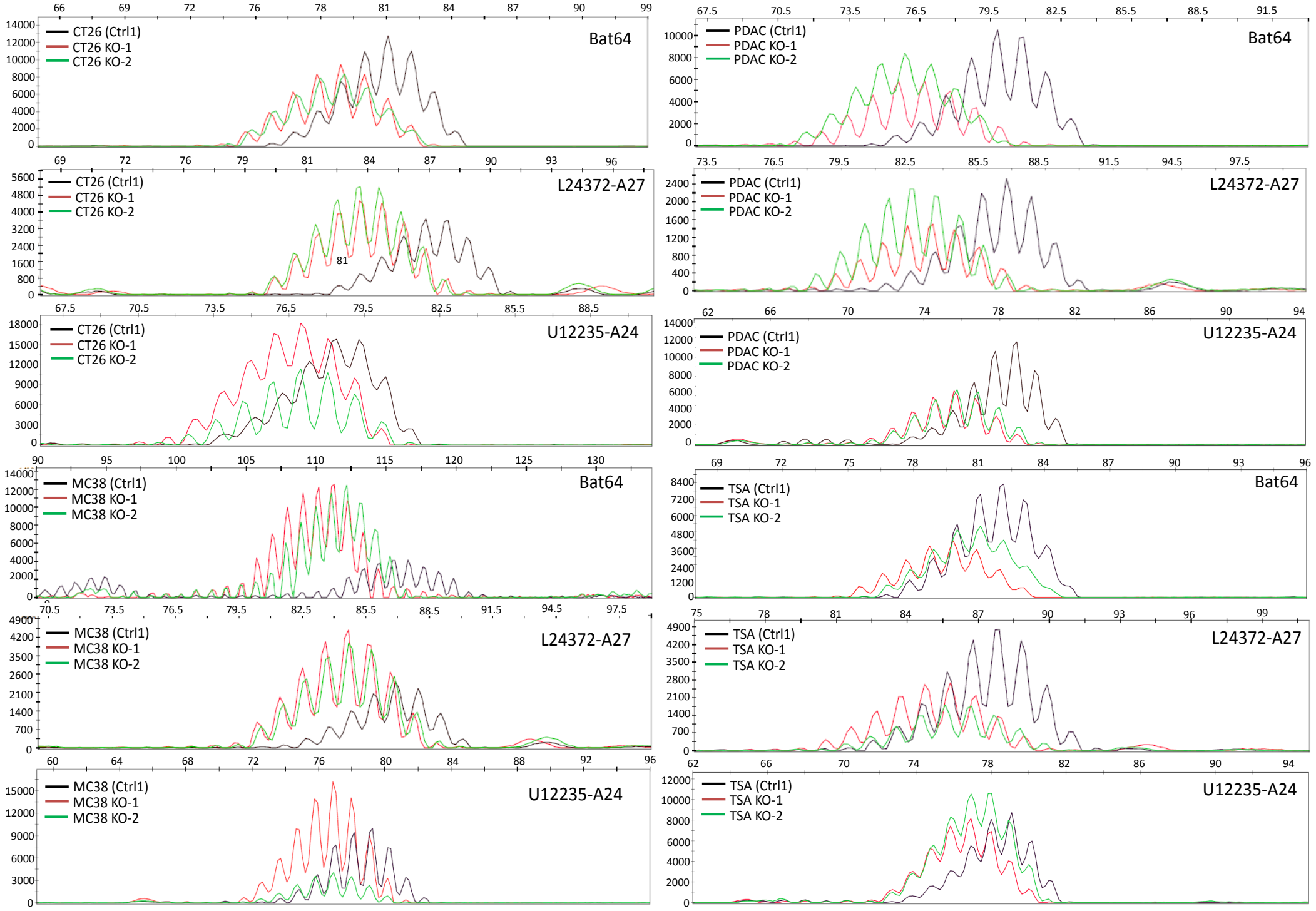

Template distribution of productive TCRs
1000
800
600
400
200
0
Mouse1 Mouse2 Mouse3 Mouse4 Mouse5 Mouse6 Mouse7 Mouse8
CT26 (Ctrl2)
CT26 KO-4



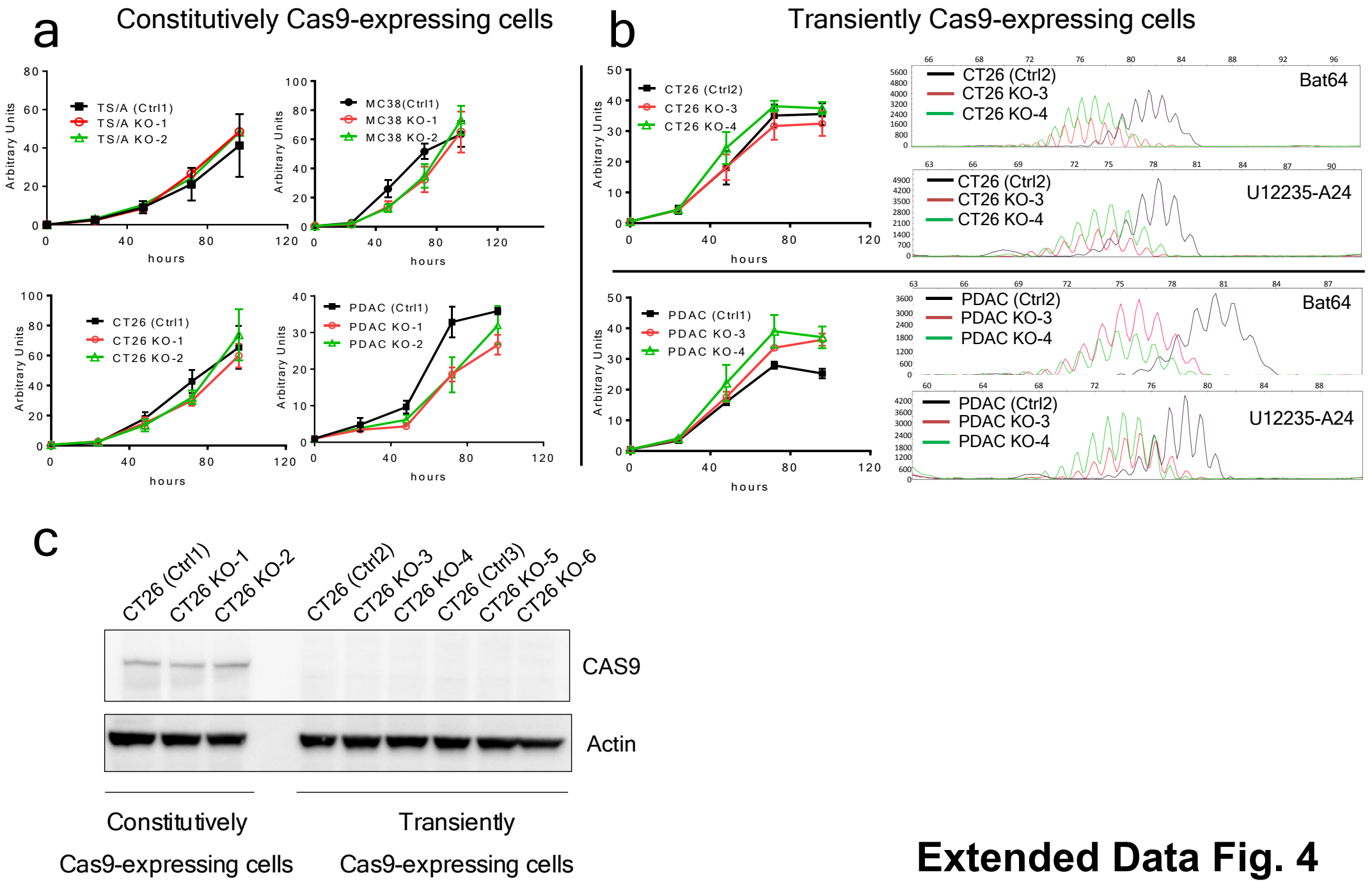
Extended Data Fig. 1



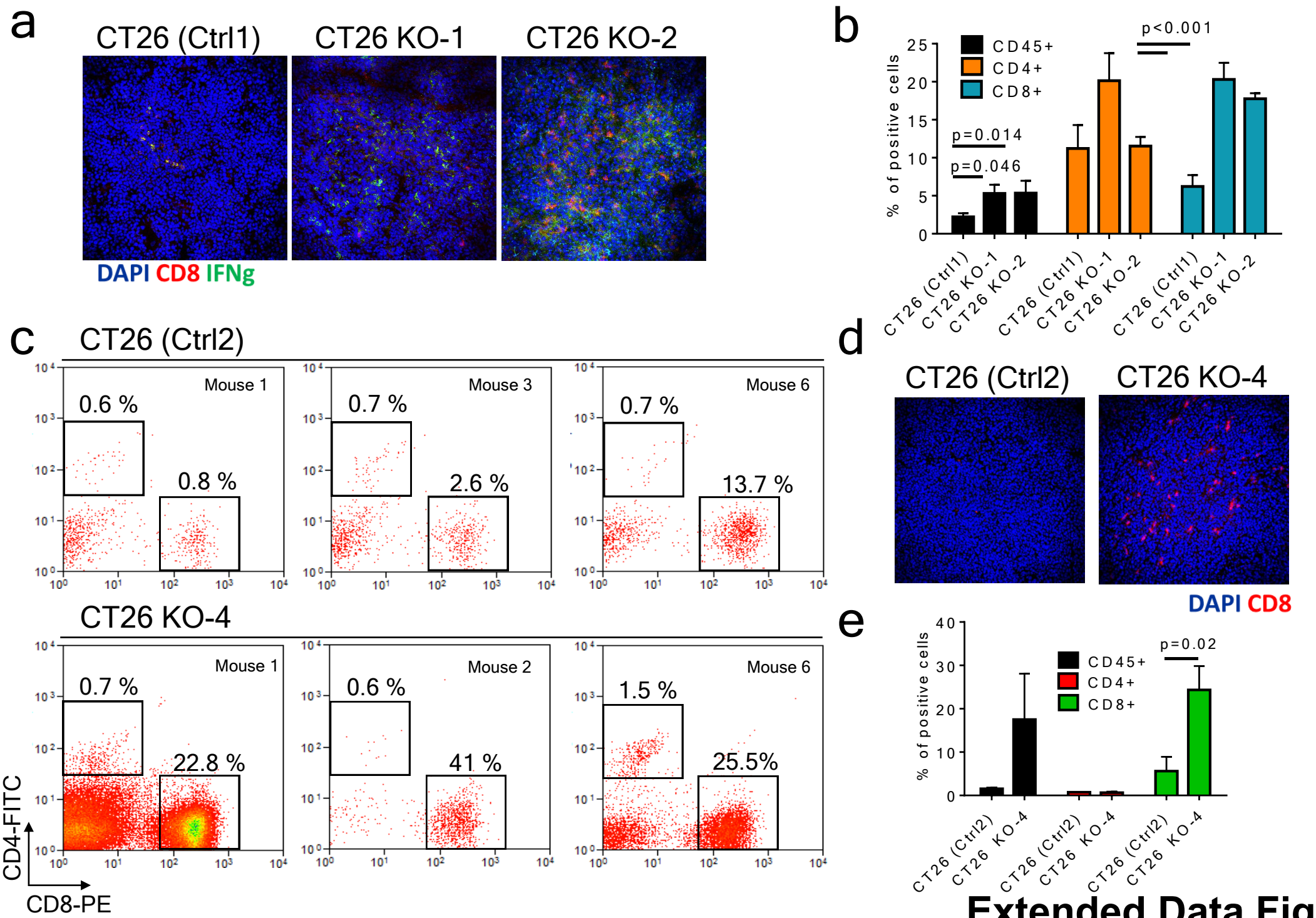
Extended Data Fig. 2



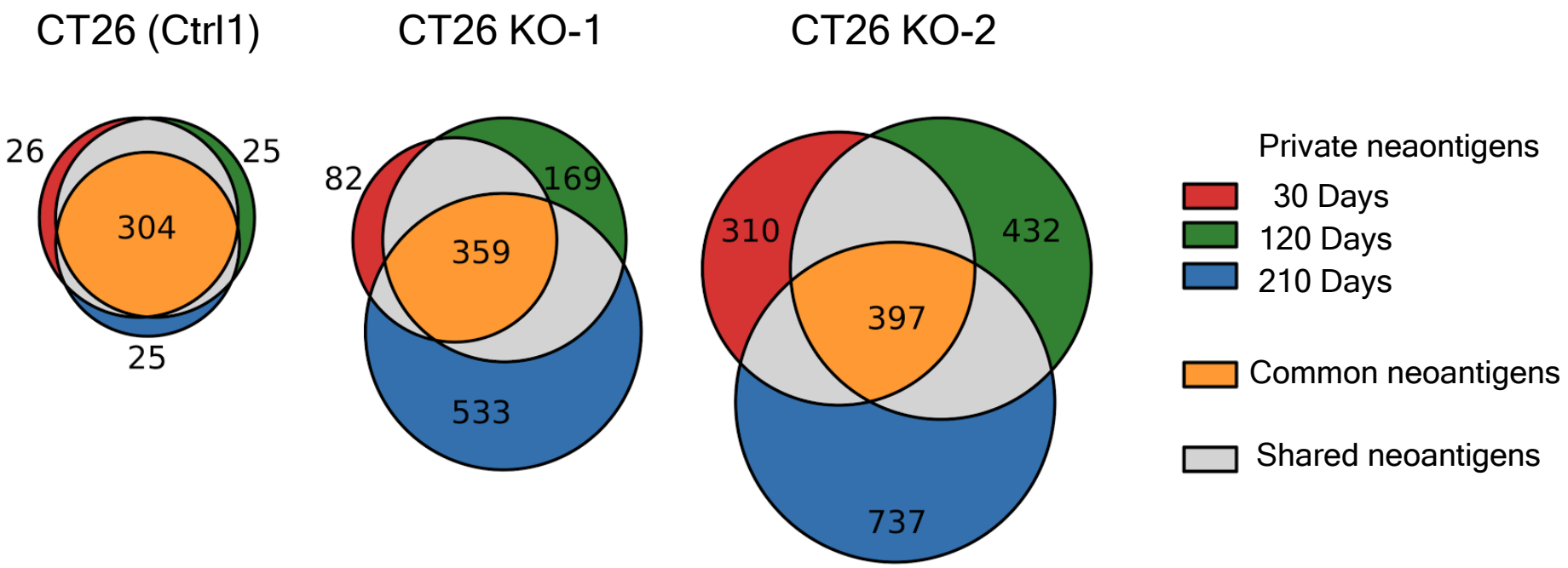
Extended Data Fig. 3



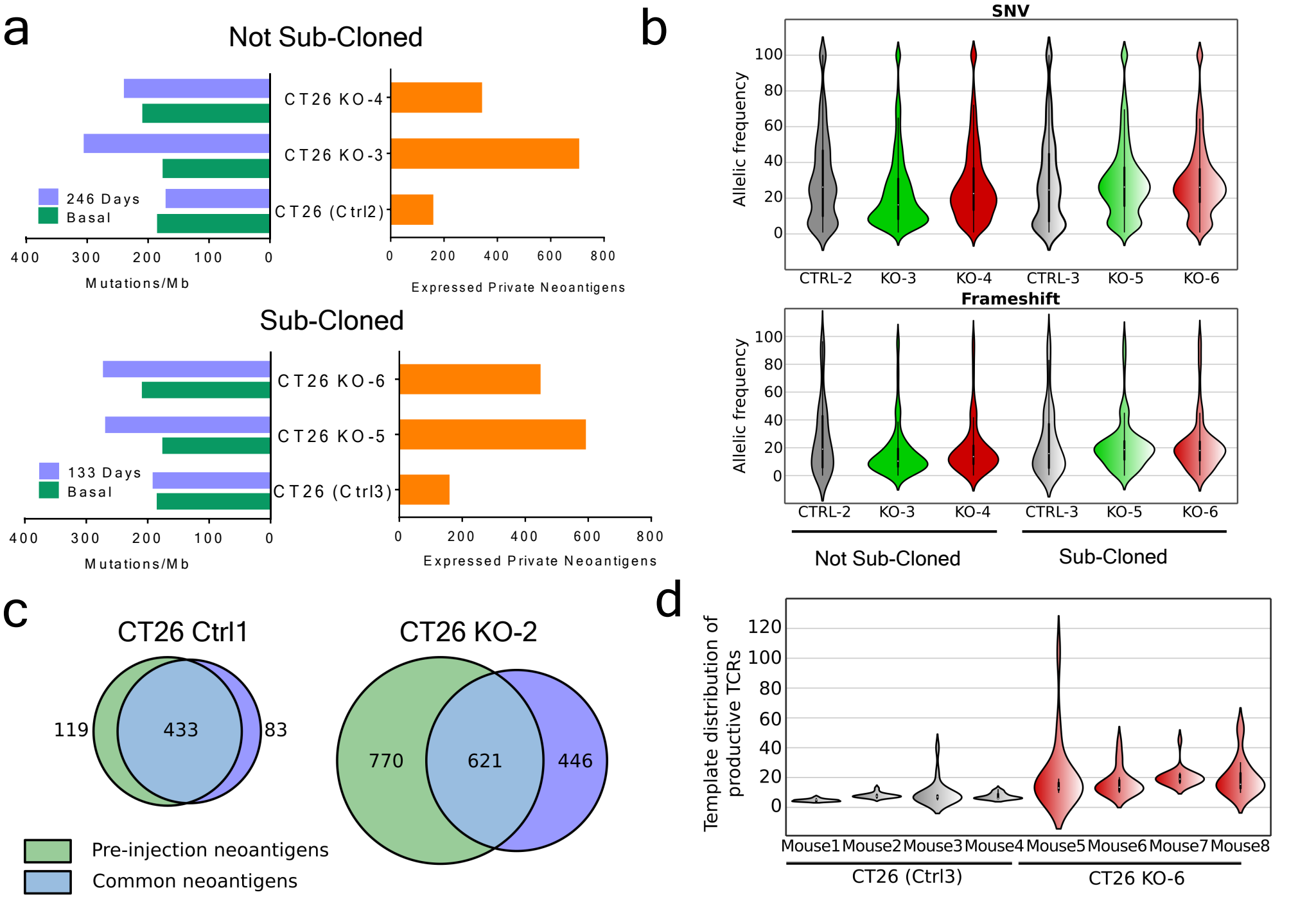
Extended Data Fig. 4



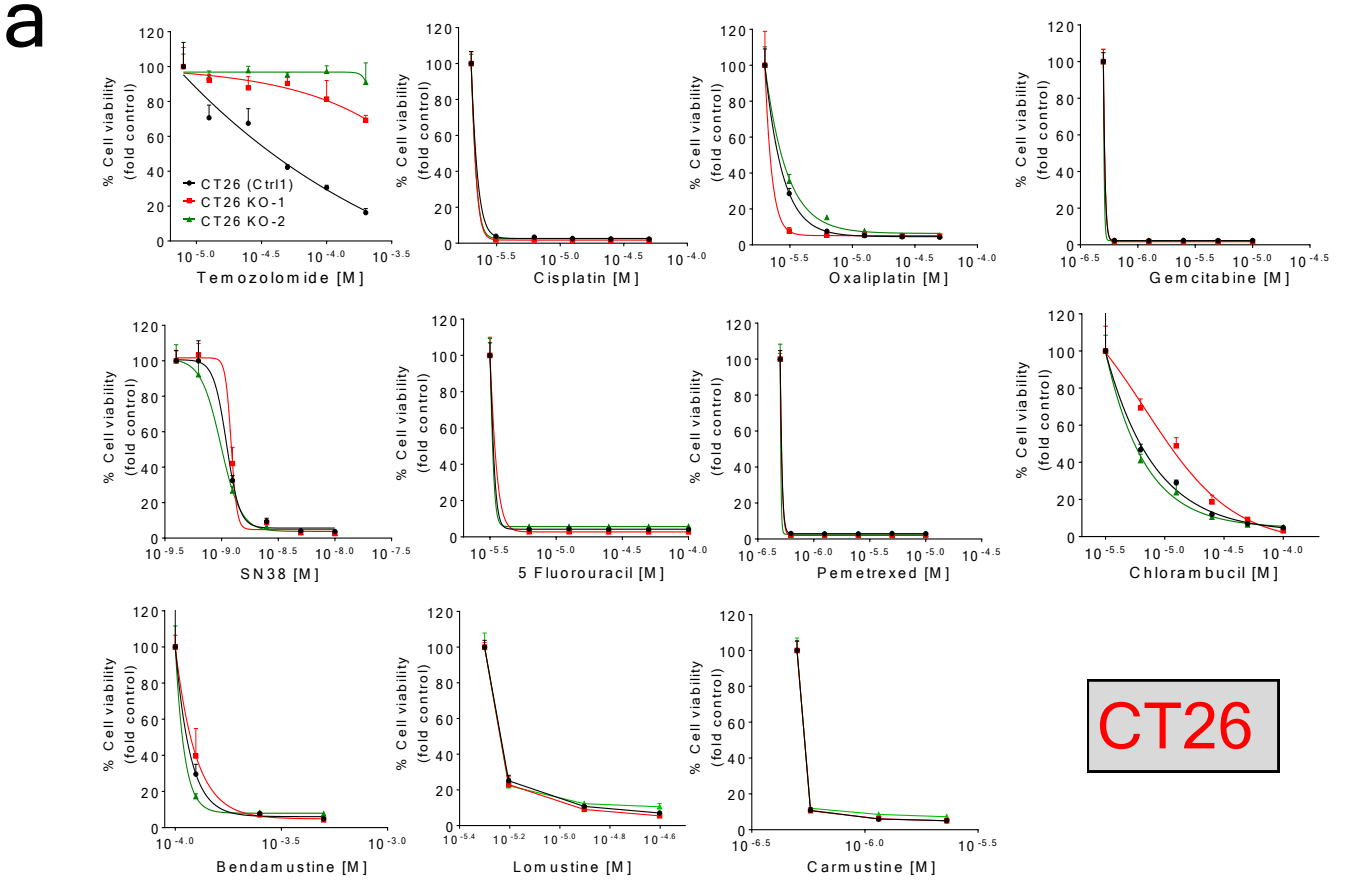
Extended Data Fig.5

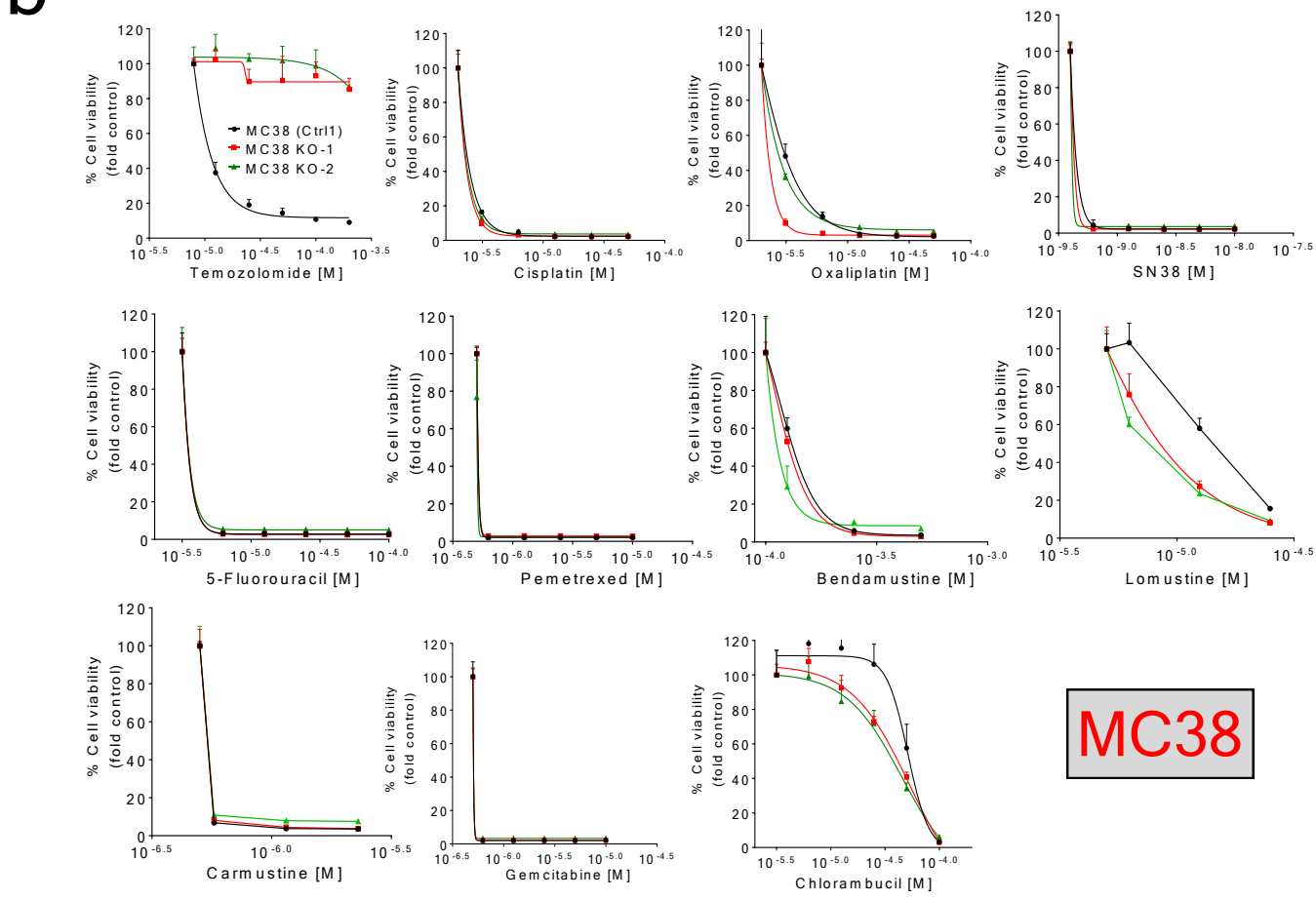


Extended Data Fig. 6

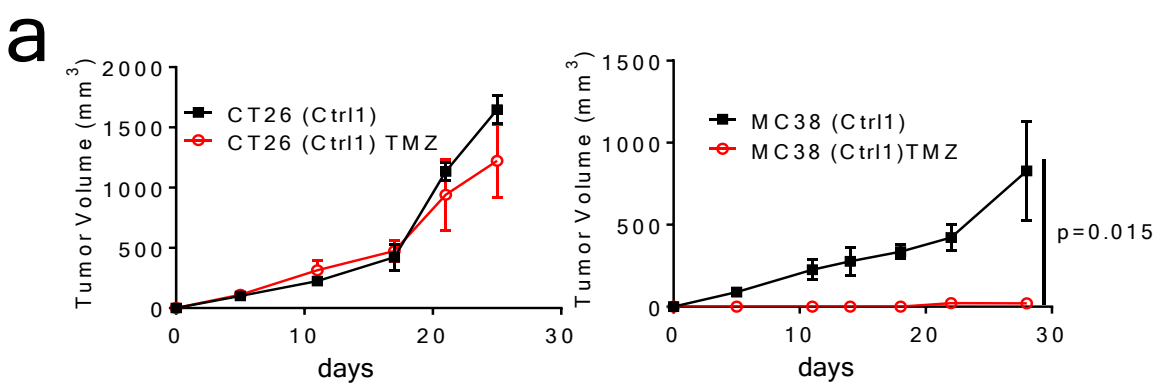


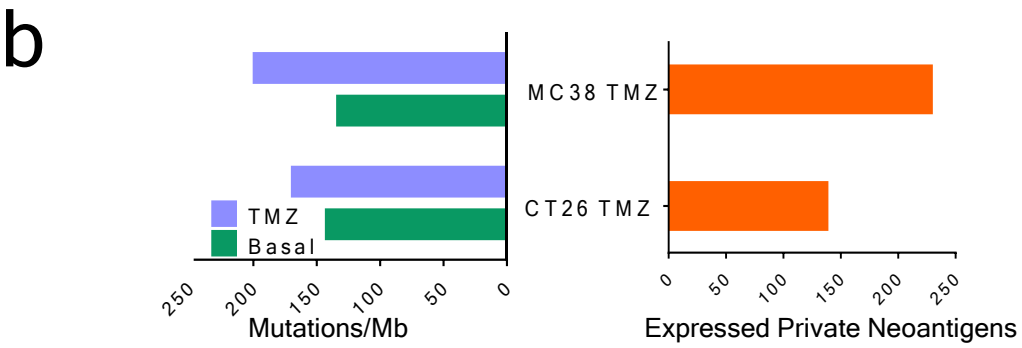
Extended Data Fig. 7

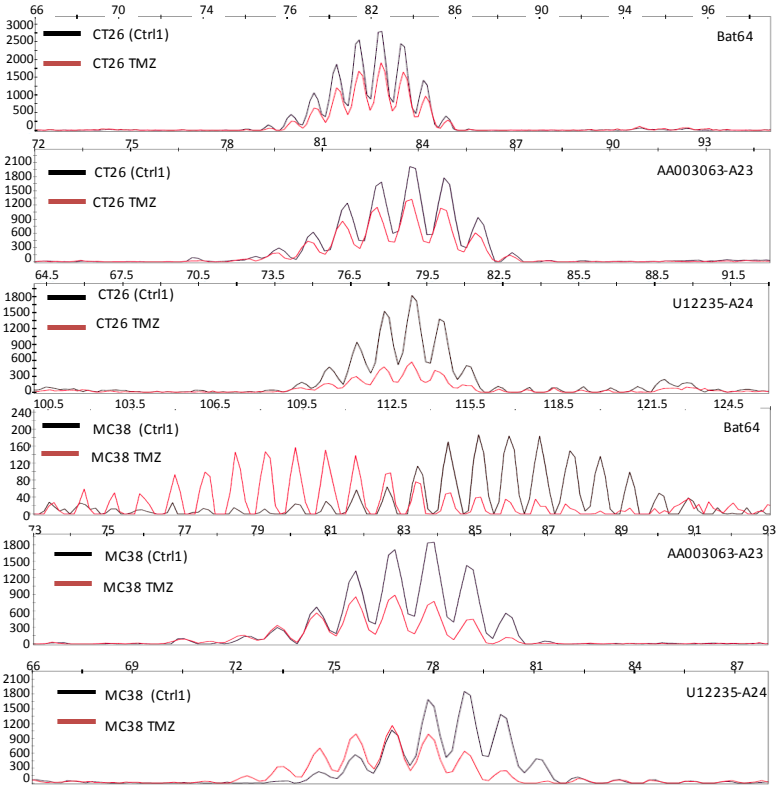


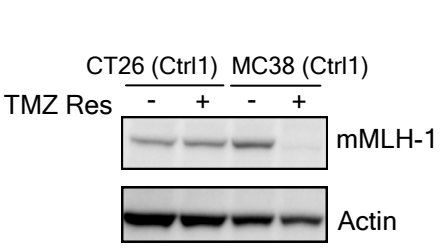


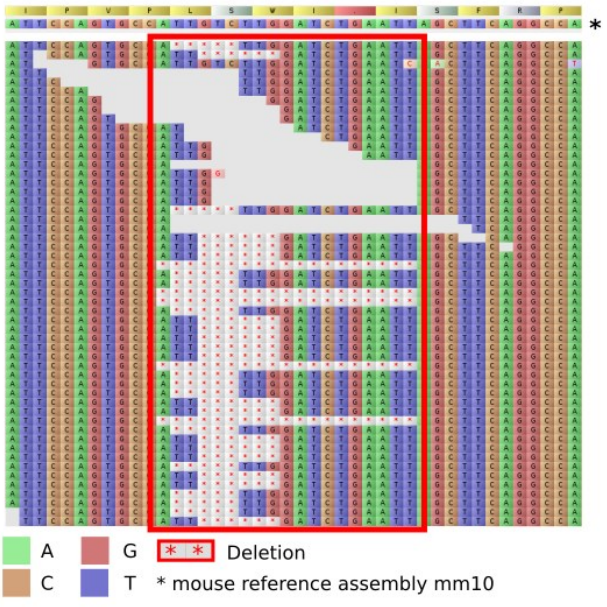
Extended Data Fig. 8



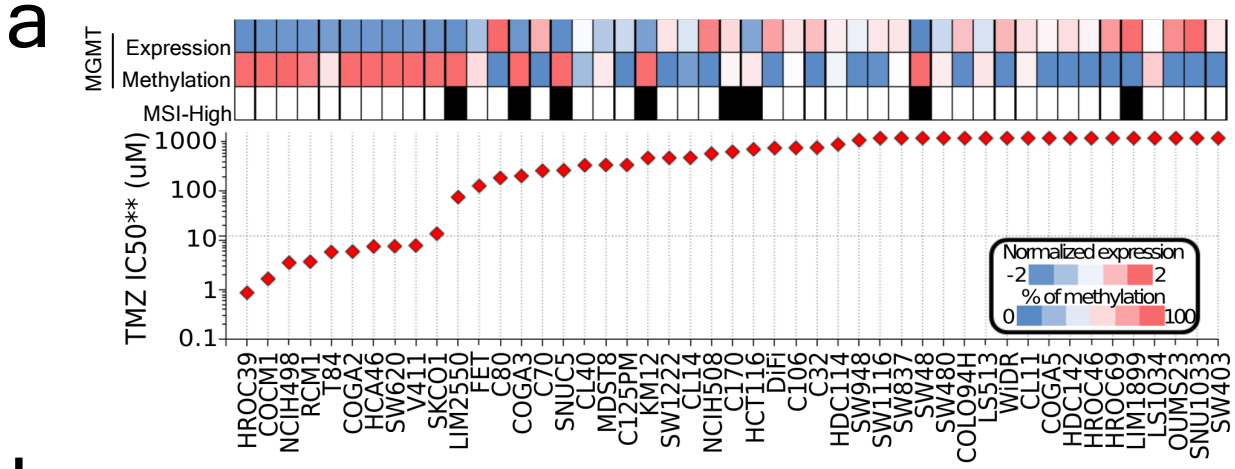


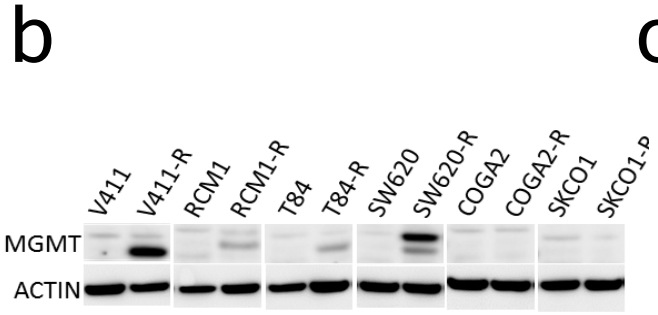


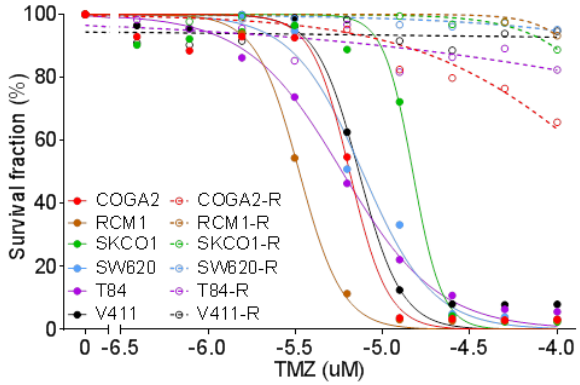


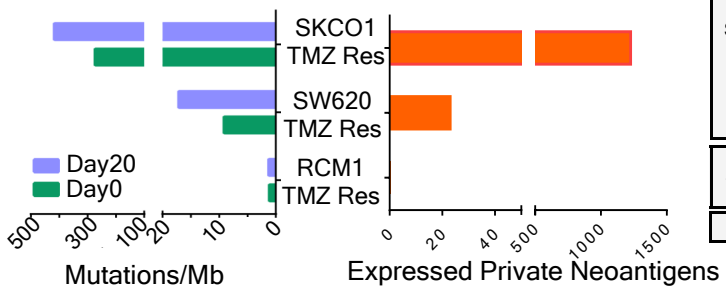


Extended Data Fig 9

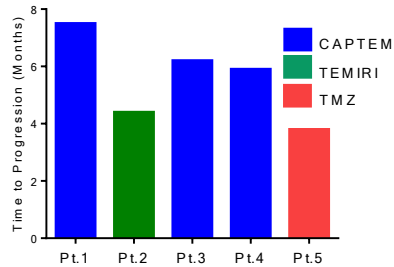




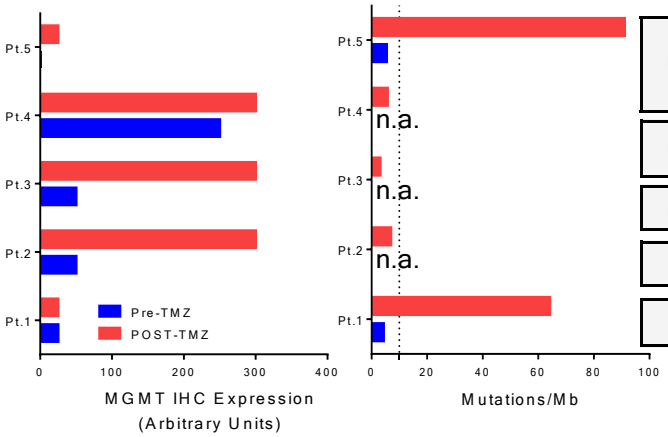




	Gene	aa	Allele Freq.
SKCO1	MSH2	p.T31I	8.1%
	MSH3	p.A654T	10.1%
	MSH6	p.W777*	10.0%
	PMS2	p.T569I	29.9%
	PMS2	p.H532Y	32.4%
	PMS2	p.G108R	45.3%
	POLE2	p.T492I	4.3%
SW620	MSH6	P.R554C	24.8%
	MSH6	**P.T1219I**	**39.3%**
RCM1	-	-	-



	Pt.1	Pt.2	Pt.3	Pt.4	Pt.5
Gender					
Recist	PR	SD	SD	PR	PR
Primary Tumor Location	Rectum	Right Colon	Rectum	Left Colon	Left Colon



	Gene	aa	Allele Freq.
Pt.5	**MSH6**	**p.T1219I**	**14.2%**
	MSH2	p.T782I	46.2%
	MSH6	p.G1224E	17.3%
Pt.4	-	-	-
Pt.3	-	-	-
Pt.2	-	-	-
Pt.1	**MSH6**	**p.T1219I**	**9.4%**

Extended Data Fig.10