SUPPLEMENTARY INFORMATION

FIGURES

Supplementary figure 1. Stage-related C3aR expression in CRC patients.

C3aR expression in stage I to stage IV CRC patients from the TCGA dataset. Significance was calculated using 1-way Anova with Bonferroni post-test (**p<0.01).

Supplementary figure 2. Loss of C3aR exacerbate inflammation and tumor development in inflammation-driven CRC.

(A) Weight loss in WT and C3aR-/- mice treated with AOM/DSS. (B) The total number of tumors in WT and C3aR-/- mice colon. (C) Single tumor diameters in WT and C3aR-/- mice were measured with a sliding caliper and assigned to the groups <2mm or >2mm and used to calculate (D) tumor load. Display of total numbers of dendritic cells and macrophages, Tregs (CD3+ CD4+ CD25+ FoxP3+), Th1 (CD3+CD4+ IFN-γ+) cells, Th17 (CD3+CD4+ IL-17A+) cells, Th1/Th17 (CD3+CD4+ IFN-γ+IL-17A+) cells and Tc (CD8+IFNγ+) in mLN and in the tumors (H-J) by flow cytometry. Results are pooled from two independent experiments with a minimum of 11 mice/group in A-D and five mice/group in E-J. Significance was calculated in A, B and D using t-test; 2-way ANOVA with Bonferroni post-test was used in panel C and E-J (* p< 0.05; ** p< 0.01; *** p< 0.001).

Supplementary figure 3. Loss of C3aR expression does not significantly impact the functional and phenotypic composition of mLN-infiltrating lymphocytes.

Single cell suspensions from mLN of WT, $APC^{Min/+}$, $APC^{Min/+}$/C3aR-/- and C3aR-/- mice were analyzed by FACS and total number of (A) CD4+T cells, (B) Th17 cells, (C) Th1 cells, (D) Th1/Th17 cells, (E) CD8+T cells, (F) Tc cells were obtained. 9 animals/group were used, and

significance was calculated using 1-way Anova with Bonferroni post-test (** p< 0.01; ****p>0.0001).

Supplementary figure 4. Transplantation of the APC$^{Min/+}$/C3aR-/- microbiota does not affect the number of tumors in the small intestine (SI) and the functional profile of mLN-infiltrating lymphocytes.

5-week-old APC$^{Min/+}$ mice (recipients) were treated for 1 week with broad spectrum antibiotics. 48 h after antibiotics treatment, the recipient mice were transplanted via oral gavage with the gut microbiota of 12-week-old APC$^{Min/+}$ or APC$^{Min/+}$/C3aR-/- mice (donors) for 3 consecutive days and again once/week for 7 weeks. (A) Total tumor count in the SI and (B) in the different portions of the SI: duodenum (D), jejunum (J) and ileum (I) of recipient mice. (C-H) Flow cytometry analysis of mLN infiltrating lymphocytes showing in (C) total CD4+T cells, (D) Th17 cells, (E) Th1 cells, (F) Th1/Th17 cells, (G) CD8+ T cells and (H) Tc cells. Shown are results of 2 independent experiments with 11 mice/group. Significance was calculated in A and B using 1-way Anova and 2-way Anova respectively with Bonferroni post-test and in C-H using unpaired T test (** p< 0.01; *** p< 0.001; **** p<0.0001).

Supplementary figure 5. Impact of C3aR on tumor- and mucus-associated microbiota species.

Bacterial DNA was extracted from colon tumors and mucus of 12-week-old APC$^{Min/+}$ and APC$^{Min/+}$/C3aR-/- mice and metagenomic analysis was performed. Doughnut charts showing the bacterial composition of tumors (A) and mucus (B) in 12-week-old APC$^{Min/+}$ and APC$^{Min/+}$/C3aR-/- mice. Significance was calculated in A by DESeq R-package.

Supplementary figure 6. Differential gene expression analysis in distal colon of APC$^{Min/+}$, APC$^{Min/+}$/C3aR-/-, C3aR-/- and WT mice.

Summary plot showing the number of significantly differentially expressed genes in distal colons of (A) C3aR-/- vs. WT mice (n=4); (B) $APC^{Min/+}$ vs. WT mice; (C) $APC^{Min/+}$/C3aR-/- vs. C3aR-/- mice; (D) $APC^{Min/+}$/C3aR-/- vs. WT mice; (E) $APC^{Min/+}$ vs. C3aR-/- mice. The plots display log fold change against log counts per million for each gene with the red points representing significant differentially expressed genes, and the horizontal blue lines indicating a 2-fold increase or decrease in expression. Significance was calculated using the edgeR function decideTestsDGE with default p-value adjustment method (Benjamini-Hochberg false discovery rate, FDR), default p-value threshold (0.05), and a custom log fold change threshold of 1 (1 on the log2 scale, equivalent to a relative change of 2 or 0.5 in non-logarithmic terms).

Supplementary figure 7. Loss of C3aR results in transcriptional up-regulation of innate and adaptive immune pathways in the healthy distal colon of $APC^{Min/+}$/C3aR-/- mice.

(A) Heatmap from (row-wise z-transformed) log counts per million values, using data from distal colon of indicated mouse strains and all genes that are significant at 5% FDR in at least one contrast. Rows represent genes, and columns represent individual samples. Each row (gene) is z-transformed to have mean zero and standard deviation one. The heatmap also includes column annotation labels indicating the mouse strain for each sample. (B) Visualization of the gene-associated Gene Ontology (GO) biological processes in healthy colon. (C) Volcano plot illustrating the magnitude of fold change for all genes differentially expressed in the distal colon of $APC^{Min/+}$ vs. $APC^{Min/+}$/C3aR-/- mice (n=4). Representative significantly up-regulated and down-regulated genes belonging to the GO biological processes shown in B are highlighted in yellow. Significance was calculated using the edgeR function decideTestsDGE with Benjamini-Hochberg correction for false discovery rate (FDR); a default FDR threshold of 0.05 and a log2 fold change (log2FC) threshold of 0.6 were applied.

Supplementary figure 8. Validation of the mass cytometry panel

The antibody panel was validated by solution mass cytometry on fresh cells isolated from WT spleen and colon lamina propria digested with collagenase VIII. For cytokines, the cells were stimulated with PMA/Ionomycin, stained with surface antibodies and then permeabilized to perform the intracellular staining following the procedures described in the material and methods of our manuscript. Metal-tagged antibodies against PE and APC were used to detect IL-17A-PE and IFN-γ-APC, respectively.

Supplementary Figure 9. Small intestinal tumors in $APC^{Min/+}$ and $APC^{Min/+}$/C3aR-/- mice

Tumor burden in the small intestine of $APC^{Min/+}$ and $APC^{Min/+}$/C3aR-/- mice from 5 to 24 weeks. Significance was calculated in a by 1-way Anova with Bonferroni post-test (ns= not significant; **p<0.01; **** p<0.0001).

MATERIALS AND METHODS

Animals

C3aR-/- mice on C57BL/6J background were a kind gift from Dr. Bao Lu and Dr. Carl Atkinson. $APC^{Min/+}$/C3aR-/- mice were generated by crossing $APC^{Min/+}$with C3aR-/- mice in our animal facilities. Male and female mice were used throughout the study unless otherwise specified. All experiments were performed during the light cycle. The animals were housed in individually ventilated cages with corn cob bedding. Non littermate mice were co-housed for a minimum of two weeks before performing any intervention.

CRC models

For inflammation-driven CRC, mice were treated 10 mg/Kg of azoxymethane (AOM; Sigma-Aldrich). After 7 days, mice were administered with 1.5% (w/v) DSS (TdB Consultancy) in their drinking water for 1 week followed by 2 weeks of recovery with water. DSS/recovery cycles were

repeated three times as previously described[1]. For the spontaneous tumor model $APC^{Min/+}$ and $APC^{Min/+}$/C3aR-/- mice were euthanized at indicated time points, tumor counted and used for downstream applications.

Flow cytometry and preparation of single cell suspensions

Single cell suspensions were prepared from mLN following standard protocols. Colon lamina propria cells were isolated as previously described with minor modifications[2]. For isolation of tumor-associated immune cells, colon polyps were shaken in PBS, 1% BSA, 10 mM EDTA to remove epithelial cells, and further digested for 30 minutes at 37°C with Collagenase VIII (Sigma-Aldrich) in complete medium with shaking.

For flow cytometry staining, cells were incubated with anti-FcR antibody (clone 24G2) and stained with the following surface antibodies: anti-CD45.2 (clone 104, eBioscience), Ly6G (clone 1A8), CD3 (clone 17A2, eBioscience), Ly6C (clone AL-21), CD11b (clone M1/70), CD4 (cloneRM4-5), CD8a (53-6-7), B220 (clone RA3-6B2), I-A/I-E (clone M5/114.15.2), CD11c (clone HL3), F4/80 (MB8, eBioscience), CD103 (clone 2E7), CD25 (clone PC61), anti- IFN-γ (clone XMG1.2), anti-IL-17A (clone TC11-18H10), anti-FoxP3 (clone FJK-16s eBioscience). For cytokine staining, cells were permeabilized with Cytofix/Cytoperm buffer according to manufacturer instructions. For Foxp3 staining the permeabilization was performed using the FoxP3 permeabilization buffer (eBioscience). All antibodies were purchased from BD Pharmingen unless otherwise specified. Samples were acquired with FACSCanto II or Fortessa LSR (BD Bioscience) and analyzed with FlowJo software (TreeStar).

Fecal microbiota transplantation

For fecal microbiota transplantation (FMT), fecal material was obtained from 12-week-old $APC^{Min/+}$ and $APC^{Min/+}$/C3aR-/- donor mice. Briefly, fecal pellet and cecal contents were harvested

from the mice, dissolved in sterile PBS at 100 mg/ml, passed through a 70-μm strainer in order to eliminate insoluble debris, and frozen in Columbia broth with 20% glycerol. At the time of administration, aliquots were thawed, centrifuged to eliminate the glycerol-containing medium and resuspended at 100 mg/ml for gavage. This procedure allowed us to use the same material throughout the experiment, thereby avoiding confounding effects due to fecal material coming from different animals. Five-week-old $APC^{Min/+}$ mice were used as FMT recipient. Recipient mice were treated for one week with the broad-spectrum antibiotics in the drinking water, rested for 48 h with water without antibiotics, followed by gavage with 200 μl /mouse of the previously prepared fecal material for 3 consecutive days in the first week and once a week for the next 6 weeks.

Microbiota profiling

V5-V6 hypervariable regions of bacterial 16S rRNA gene were amplified and processed with a modified version of the Nextera protocol[3]. The obtained metabarcoding libraries were sequenced by using the MiSeq Illumina platform with a 2x250 paired end (PE) approach. Metagenomic amplicons were analyzed by applying the BioMaS[4] pipeline as follows: (i) The paired-end reads were merged into consensus sequences using Flash[5] and subsequently dereplicated as previously described[6] maintaining the consensus sequence; (ii) The remaining non overlapping PE reads were considered for further analysis only if after the low-quality region trimming (Phred quality cut-off = 25) both read ends were ≥50 bp long[7]; Both the merged sequences and the unmerged reads were matched against the RDP (Ribosomal Database Project) database (release 10.29)[8] by Bowtie2[9]. The mapping data were filtered according to two parameters: identity percentage (≥97%) and query coverage (≥70%); (iv) Finally, all mapped reads fulfilling the settled filters were taxonomically annotated using the Tango tool[10]. Assigned genera were filtered considering as present only the ones for which at least five reads per samples were present. The read counts were

normalized using an approach similar to the RPKM (Reads per kilo-base per million): normalized count = assigned reads / (total assigned reads at the rank level/1,000,000). Significant differences at the genus and species level were calculated with the DESeq R-package[11]. A tree representing the phylogenetic relationship between the Amplicon Sequence Variants (ASV) was produced by using the QIIME1 package[12]. The rarefaction curves were inferred and plotted by using an in house developed R script relying on the rarefy function of the vegan package. Alpha and Beta diversity analysis were performed using the phyloseq R[13]. In particular, the observed ASVs, the Shannon and the Faith Indexes were used as richness and alpha diversity measures. The weighted and unweighted UniFrac and the Bray-Curtis metrics were used for the Beta diversity inference[14]. Difference in the inferred alpha diversity indices were measured by using the Kruskal-Wallis test followed by a pairwise Wilcoxon as post-hoc test (p-values corrected by using the Benjamin-Hochberg procedure). The PERMANOVA test was used to compare groups in beta diversity data, using 999 permutations. To isolate bacterial DNA from mucus scraped from the colon we used the same protocol used to isolate bacterial DNA from the feces. For the isolation of tumor-associated bacteria, tumors were homogenized in sterile phosphate buffered saline solution, centrifuged, and both supernatants and pellets were subjected to depletion of host DNA by using the QIAamp DNA Microbiome kit (Qiagen) before proceeding with bacterial DNA extraction. Sequencing libraries were prepared using the Nextera XT Library Preparation kit (Illumina) following the manufacturer's guidelines and ran on HiSeq 2500 platform (Illumina).

RNA-Seq

For RNASeq analysis, 12-week-old $APC^{Min/+}$, WT, C3aR-/- and $APC^{Min/+}$/C3aR-/- mice (4 mice/group) were sacrificed. The proximal (first 3 cm adjacent to the caecum) and distal colon (the final 3.5 cm adjacent to the rectum) and the single tumors were immediately preserved in 2

ml of RNAlater stabilization solution (Qiagen) according to manufacturer instructions and stored at -20°C. For RNA extraction, tumors and colons were homogenized in Trizol (500 ul/50 mg weight), centrifuged at 13000 rpm. The supernatant was loaded on Qiagen Mini Kit columns and treated as indicated in the manufacturer instructions. RNA concentration was quantified by Nanodrop and quality and integrity were evaluated by using a bioanalyzer (Agilent Technologies). Only samples with a RIN>8 were used for downstream applications. Briefly, 1µg of RNA was used to prepare cDNA with the TruSeq RNA kit following the recommendations for low sample preparation protocol (Illumina). Samples were sequenced on an Illumina MiSeq instrument at a depth of 35x10^6 reads per sample. Reads in FASTQ format were quantified at the gene level using featureCounts and the count table was delivered to edgeR for the differential expression analysis using the GLM functionality[15, 16]. GO analysis was conducted to identify DEG at the biologically functional level. The identified DEGs were uploaded to the online software Advaita Bioinformatics to integrate functional genomic annotations. A false discovery rate (FDR) cutoff of 5% and minimum fold change of 1.5 (log2FC=0.6) was applied to determine differential expression.

Analysis of C3aR expression, methylation and correlation with immune cells in patients with CRC

Data on C3aR expression were obtained from the R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl) or TCGA database. C3aR expression data were plotted with PRISM (GraphPad Software). Data on C3aR methylation were obtained from the human pan-cancer methylation database (MethHC, http://methhc.mbc.nctu.edu.tw/php/index.php)[17].

Spearman and Pearson correlations between C3aR (queried as c3ar1) mRNA expression and diverse immune cell populations (assessed by CIBERSORT) were made available for this study

and downloaded through the private S:CORT cBioPortal for 179 colon cancer patients (FOXTROT cohort) and 231 rectal cancer patients. All cases in the S:CORT WS3 Grampian Set (retrospective cohort) were included in the study.

Immune cell populations queried: CIBERSORT: Resting dendritic cells, naïve B-cells, resting and activated NK cells, regulatory T cells, CD8 T-cells, plasma cells, neutrophils, monocytes, M1 and M2 macrophages, CD4 T cells.

Mass cytometry

Mass cytometry antibodies were either labelled in-house using antibody-labelling kits and protocols according to manufacturer's instructions or purchased from Fluidigm. Antibodies were individually titrated and optimized into the final panel before use, ensuring that each parameter was informative.

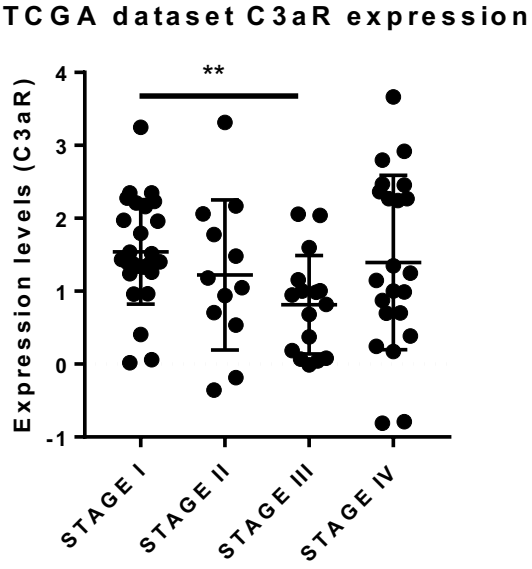
After barcoding, the composite sample was stained with the cocktail of primary antibodies (Table 1), washed and fixed with 1.6% paraformaldehyde (PFA; Electron Microscopy Sciences) at 4 °C. Fixed cells were permeabilized and resuspended in 400 µl of intracellular antibody mixture (Suppl. Table 1) in permeabilization buffer at 4°C. After washing, samples were incubated with DNA-intercalating solution (Iridium (Sigma) in MaxPar Fix/Perm buffer (Fluidigm) overnight at 4°C. Before the acquisition, the samples were washed twice with MilliQ water. Barcoded composite samples were acquired on a Helios mass cytometer (Fluidigm). Quality control and tuning processes on the Helios are performed daily before acquisition. Data from different days and across acquisition time was normalized by adding five-element beads to the sample immediately before acquisition and using the MATLAB-based normalization software, as described previously[18].

REFERENCES

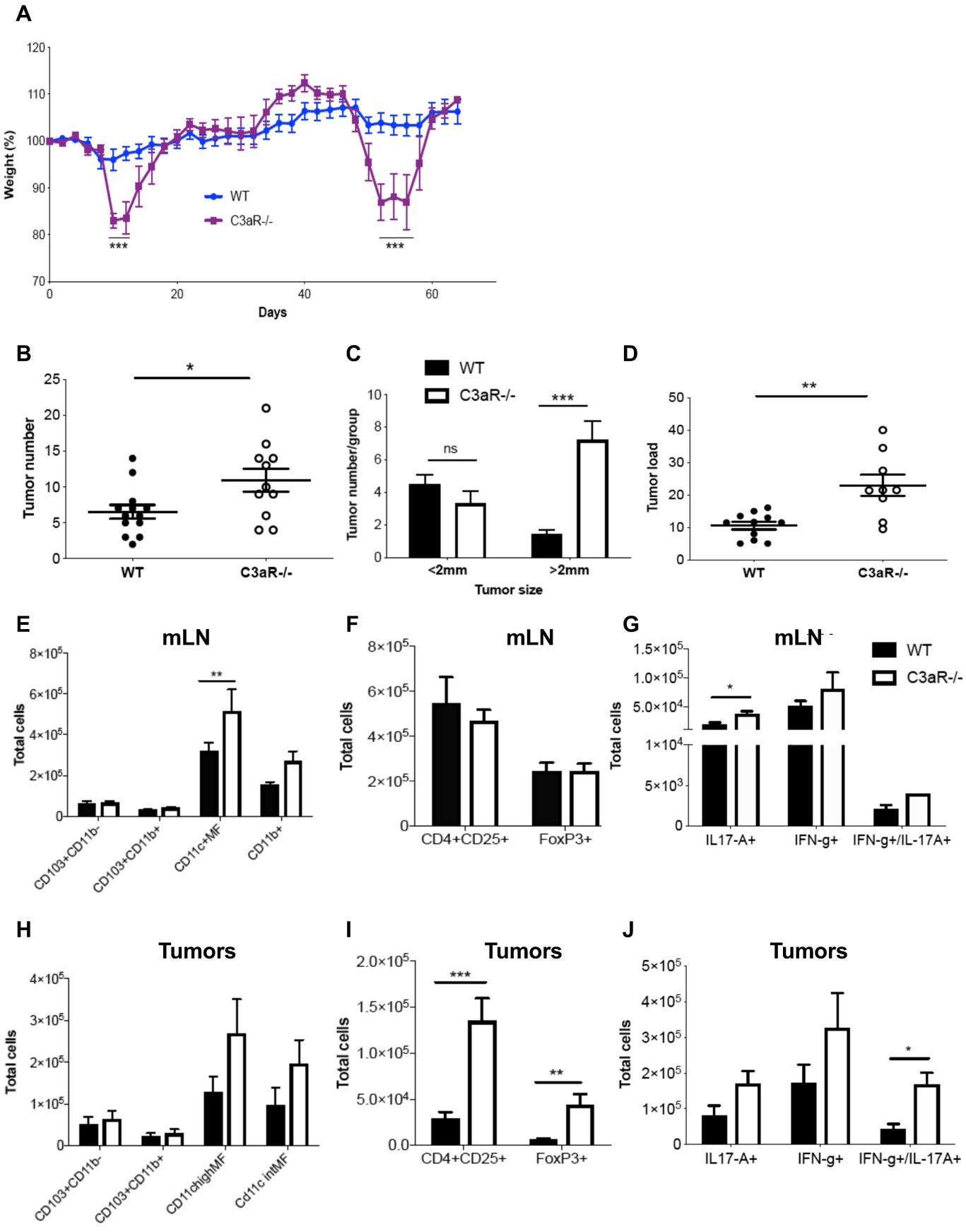
1. Neufert C, Becker C, Neurath MF. An inducible mouse model of colon carcinogenesis for the analysis of sporadic and inflammation-driven tumor progression. Nat Protoc 2007;2:1998-2004.
2. Lefrancois L, Lycke N. Isolation of mouse small intestinal intraepithelial lymphocytes, Peyer's patch, and lamina propria cells. Curr Protoc Immunol 2001;Chapter 3:Unit 3 19.
3. Manzari C, Chiara M, Costanza A, et al. Draft genome sequence of Sphingobium sp. strain ba1, resistant to kanamycin and nickel ions. FEMS Microbiol Lett 2014;361:8-9.
4. Fosso B, Santamaria M, Marzano M, et al. BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. BMC Bioinformatics 2015;16:203.
5. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 2011;27:2957-63.
6. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460-1.
7. Mariathasan S, Turley SJ, Nickles D, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 2018;554:544-548.
8. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;37:D141-5.
9. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.
10. Alonso-Alemany D, Barre A, Beretta S, et al. Further steps in TANGO: improved taxonomic assignment in metagenomics. Bioinformatics 2014;30:17-23.

11. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
12. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335-6.
13. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8:e61217.
14. Chang Q, Luan Y, Sun F. Variance adjusted weighted UniFrac: a powerful beta diversity measure for comparing communities based on phylogeny. BMC Bioinformatics 2011;12:118.
15. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923-30.
16. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;40:4288-97.
17. Huang WY, Hsu SD, Huang HY, et al. MethHC: a database of DNA methylation and gene expression in human cancer. Nucleic Acids Res 2015;43:D856-61.
18. Finck R, Simonds EF, Jager A, et al. Normalization of mass cytometry data with bead standards. Cytometry A 2013;83:483-94.

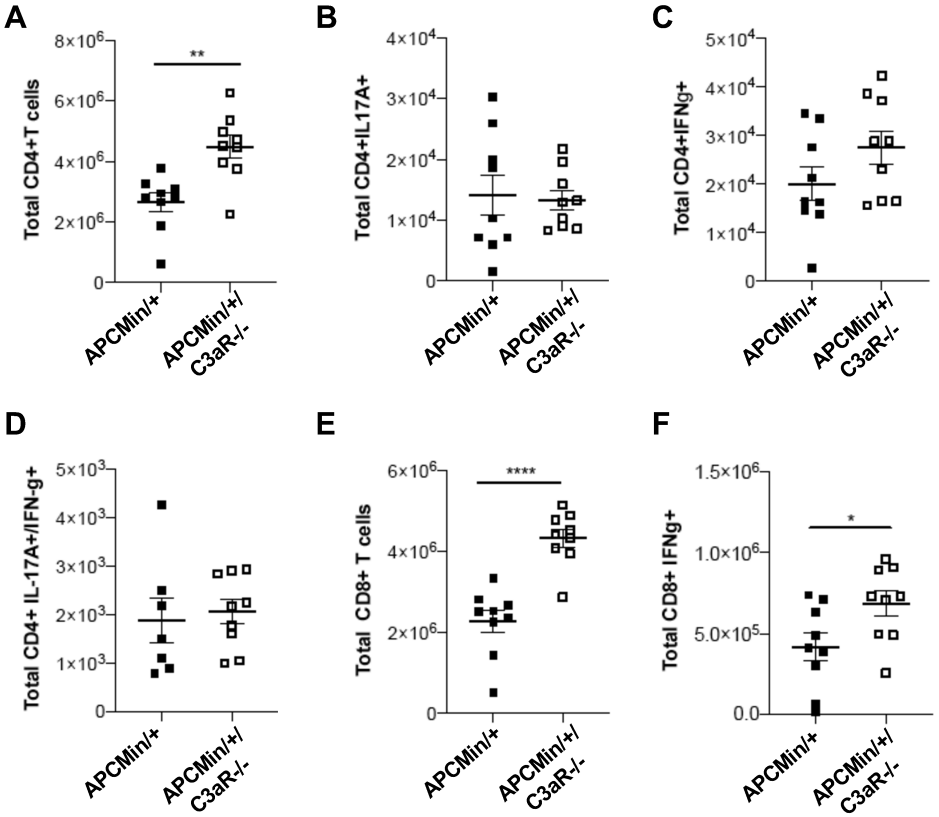
Supplementary Figure 1



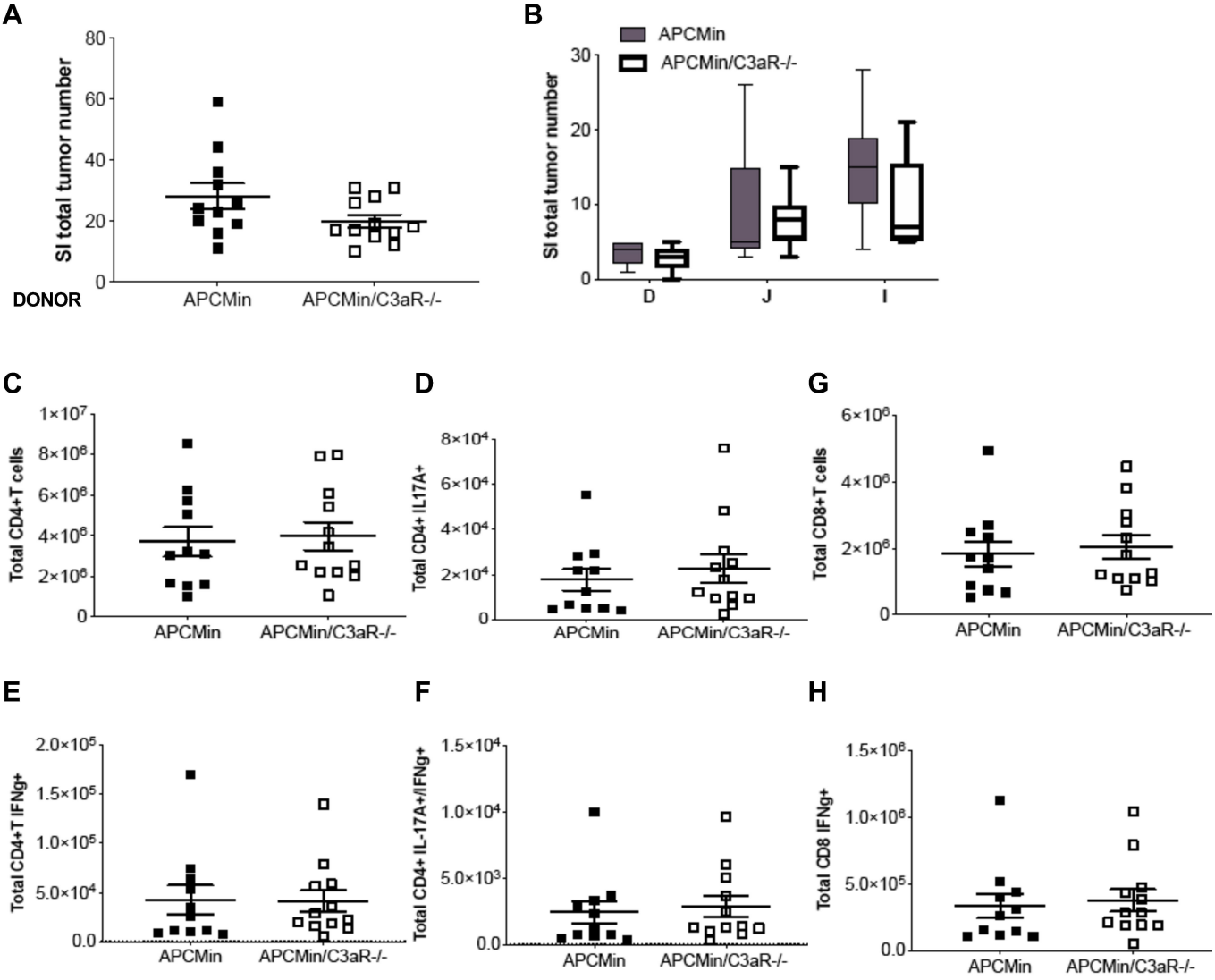
Supplementary Figure 2



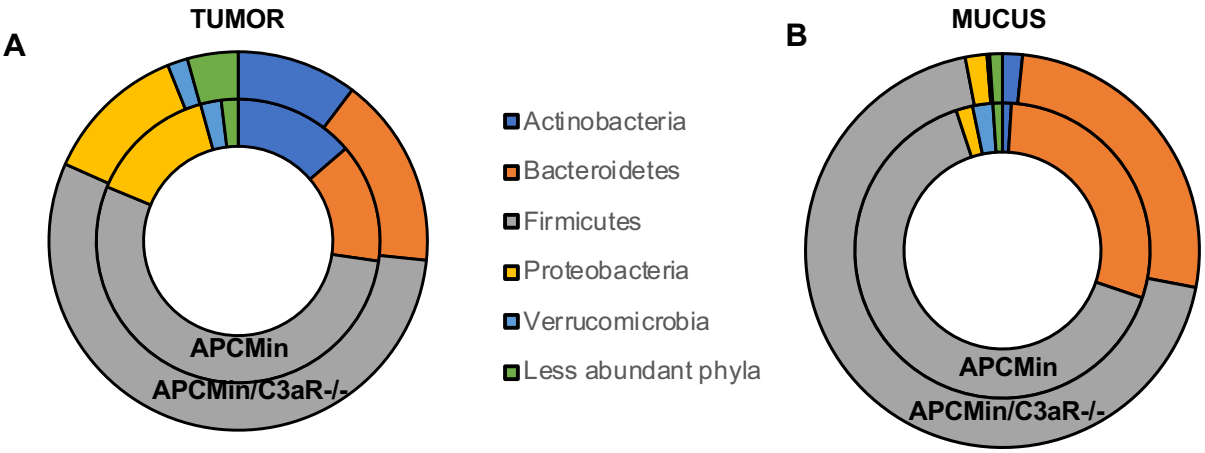
Supplementary Figure 3



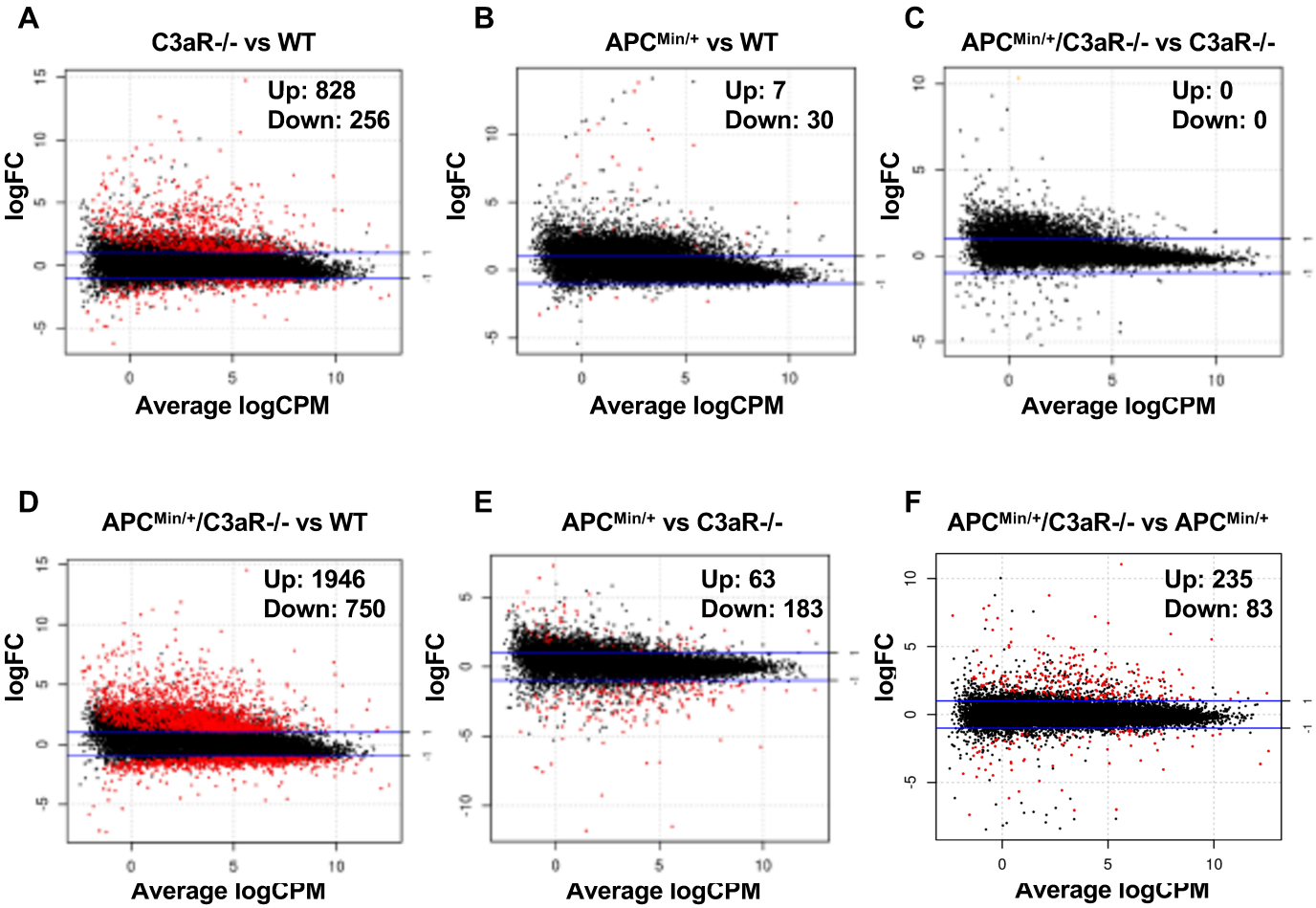
Supplementary Figure 4



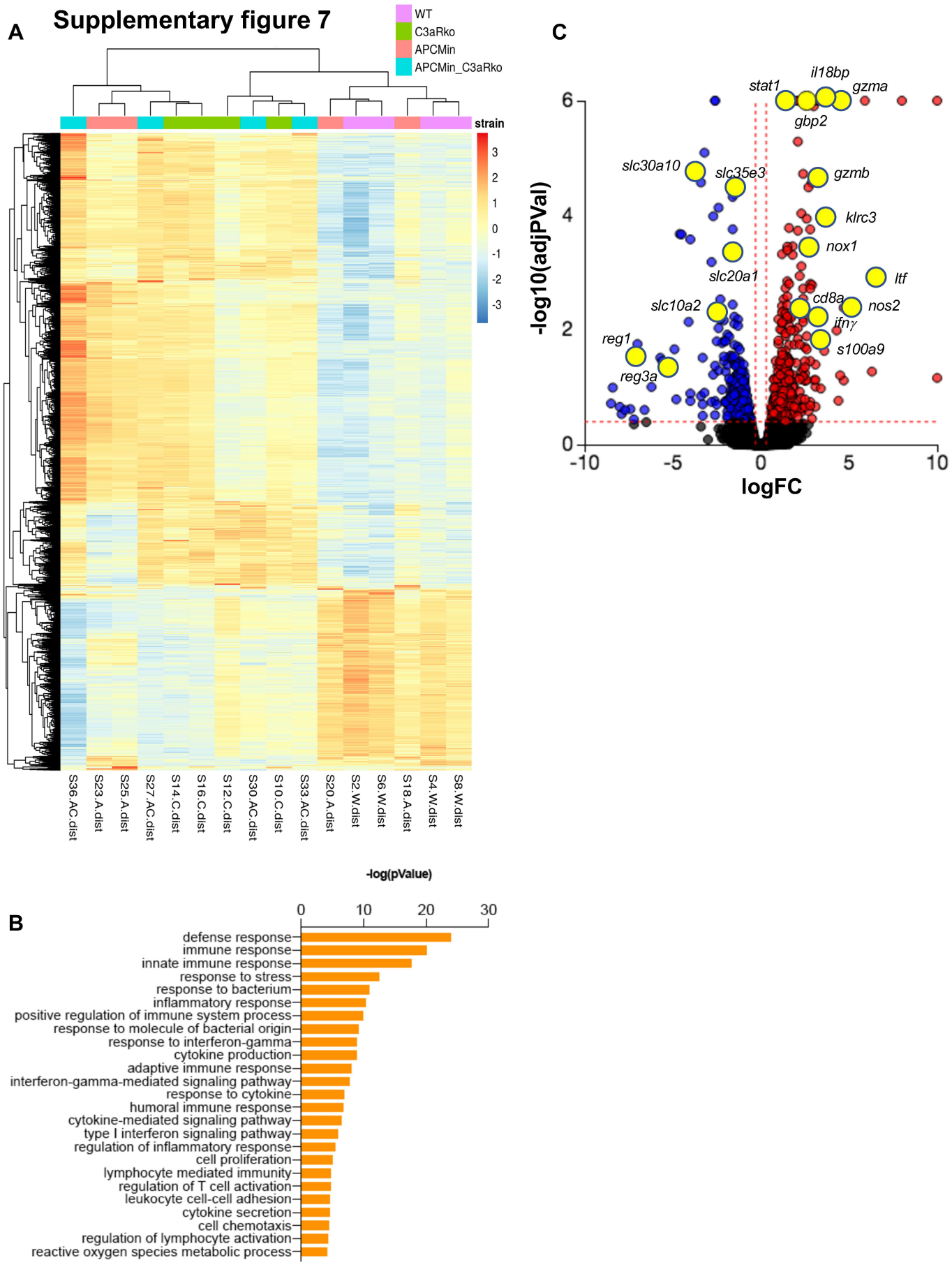
Supplementary figure 5



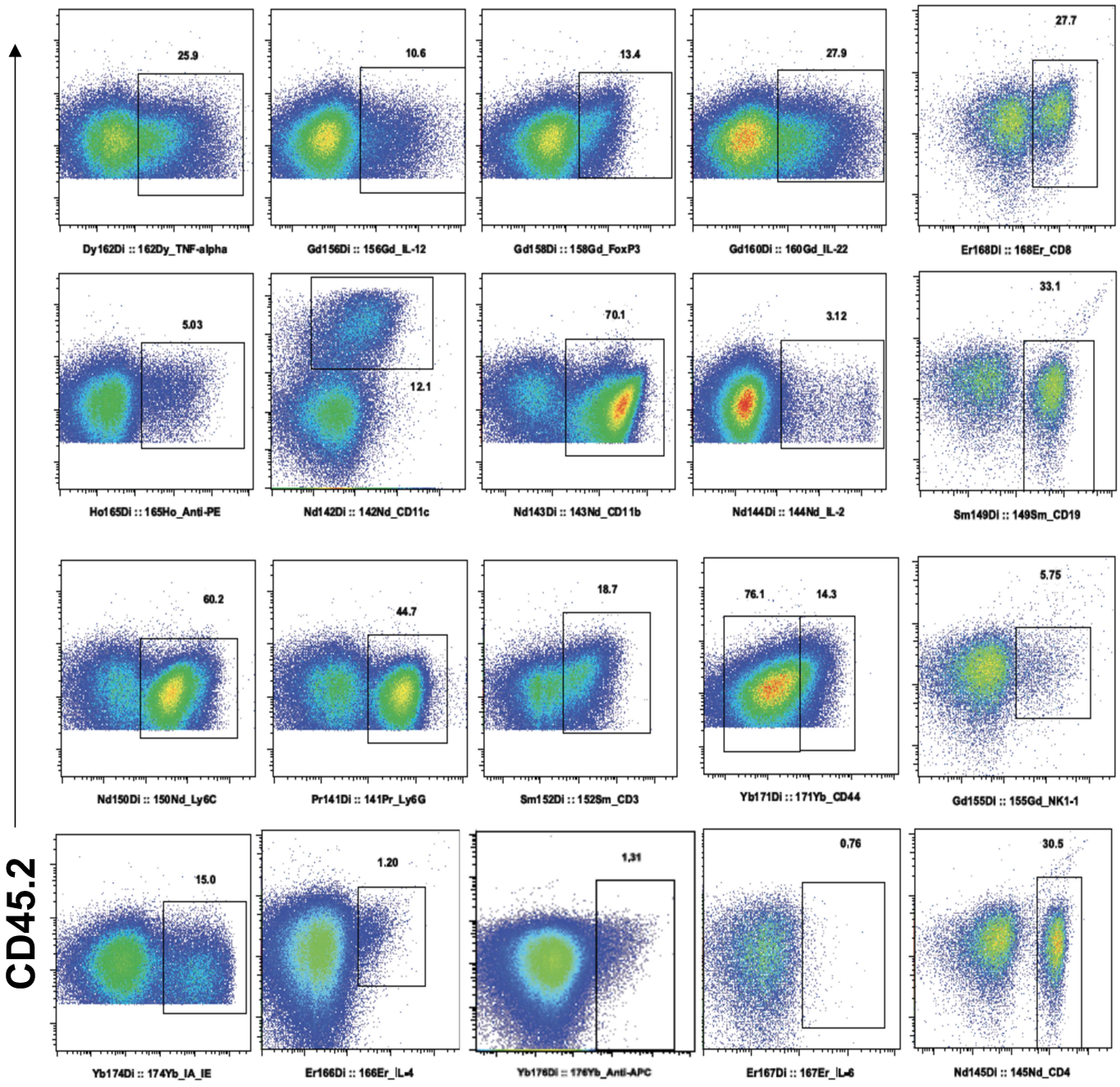
Supplementary Figure 6



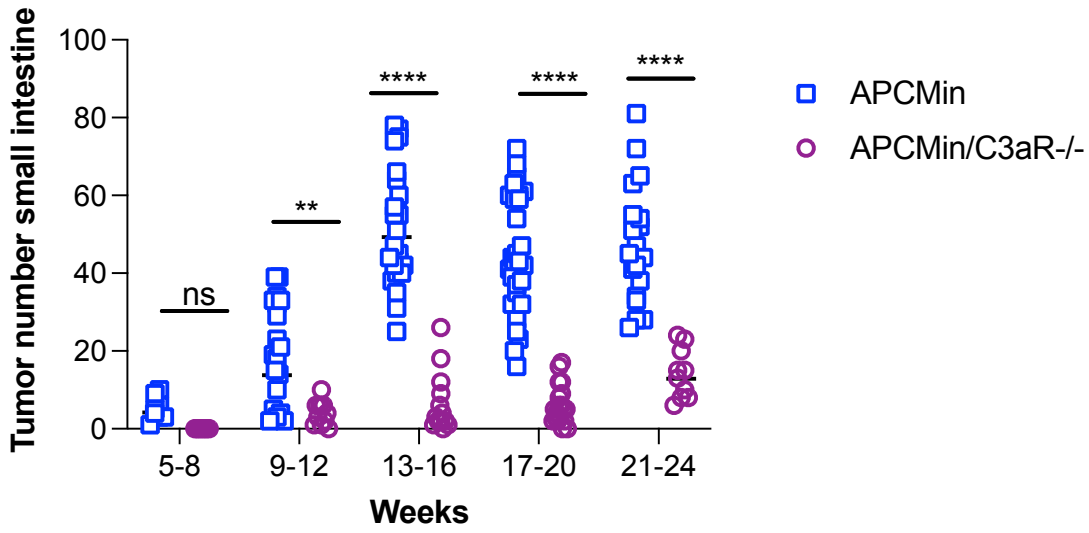
Supplementary figure 7



Supplementary figure 8



Supplementary figure 9



CRC	Cibersort		xcell		MCP	
Population name (Cibersort overlap)	R	p-value	R	p-value	R	p-value
Neutrophils	0.33	7.54E-06	0.50024663	1.81E-07	0.532575	1.97E-08
Macrophages_M0	0.33	6.47E-06	0.40681445	3.55E-05		
CD4 resting	0.28	1.47E-04	0.32985489	0.000968796		
Macrophages_M1	0.25	6.12E-04	0.40650321	3.60E-05		
CD8 T cells	-0.3	4.80E-05	-0.28081781	0.00533336	-0.1731918	0.089796
NK cells activated	-0.28	5.78E-08	-0.34205869	0.000605206	-0.0296307	0.773263
Tregs	-0.39056755	8.30E-08	-0.26286856	0.009286821		
CD4 activated	-0.14	5.79E-03	-0.25352064	0.012224454	-0.0542505	0.59766

RC	Cibersort		xcell		MCP	
Population name (Cibersort overlap)	R	p-value	R	p-value	R	p-value
Dendritic_cells_resting	0.347460005	1.00E-07	0.24110623	0.000279053		
Neutrophils	0.313791277	1.75E-06	0.39828579	6.76E-10	0.3459545	1.15E-07
Macrophages_M1	0.25602375	0.0001105	0.61546371	1.23E-24		
NK_cells_resting	0.219192585	0.0009842	0.26347605	6.80E-05		
T_cells_CD4_memory_resting	0.211791833	0.0014671	0.35240985	6.39E-08	0.0617925	0.358383
B_cells_naive	-0.31469163	1.62E-06	-0.20412594	0.002188083	-0.1006696	0.133957
T_cells_CD8	-0.16122597	0.0159581	-0.35656168	4.36E-08	-0.2197795	0.000953
NK_cells_activated	-0.22254934	0.0008176			-0.1690961	0.011433

CRC= colorectal cancer, RC=rectal cancer, R and p-values from Pearson