Cell Reports, Volume *16*

Supplemental Information

Epithelial IL-23R Signaling Licenses Protective IL-22 Responses in Intestinal Inflammation

Konrad Aden, Ateequr Rehman, Maren Falk-Paulsen, Thomas Secher, Jan Kuiper, Florian Tran, Steffen Pfeuffer, Raheleh Sheibani-Tezerji, Alexandra Breuer, Anne Luzius, Marlene Jentzsch, Robert Häsler, Susanne Billmann-Born, Olga Will, Simone Lipinski, Richa Bharti, Timon Adolph, Juan L. Iovanna, Sarah L. Kempster, Richard S. Blumberg, Stefan Schreiber, Burkhard Becher, Mathias Chamaillard, Arthur Kaser, and Philip Rosenstiel

Supplemental Figures:

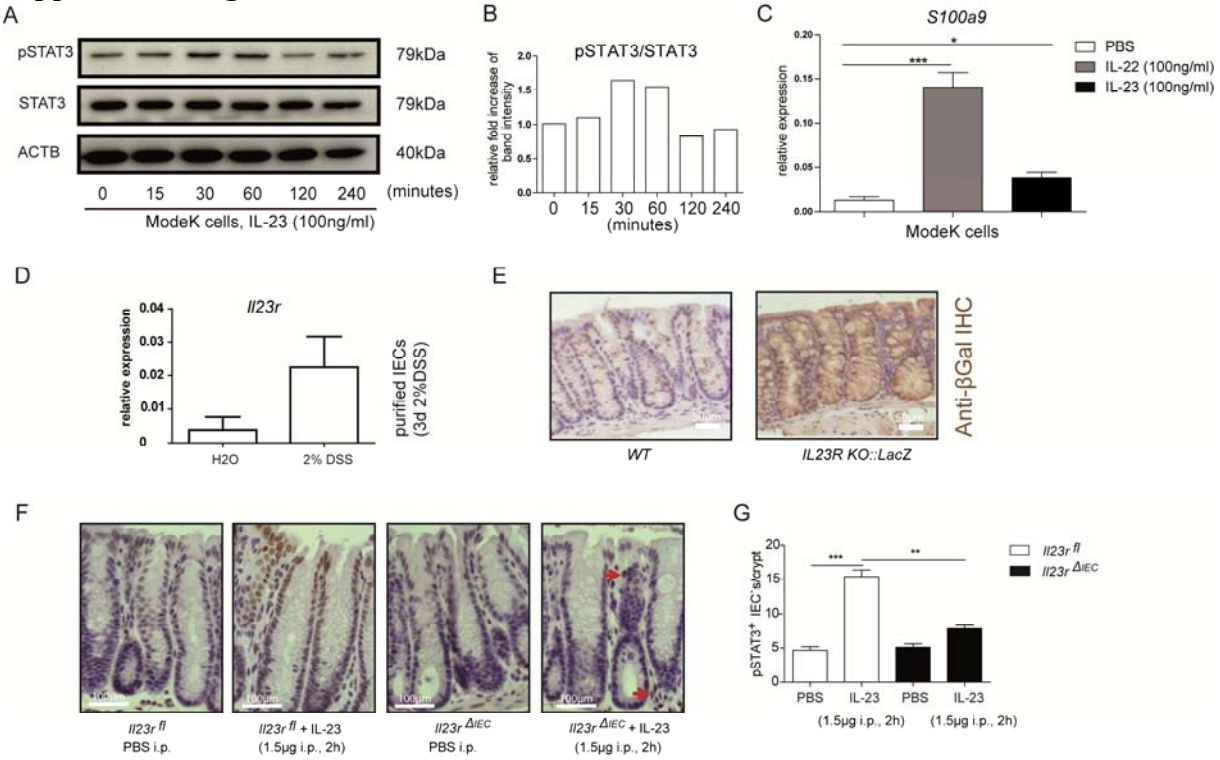

Figure S1, related to Figure 1 *Confirmation of functional IL-23R signaling in intestinal epithelial cells*
Overnight serum starved (FCS 0,1%) Mode-K intestinal epithelial cells were stimulated with 100ng IL-23 for indicated time points. Time dependent activation of pSTAT3 in response to IL-23 stimulation (A). Quantification of immunoblot in (A) normalized to STAT3 and depicted as relative increase of band intensity compared to untreated Mode-K cells (B). Stimulation of Mode-K cells with indicated concentrations of IL-22 or IL-23 for 6h. Cells were subjected to RNA isolation and expression of *S100a9* was determined by real-time PCR (C). Gene expression of *IL23r* in highly purified intestinal epithelial cells after 3d treatment with water or 2% DSS (D). Staining of anti-beta galactosidase in WT or IL-23R KO mice, carrying a lacZ reporter gene integrated in the deletion cassette of the IL-23R gene under the control of the enbdogenous promoter (E). $Il23r^{fl}$ and $Il23r^{\Delta IEC}$ were treated with IL-23 (1,5µg) or PBS for 1h and STAT3 phosphorylation was detected by pSTAT3 staining of intestinal mucosa. Representative pictures showing increase of $pSTAT3^+$ IEC`s in $Il23r^{fl}$ + IL-23 but not in $Il23r^{\Delta IEC}$ + IL-23 mice. Arrows indicate non-epithelial $pSTAT3^+$ cells in lamina propria (F). Statistical evaluation of $pSTAT3^+$ IEC`s per crypt (G). Data are representative of a minimum of 3 animals per group. Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM. *, $p<0.05$.

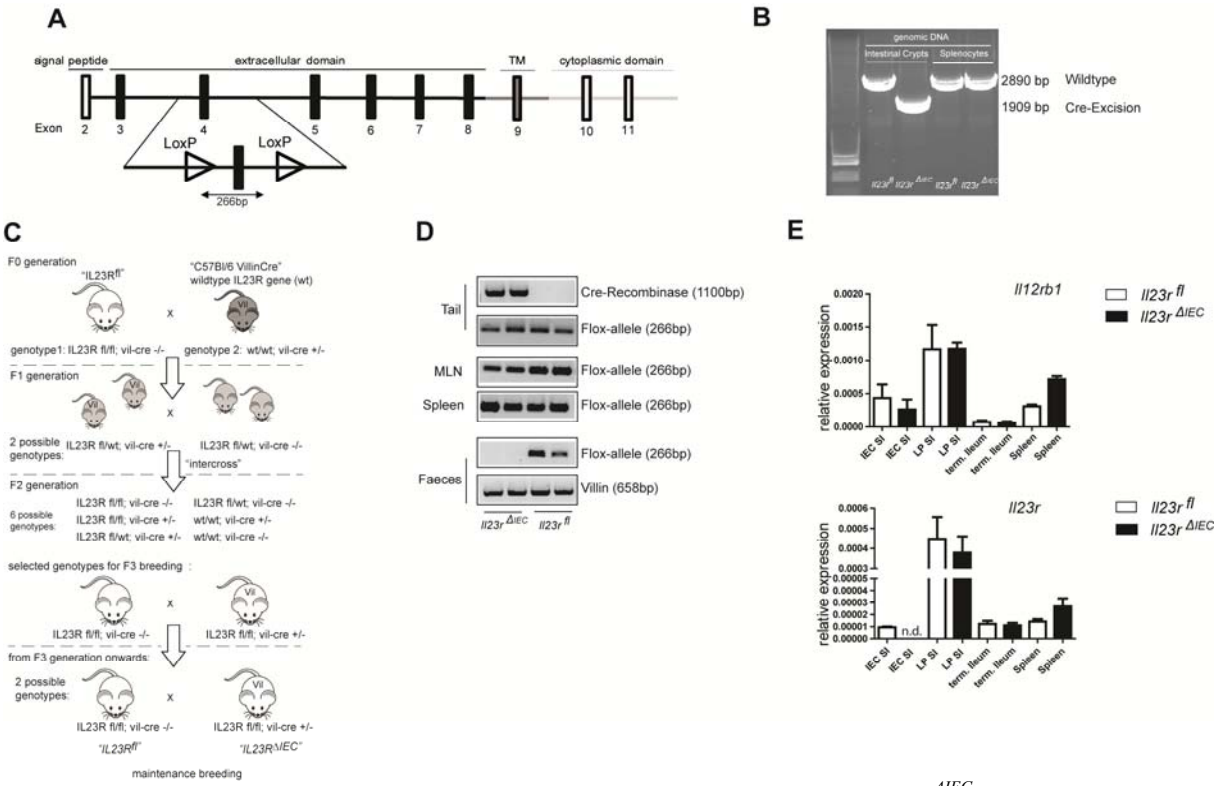

Figure S2, related to Figure 1 *Deletion strategy and confirmation of genotype in Il23r$^{\Delta IEC}$*

Strategy for targeted insertion of LoxP sites into the *Il23r* gene. LoxP sites were inserted around the exon 4, which encodes for the extracellular domain of the *IL23R* (A). Representative gel pictures of genotyping, detecting the epithelial-specific cre-mediated recombination in isolated intestinal crypts but not in splenocytes (B). Breeding scheme for generation of *Il23r^{fl}* (WT) and *Il23r$^{\Delta IEC}$*(KO) mice (C). Representative gel pictures of genotyping, detecting the presence of Cre-Recombinase and the fl/fl region in the genomic tail DNA of *Il23r$^{\Delta IEC}$* and *Il23r$^{fl/fl}$* mice. Epithelial DNA was recovered from colonic fecal samples. PCR with genomic DNA detects presence of the fl/fl region and the IL23R gene only in *Il23r^{fl}* but not in *Il23r$^{\Delta IEC}$* mice (D). Small intestinal epithelial cells (SI IEC), small intestinal lamina propria cells (SI LP) and splenocytes were isolated from *Il23r^{fl}* and *Il23r$^{\Delta IEC}$* mice and Il23R expression in isolated cells and whole terminal ileum (term. Ileum) was determined by qRT-PCR (E). Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM. **, $p<0.01$; ***, $p<0.001$.

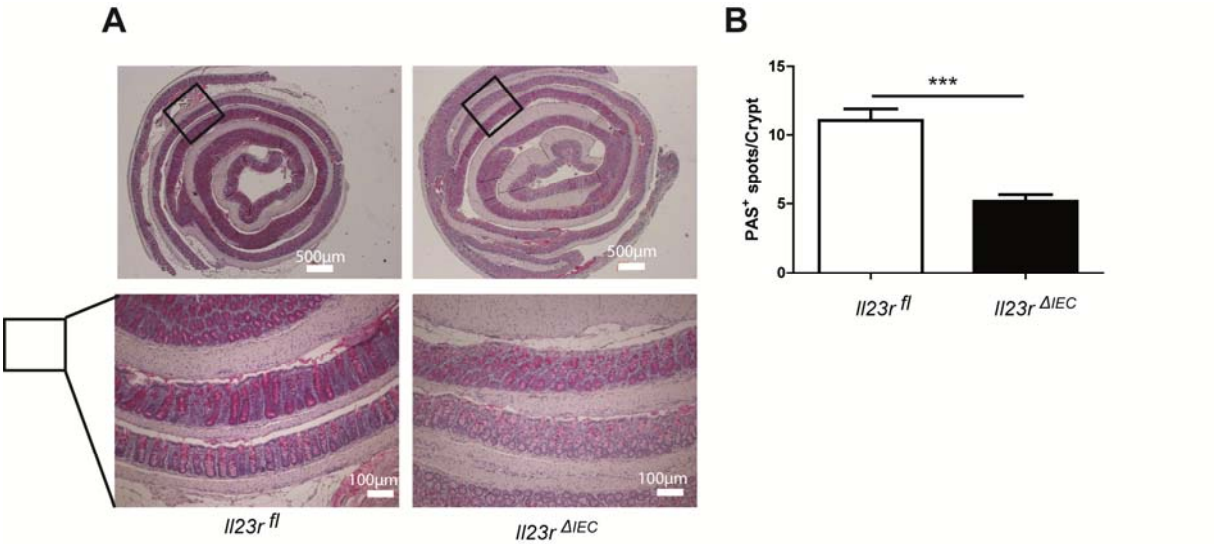

Figure S3, related to Figure 1 *Il23r*$^{\Delta IEC}$ *mice display reduced colonic PAS*$^{+}$ *cells*
Colon swiss rolls of untreated *Il23r*$^{\Delta IEC}$ and *Il23r*^{fl} mice were immunostained with PAS to detect mucus producing goblet cells (A). *Il23r*$^{\Delta IEC}$ show reduced number of PAS^{+} cells per colon crypt (B). PAS^{+} cells were assessed in a minimum of 20 crypts per slide. Data are representative of a minimum of 3 animals per group. Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM. **, $p<0.01$; ***, $p<0.001$.

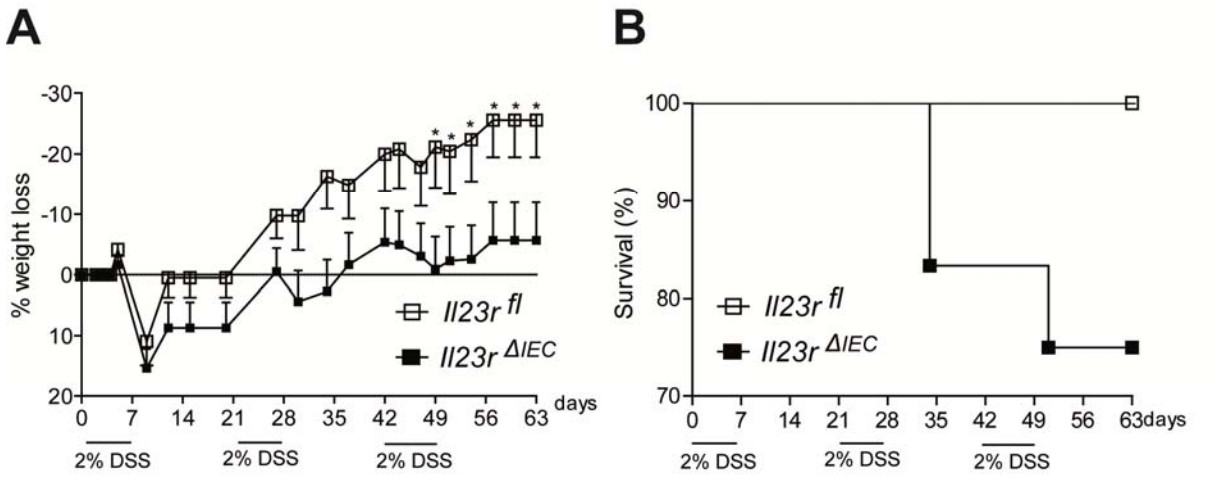

Figure S4, related to Figure 1 *Il23r$^{\Delta IEC}$ mice develop increased mortality and disease activity in a long term chronic DSS-Colitis*
Chronic DSS-Colitis was induced by three cycles of 2%DSS for 7 days followed by 14 days of water in *Il23r$^{\Delta IEC}$* (n = 12, male) and *IL23R^{fl}* (n = 7, male) mice. Weight loss (A) and overall survival (B) were monitored until day 63. Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM.*, p<0.05.

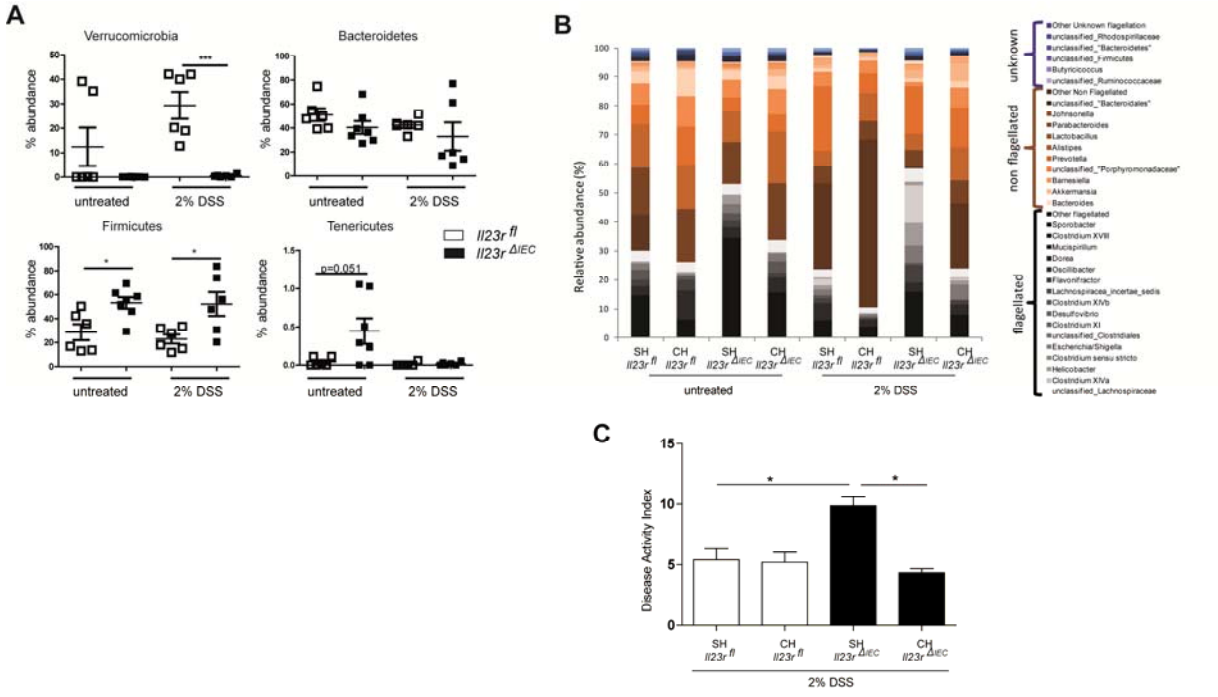

Figure S5, related to Figure 2 *Epithelial IL-23R signaling affects the colonic microbiota*
(A) Relative distribution of dominant bacterial phyla in *Il23r$^{\Delta IEC}$* and *IL23R^{fl}* feces collected before and after DSS treatment. (B) Bar graph showing relative abundance of bacterial groups (with flagellated status) in single and co-housed *Il23r$^{\Delta IEC}$* and *IL23R^{fl}* animals. (C) Disease activity index after induction of chronic colitis according to housing condition (SH: single housed; CH: co-housed).
Non parametric Mann Whitney test was applied to the significance of differences. *p<0.05, **p = 0.005.

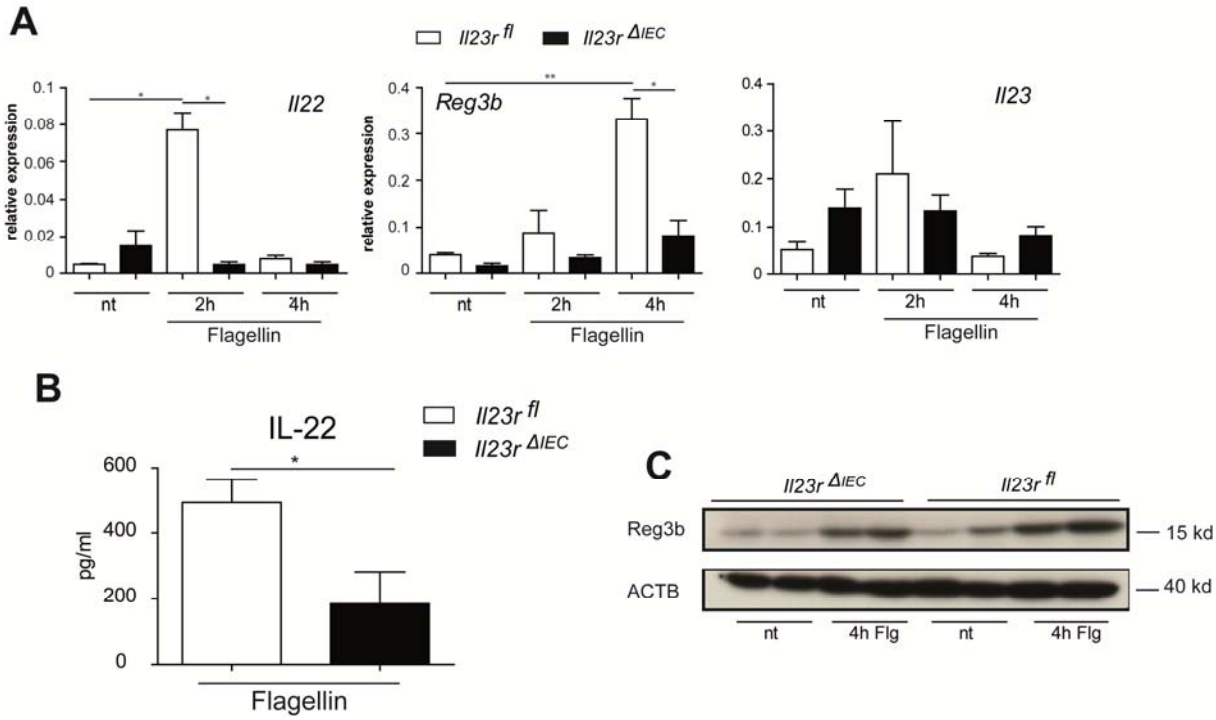

Figure S6, related to Figure 5 *Epithelial IL-23R is involved in the endogenous Il23-dependent IL-22 induction by flagellin stimulation*

Il23r^{fl} and *Il23r$^{\Delta IEC}$* animals were injected (i.p.) weight adapted (1µg/20g body weight) with Flagellin for 2h and 4h to induce endogenous IL23 [1]. Colon intestinal epithelial cells were isolated and assessed for gene transcript levels of *Reg3b* and *IL-22*. Flagellin stimulation induced upregulation of intestinal *IL-22*, which was reduced in epithelial cells from *Il23r$^{\Delta IEC}$* animals (A). IL-22 ELISA of supernatants from post mortem ex vivo colon cultures shows higher IL-22 production in Il23r^{fl} animals. (B). Flagellin induced epithelial Reg3b production is reduced in *Il23r$^{\Delta IEC}$* animals (C). Data are representative of minimum n=3 animals per group. Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM. *, p<0.05; **, p<0.01.

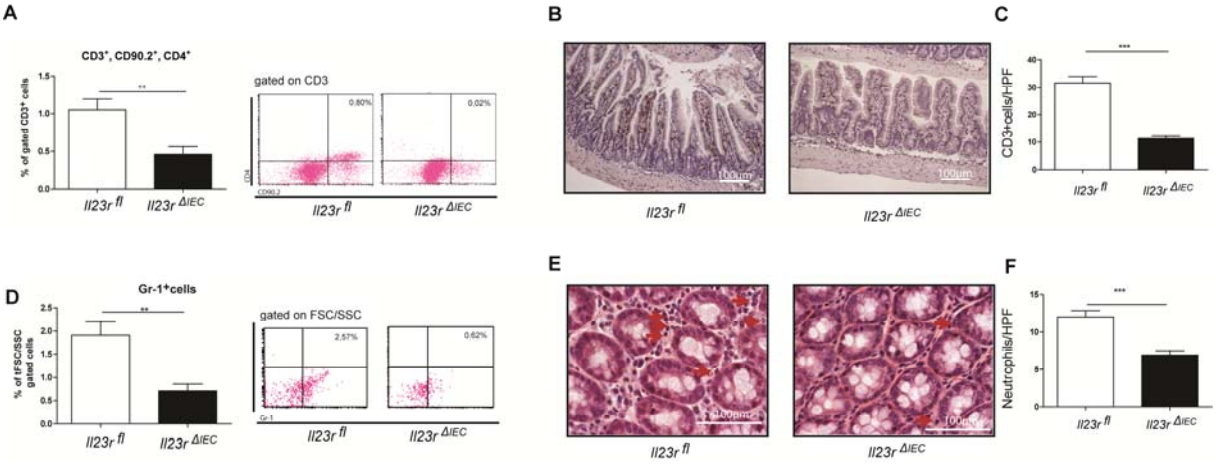

Figure S7, related to Figure 6 *Reduced CD3$^+$ cells and neutrophilic granulocytes in the intestinal lamina propria of IL23R$^{\Delta IEC}$*

Small intestinal lamina propria cells of *IL23R^{fl}* and *Il23r$^{\Delta IEC}$* were isolated and assessed by FACS. LP cells were CD3$^+$ gated and then assessed for the expression of CD4 and CD90.2 (Thy-1) (A). Ileal CD3$^+$ staining in swiss rolls from *Il23r^{fl}* and *Il23r$^{\Delta IEC}$* mice (B) and corresponding statistical analysis (C). Numbers of neutrophilic granulocytes were determined by Gr-1 staining (D) and in histological sections by two independent blinded investigators (neutrophils/HPF) (E,F). Significance determined using two-tailed Student's t test and expressed as the mean +/- SEM. **, p<0.01; ***, p<0.001.

Extended experimental procedures:

FACS

Flow cytometry was performed on a FACSCalibur (B&D, Heidelberg, Germany) and data were analyzed using CellQuest software (B&D, Heidelberg, Germany) . Following anti-murine antibodies were used for analysis: Anti-EpCam-PE, Clone: G8.8 (Biolegend, San Diego, USA); anti- CD3e, Clone: 145-2C11 (Immunotools, Friesoythe, Germanny), anti-CD4, Clone: YTS 191.1.2 (Immunotools, Friesoythe, Germanny); anti-CD90.2, Clone: 53-2.1 (eBioscience, San Diego, USA); anti-Gr-1, Clone: RB6-8C5 (eBioscience, San Diego, USA).

Isolation of primary cells, cultivation of intestinal organoids.

Murine intestinal epithelial cells were isolated using HBSS-EDTA method. In brief, small or large intestine was removed and cleared of intestinal content by flushing the intestine with HBSS (w/o). After removal of residual fat and Peyers patches the intestine was cut longitudinally and then laterally in pieces of 0.5 cm. Intestinal pieces were incubated in HBSS (w/o) containing 5mM EDTA, 5% FCS and 1mM DTT for 20 min. at 37°C and constant rotation. After vigorous shaking epithelial cells were transferred into fresh advanced DMEM/F12, containing 1% P/S, 1% NEAA, 10mM HEPES and 10% FCS. For some experiments $EpCAM^+$ IEC´s were further purified using fluorescent staining with EpCam-PE antibody followed by immunomagnetic beads directed against PE.

For isolation of small intestinal lamina propria, the remaining epithelium-depleted intestinal tissue was further subjected to enzymatic digestion using Lamina propria dissociation kit (Miltenyi®, Bergisch Gladbach, Germany), according to manufacturers protocol. Mouse intestinal organoids were cultivated as described before [2]. In brief, small intestine was removed and cleared of intestinal content by flushing the intestine with HBSS (w/o). After removal of residual fat and Peyers patches the intestine was cut longitudinally and then laterally in pieces of 0.5 cm. Intestinal pieces were incubated in ice cold PBS + 10nM EDTA for 10 min for 4x times, intermitted by vigorous shaking and replacement with fresh PBS + 10nM EDTA after every shaking process. The crypt suspension were then strained through a 100µm strainer, followed by a spin with 1200rpm at +4°C. Pure epithelial crypts were resuspended in BD Matrigel® (BD Bioscience/Heidelberg, Germany), (5-10 crypts/1µl Matrigel) and embedded in 24-well plates and cultivated in Advanced DMEM F12, supplemented with 1% P/S, NEAA, 10mM EDTA, 10%FCS and growth factor cocktail: 500ng/ml R-spondin (R&D,MN, USA), 100ng/ml Noggin (R&D,MN, USA) and 10ng EGF (Peprotech, Hamburg, Germany). Medium was changed every other day and organoids were stimulated after 7 days of cultivation.

Peritoneal cellular recruitment, FACS analysis

Six hours after intraperitoneal administration of purified, endotoxin-free Reg3b (100µg/kg), peritoneal cells were harvested and centrifuged at 400 x g for 10 min at 4°C. An aliquot of 200,000 cells was centrifuged on microscopic slides and relative cellular composition was determined using FACS and light microscopy as described in supplemental methods.

Flagellin Quantification

Fecal flagellin using HEK-Blue hTLR5 cells (Invivogen, Toulouse, France) as reporter assay was determined as described before [15]. Purified FLAG-ST ultrapure flagellin from *Salmonella typhimurium* was used for standard curves.

16S rRNA gene sequencing and analysis

Genomic DNA from feces was extracted with Mobio power soil DNA extraction kit. 16S rRNA gene variable region (V4) was amplified using barcoded (12 base) primers to tag the individual samples [41] and sequenced using standard methods on a Illumina Miseq instrument (for details, see supplementary method section) Sequence analysis was performed in Mothur [42]. Metagenomic contents of microbial communities were predicted from 16S rRNA gene profiles by implementing PICRUSt [14].

Induction of colitis and determination of clinical scores

For chronic colitis induction, mice were supplied with 2% of DSS (molecular mass 40 kDa; TdB consultancy, Uppsala, Sweden) dissolved in drinking water for 5 days followed by 5 days of regular drinking water as described [3]. Body weight, stool consistency and fecal blood loss using the Haemoccult test (Beckman Coulter, Krefeld-Fischeln, Germany) were recorded every other day in order to calculate the Disease Activity Index (DAI), as described previously [4]. Consumption of drinking water was measured daily. High resolution mouse video endoscopic system was used at day 30 of the experiment (Karl Storz AIDA™ VET) and endoscopic scores were obtained as described previously [5]. For induction of short colitis mice were supplied with 2% DSS

dissolved in drinking water for 3 days and were immediately sacrificed for intestinal epithelial cell isolation as described elsewhere [6].

Histopathological analyses of mouse colon tissue

Postmortem, the entire colon was excised and separated longitudinally into two equal parts. The longitudinal section was rolled up, starting with the distal part thereby having the distal colon at the very inner layer and the proximal colon at the very outer layer. The entire specimen was fixed in 4% PFA. Paraffin sections were cut and stained with hematoxylin/eosin. Histological scoring was performed in a blinded fashion by two independent observers. The histologic score displays the combined score of inflammatory cell infiltration and tissue damage as described elsewhere[4]. The percentage of intact $BrdU^+$ mucosa was expressed as *1- (length of BrdU-negative mucosa in µm/total mucosal length in µm)*100*.

Immunohistochemistry and immunofluorescence

For immunohistochemical staining, 5 µm sections of paraffin-embedded colon/ileum swiss rolls were deparaffinized with Xylol substitute (Roth), incubated in citrate buffer for 3 minutes and subsequently blocked with blocking serum (Vectorlab, Peterborough,UK) for 20 minutes. Primary BrdU- (BD Pharmingen, Heidelberg, Germany) and pSTAT3 (Cell Signaling, Cambridge, UK), CD3 (Abcam, Cambridge, UK) or anti-beta galactosidase (Abcam, Cambridge, UK) were incubated over night. Sections were washed, incubated with secondary antibodies and DAB substrate with Vectastain ABC Kit (Vectorlab, Peterborough,UK). For PAS staining slides were dehydrated, rinsed for 5 min. with aqua and 0.5% periodic acid respectively and incubated with Schiff`s reagent for 1 min. Schiff`s reagent was rinsed of for 5 min and slides were counterstained with hematoxylin. Slides were visualized by an AxioImager Z1 microscope (Zeiss, Jena, Germany). Pictures were captured by a digital camera system (Axiocam HrC/HrM, Zeiss). Measurements were made using a semi-automatted image analysis software (Axiovision version 08/2013).

Colon Culture

Colons were washed in cold PBS supplemented with penicillin and streptomycin (GIBCO, Darmstadt, Germany) and cut longitudinally. 4 segments of 0,5cm length were cultured in 24-well flat-bottom culture plates in RPMI 1640 media (GIBCO, Darmstadt, Germany) for 24 h at 37 °C. Supernatants were analysed IL-22 concentration using IL-22 ELISA Kit, according to manufacturers protocol from Peprotech (Hamburg, Germany).

Peritoneal cellular recruitment, FACS analysis

Six hours after intraperitoneal administration of 100 microg/kg of purified, endotoxin-free RegIIIbeta, peritoneal cells were harvested and centrifuged at 400 x g for 10 min at 4°C. An aliquot of 200,000 cells was centrifuged on microscopic slides (cytospin at 1000 rpm for 10 min, at room temperature) and fixed. To determine the relative percentage of each cell type present in the peritoneal lavage fluid, one hundred Diff-Quick stained cells were counted twice under light microscopy. Collected peritoneal cells were washed once in phosphate buffered saline solution containing 0.5% Bovine Serum Albumin (PBS/ 0.5% BSA) and stained on ice at 10^6 cells/100 µL with optimal concentrations of fluorochrome-conjugated antibodies or the corresponding isotype controls. The following primary antibodies were used: anti-CD11b PerCP (clone M1/70), anti-F4/80 PE (clone BM8) for resident macrophages, anti-GR1 PeCy7(clone 1A8) for granulocytes and anti-Ly6C FITC (RBC-8C5) for 20 min in the dark. All antibodies were purchased from BD Pharmingen. After washing with PBS/ 0.5% BSA, cells were analyzed on a Becton Dickinson LSR I analyzer.

Western Blot

Total protein lysates prepared from epithelial cells were separated on 12% polyacrylamide gels under standard SDS-polyacrylamide gel electrophoresis conditions. Proteins were then transferred onto 0.45 µm PVDF membrane and probed pSTAT3/STAT3 (Cell Signaling, Cambridge,UK), Reg3b (Gift of W.D. Hardt, University of Zurich).

Bacterial DNA extraction and 16S rRNA gene sequencing

Total genomic DNA from fecal pellets of single-housed $Il23r^{fl}$ and $Il23r^{AIEC}$ animals was extracted using MoBio Powersoil DNA Isolation kit (*Dianova* GmbH, *Hamburg*, *Germany*) with slight changes: Solution C1 and 20 µl of Proteinase K was added to fecal material and incubated for 2 hrs at 50°C to enhance the bacterial cell lysis. Rest protocol was followed as per the manufacturer instructions. Bacterial communities were profiled by 16S rRNA gene amplicon sequencing. 16S rRNA gene variable region V4 was amplified using 515F and 806R primers [7]. Forwrd primer includes 5 prime illumina adaptor (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), linker (GT) and 16S rRNA gene specific primer (GTGCCAGCMGCCGCGGTAA) and reverse primer consist of reverse complement of 3' Illumina adapter (CAAGCAGAAGACGGCATACGAGAT), golay barcode primer pad (AGTCAGTCAG), linker (CC), and 16S rRNA gene specific reverse primer (GGACTACGCGGGTWTCTAAT).

2 µl of DNA was amplified using above mentioned composite primers, duplicate PCR reactions were performed for each sample. Amplification was performed by Phusion® Hot start Flex 2X master Mix (New England Biolab, Germany) in GeneAmp PCR system 9700 (Applied Biosystems, Foster City, California, USA) using following cycling conditions: an initial denaturation of 3 min at 98°C followed by 30 cycles, denaturation at 98°C for 10 seconds, annealing at 55°C for 30 seconds, elongation at 72°C for 30 seconds and a final extension at 72°C for 10 minutes. PCR performance for quality (expected amplicon size) and quantity (band intensity) was assessed by running aliquot of amplified products on agarose gel. Quantitative normalization was performed using SequalPrep kit (Invitrogen) to pool equal amount of amplicons per sample. Sequencing was performed by Illumina MiSeq (2 x 250 sequencing kit), which allowed us to generate paired end reads with entire overlapping in contig generation.

Analysis of sequences

Sequencing reads were primarily processed for quality control using the software mothur package (Schloss et al., 2009). Forward and reverse reads (fastq) were assembled to contigs sequences, and discarded if, were more than 275 bases in length, having any ambiguous base or more than 8 homopolymers. Sequences were aligned against mothur curated silva alignment database and screened to have alignment in amplified specified (V4 region) region only. Chimeric sequences were detected with Uchime [8] algorithm and were also removed. Sequences were assigned taxonomically using mothur formatted Ribosomal Data base Project (RDP) training sets [9] and eliminated if classified as unknown, archaea, eukaryotes, chloroplast and mitochondria. Phylip formatted distance matrix was computed from remaining quality aligned sequences. Sequences with at least 97% similarity were clustered in into species Operational Taxonomical Units (OTUs) using neighbor joining algorithm within mothur.

Multivariate analysis and statistics

OTU abundance were normalised by subsampling to the lowest number of sequences within analysed groups. Percent abundance of OTUs was log transformed to perform Principal coordinate Analysis (PCoA). One way Non Parametric MANOVA (NPMANOVA) was performed to test the statistical significance of microbial community differences among groups (genotype and treatment). Both PCoA and NPMANOVA were performed in PAST software and principle coordinates were visualized in the vegan package v.2.0-10 in R softwarev 3.0.3 (ref). Statistical significance of bacterial groups in experimental groups was calculated by Nonparametric Kruskal wallis test and Mann-Whitney U test in Graphpad Prism 5.

Functional metagenome prediction

Functional metagenomics contents of microbial communities were predicted from 16S rRNA gene profiles by implementing PICRUSt [10]. In summary, 16S rRNA gene sequences were mapped against Green Genes reference database (August 2013, gg_13_8_99) to pick close reference OTUs. Shared OTUs file was rarefied to equal depth of sequences per samples and a biom file was generated from shared OTU table using Picrust parameter. Both steps were performed in Mothur. Resulted biom file was subjected to online PICRUST galaxy terminal (http://huttenhower.sph.harvard.edu/galaxy/). OTUs were further normalized for the predicated 16 S RNA gene copy number and KEGG pathway functions were predicted at level 1 - 3 with the use of KEGG Orthologs Green Genes 13_5 database in Picrust. Finally predicted metagenomic profile was analyzed in STAMP [11] to test the significance of predicted functional differences in microbial communities. One-way analysis of variance (ANOVA) with Benjamini Hochberg FDR correction and two sided Welch's t test was performed for multiple and two group comparision.

Analysis of flagellated bacteria

16S rRNA gene sequences were classified at genus level using ribosomal database project (RDP) reference and taxonomy training sets. Bacterial genus or group was considered to be flagellated if mentioned in Bergeys manual of systematic bacteriology 2nd edition as suggested or were alternatively searched for its flagellated status in already published work Genus was considered to be flagellated if representative species was flagellated[12]. Clostridium groups XIVa, sensu strict, XI, XIVb were considered to be flagellated as most of the bacterial genera within these clostridium groups are flagellated. Sequences not classified to genera level were further annotated by RDP seqmatch to identify closest known isolate and flagellated status was defined as stated above. Bacterial groups either not available in Bergeys manual or without precise information about flagellation were searched for their flagellation status in original publications. Bacterial genus or groups without specific information were considered to be unknown.

Real Time TaqMan

TaqMan assay was purchased from Applied Biosystems. A list of all TaqMan is below. Total RNA (1µg) was reverse-transcribed to cDNA according to the manufacturer`s instructions (MultiScribe Reverse Transcriptase; Applied Biosystems, Foster City, USA). Reactions were carried out on the PRISM Sequence 7700 Detection

System (Applied Biosystems, Foster City, USA). Error bars represent SEM of min. 3 individual samples and expression levels are depicted as relative to expression of β-actin.

TaqMan Probes and Primers

Gene Name	Symbol	Species	TaqMan Probe ID
Interleukin 22	IL-22	Mouse	Mm01226722_g1
regenerating islet-derived 3 beta	Reg3b	Mouse	Mm00440616_g1
regenerating islet-derived 3 gamma	Reg3g	Mouse	Mm01181783_g1
Interleukin 23 receptor	Il23r	Mouse	Mm00519943
ß-actin	Actb	Mouse	NM_007393.1
Interleukin 23	IL-23	Mouse	Mm01160011_g1

Supplemental References:

1 Kinnebrew, Melissa A. *et al.* Interleukin 23 Production by Intestinal CD103+CD11b+ Dendritic Cells in Response to Bacterial Flagellin Enhances Mucosal Innate Immune Defense. *Immunity* **36**, 276-287, (2012).

2 Sato, T. *et al.* Single Lgr5 stem cells build crypt villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265, (2009).

3 Sina, C. *et al.* Extracellular cathepsin K exerts antimicrobial activity and is protective against chronic intestinal inflammation in mice. *Gut* **62**, 520-530, doi:10.1136/gutjnl-2011-300076 (2013).

4 Siegmund, B., Lehr, H.-A., Fantuzzi, G. & Dinarello, C. A. IL-1β-converting enzyme (caspase-1) in intestinal inflammation. *Proceedings of the National Academy of Sciences* **98**, 13249-13254, doi:10.1073/pnas.231473998 (2001).

5 Wirtz, S., Neufert, C., Weigmann, B. & Neurath, M. F. Chemically induced mouse models of intestinal inflammation. *Nat. Protocols* **2**, 541-546 (2007).

6 Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *The Journal of Experimental Medicine* **206**, 1465-1472, doi:10.1084/jem.20082683 (2009).

7 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621-1624, (2012).

8 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200, doi:10.1093/bioinformatics/btr381 (2011).

9 Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37**, D141-D145, doi:10.1093/nar/gkn879 (2009).

10 Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotech* **31**, 814-821, doi:10.1038/nbt.2676 (2013)

11 Parks, D. H. & Beiko, R. G. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**, 715-721, doi:10.1093/bioinformatics/btq041 (2010).

12 Liu, R. & Ochman, H. Stepwise formation of the bacterial flagellar system. *Proceedings of the National Academy of Sciences* **104**, 7116-7121, doi:10.1073/pnas.0700266104 (2007).