Supplementary Materials and Methods

Immunofluorescence and immunohistochemistry

For immunofluorescence, unstained slides were deparaffinized, followed by antigen retrieval using Trilogy (Cell Marque #920P-09; Rocklin, CA). The slides were then blocked with 1% bovine serum albumin and 0.1% Trizol in phosphate buffer saline for 1 hour, before applying primary antibody overnight. After overnight incubation, secondary antibody was applied for 45 minutes, before washed and applied Hoechst stain (1:20,000; Sigma #33258). The following primary antibodies were used: lysozyme (1:100; Santa Cruz# sc-27958), defensin 5 (1:2000; Novus #NB110-60002) and cleaved caspase-3 antibody (1:200; Cell Signaling #9661). Secondary antibodies used include donkey anti-rabbit, donkey anti-goat, and donkey anti-mouse antibodies (1:500; Life Technologies #A-21207, A-11058, and A-21202, respectively). For Ki-67 immunohistochemistry, unstained slides were deparaffinized, followed by antigen retrieval using Diva Decloaker buffer (Biocare Medical). The slides were then quenched with 10% H_2O_2 in methanol and blocked as described above. The rabbit anti-human Ki-67 antibody was then applied (1:400; Lab Vision #RB9043R7) overnight, followed by goat anti-rabbit biotinylated secondary antibody (1:200; Thermo Fisher Scientific #31820). Signal detection was performed by using Vectastain Elite ABC kit (Vector #PK-6100) and DAB (Vector #SK-4100). The stains were reviewed using an Olympus BX53 microscope. Representative images were taken with Olympus DP73 camera with cellSens Dimension software for adjusting brightness and contrast and image cropping.

TUNEL staining

TUNEL staining was performed using the ApopTag® Fluorescein in situ apoptosis detection kit (Millipore #S7110) following the vendor's instructions.

PCR Amplification and Sequencing of Bacterial 16S rRNA Genes

Fourteen PCR amplicons, representing all nine 16S variable regions, were constructed using the Fluidigm Access Array System. 5ng/ul of DNA were input into each reaction. The sample inlets consisted of 1X High Fidelity FastStart Reaction Buffer without MgCl2 (Roche), 4.5nM MgCl2 (Roche), 5% DMSO (Roche), 200uM PCR Grade Nucleotide Mix (Roche), 0.05 U/µL 5 U/µL FastStart High Fidelity Enzyme Blend (Roche), 1X Access Array Loading Reagent (Fluidigm), 1ul DNA, and water. The primers were added to the assay inlets at 200nM forward and reverse primers with 1X Access Array Loading Reagent. The samples were harvested after PCR amplification was performed on the BioMark HD system from Fluidigm. Each sample was harvested and indexed using unique 10 base pair sequences with 14 rounds of PCR to incorporate each index sequence. All samples were pooled into 48 sample libraries and cleaned using bead purification. The samples were loaded on Miseq instruments and sequenced.

16s rRNA sequencing data analysis

Reads from the V4 region amplicon were used for downstream analyses. The paired end reads were joined into single sequences using fastq-join (70). Each sequence was assigned as belonging to a specific amplicon by detecting the primer sequences. Analysis of the reads was performed using the QIIME pipeline version 1.9.0 (71). Every amplicon and sample pair was considered a separate sample in the QIIME analysis. Open-reference operational taxonomic units

(OTU) were called using the Greengenes May 2013 release as the reference database (72). Reads were clustered into OTUs by QIIME using UCLUST (73) at a threshold of 97% similarity. Representative sequences for each OTU were classified taxonomically with the UCLUST consensus taxonomy assigner in QIIME using a sequence similarity of 0.9. Analysis of alpha diversity, including Faith's phylogenetic diversity (74) and Shannon diversity, was performed using alpha_diversity.py in QIIME on rarefied data. Beta-diversity was determined in QIIME using beta_diversity_through_plots.py to determine unweighted UniFrac distances and to generate principal coordinates analysis plots through Emperor (75).

Laser capture microdissection

The distal ileum of the mice was prepared and fixed in methacarn as previously reported (19). The crypt base epithelial cells enriched for Paneth cells were captured and RNA isolated by previously described techniques (14, 19).

Transcriptomic profiling and analysis

For RNA-seq, fragments of mouse small intestine tissue were preserved in RNAlater. RNA was isolated using Qiagen RNeasy Mini kit, according to the kit protocol. Total RNA was quantified using the Quant-iT™ RiboGreen® RNA Assay Kit and normalized to 4ng/ul. Transcriptomic profiling was performed at the Genome Technology Access Center (GTAC) at Washington University, which has performed microarray (14) and RNA-seq before. RNA library preparation was performed with RiboZero and sequencing performed on an Illumina HiSeq 2500.

RNA-seq reads were aligned to the GRCm38.76 assembly from Ensembl with STAR version 2.0.4 (76). Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 1.4.5. Transcript counts were produced by Sailfish version 0.6.3. Sequencing performance was assessed for the total number of aligned reads, total number of uniquely aligned reads, genes and transcripts detected, ribosomal fraction and known junction saturation and read distribution over known gene models with RSeQC version 2.3.

For microarray, extracted RNA was sent to GTAC for microarray analysis. Nucleic acid hybridization was performed using the Agilent Mouse 8x60 array (G4852B-60520 v2) according to the vendor's instruction.

Gene CPM differences were further assessed for each of the 4 expression pattern groups of G (T300A vs. WT), E (Smoking vs. Non-Smoking), G+E (T300A Smoking vs. All), and G+E (WT Smoking vs. All) utilizing a two tailed equal variance t-test. For both microarray and RNA-seq, genes within each group with $P < 0.05$ were submitted for transcription factor, pathway, and GO analysis employing the Enrichr toolset (42, 43). The Enrichr categories for transcription factors (ChEA, ENCODE, ChEA and ENCODE consensus), pathways (KEGG, Wiki, Reactome, BioCarta) and GO Biological, Cellular, and Molecular Processes were generated and filtered with an adjusted p-value cutoff of 0.05 for additional stringency. These categorical data were then analyzed to identify key transcriptional and/or regulatory processes with multiple lines of evidence for further laboratory validation.

Comparison of three LCM Paneth cell microarray datasets

The microarray from LCM Paneth cells from the $Atg16l1^{T300A}$-smoking study was analyzed, and the top 500 genes that were significantly downregulated (adjusted P value < 0.05) in the $Atg16l1^{T300A}$-smoked mice (versus all three other groups, similar to the analysis performed for full thickness ileal RNA-seq) were identified by the Enrichr toolset (ChEA). Among these genes, those that were associated with the PPARγ pathway were further identified. Similar analyses were performed on two previously published LCM Paneth cell microarray datasets: $Atg16l1^{HM}$ mice with or without MNV infection (ArrayExpress accession number: E-TABM-957; (19)), and CD subjects E-MTAB-1281; (14)). Significantly downregulated $Pparg$ associated genes in the three datasets were compared. A Venn diagram demonstrating the similar and different genes amongst the three datasets were constructed.

Quantitative RT-PCR

Purified RNA was used to synthesize cDNA using SuperScript III (Invitrogen), and qPCR was performed using SYBR Green reagents (Clontech, Palo Alto, CA) on an Eppendorf Mastercycler. Primer sequences for mouse *Igf1* are: 5' → 3': CATTGCTCTAACATCTCCCATCTC, 3' → 5': AGAGGTGTGAAGACGACATGAT. Primer sequences for mouse *Pck1* are: 5' → 3': CGGATGGGCATATCTGTGCT, 3' → 5': AGGCCCAGTTGTTGACCAAA. Relative expression levels were normalized to *Gapdh*, which were expressed at similar levels in all samples. qPCR for MNV was performed as described previously (19).

Serum inflammatory markers detection

Mouse serum was tested with the following ELISA kits: CXCL1, CXCL2 (both from Ebioscience), RAGE (Raybiotech), TNF-α, IL-1β and IL-6 (from Biolegend) according to the instructions of the vendors. MPO level was detected based on previously published protocol (77).

Transmission electron microscopy

For mouse studies, tissue was fixed in a modified Karnovsky's fixative of 3% glutaraldehyde, 1% paraformaldehyde in 0.1M sodium cacodylate buffer and then post-fixed in 2% osmium tetroxide in 0.1 M sodium cacodylate buffer for 1 hr, en bloc stained with 3% aqueous uranyl acetate for 30 min, dehydrated in graded ethanols and embedded in PolyBed 812 catalog #08792-1 (Polysciences, Hatfield, PA). Tissue blocks were sectioned at ninety nanometers thick, post stained with Venable's lead citrate and viewed with a JEOL model 1400EX electron microscope (JEOL, Tokyo, Japan). Digital images were acquired using the AMT Advantage HR (Advanced Microscopy Technology, Danvers MA) high definition CCD, 11 megapixel TEM camera. Paneth cells were identified by the presence of cytoplasmic granules. Ultrastructural changes with cytoplasmic vesicles and degenerative mitochondria were used for quantification. For human samples, formalin-fixed paraffin-embedded pathology blocks were used for sectioning. Due to fixation-associated artifact, quantification was not performed.

Supplementary references

70.   Aronesty E. ea-utils : Command-line tools for processing biological sequencing data. http://code.google.com/p/ea-utils.

71.   Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods.* 2010;7(5):335-6.

72.   McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, and Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal.* 2012;6(3):610-8.

73.   Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460-1.

74.   Faith DP, and Baker AM. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary bioinformatics online.* 2006;2(121-8.

75.   Vazquez-Baeza Y, Pirrung M, Gonzalez A, and Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience.* 2013;2(1):16.

76.   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.

77.   Kim JJ, Shajib MS, Manocha MM, and Khan WI. Investigating intestinal inflammation in DSS-induced model of IBD. *J Vis Exp.* 201260).
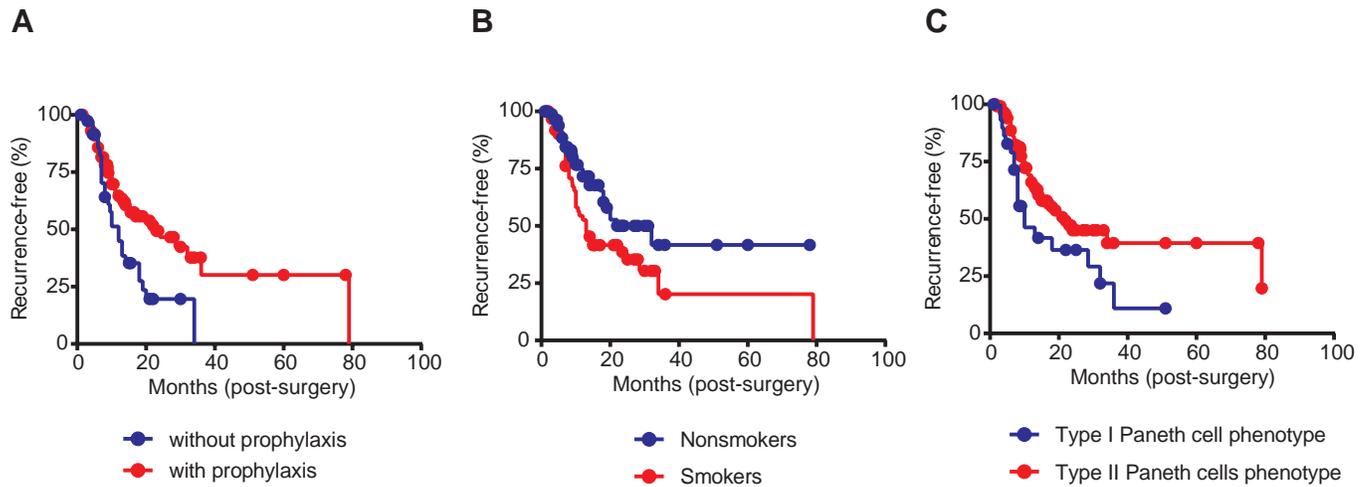
Fig. S1

Fig. S1. Various clinical and pathologic parameters affected outcome in postoperative CD. (**A**) Patients who received postoperative prophylaxis (n = 126) had longer time to disease recurrence compared to those who did not receive prophylaxis (n = 59)(P = 0.0049 by Log-rank test; multivariate analysis P = 0.0359). (**B**) Patients who were nonsmokers (n = 124) had longer time to disease recurrence compared to those who were smokers (n = 63)(P = 0.0316 by Log-rank test; multivariate analysis P = 0.0223). (**C**) Patients with Type II Paneth cell phenotype (>80% Paneth cells with normal morphology; n = 157) had longer time to disease recurrence compared to those with Type I Paneth cell phenotype (≤80% Paneth cells with normal morphology; n = 30)(P = 0.0251 by Log-rank test; multivariate analysis P = 0.0232). Only subjects with prophylaxis (n=126) were used in the outcome analysis in Fig. 1A.
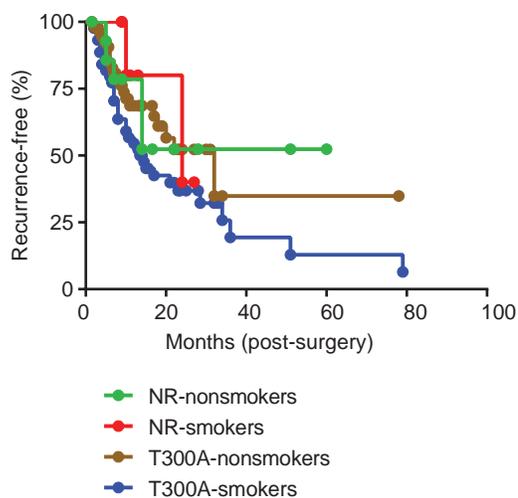
Fig. S2

Fig. S2. *ATG16L1*[T300A] genotype, smoker status, and correlation with post-operative recurrence. When the CD subjects in Fig. 1A were re-classified by *ATG16L1*[T300A] genotype and smoker status, the *ATG16L1*[T300A]-smoker group showed significantly shorter time to recurrence after surgery compared to all three other groups combined ($P = 0.0371$ by Log-rank test).
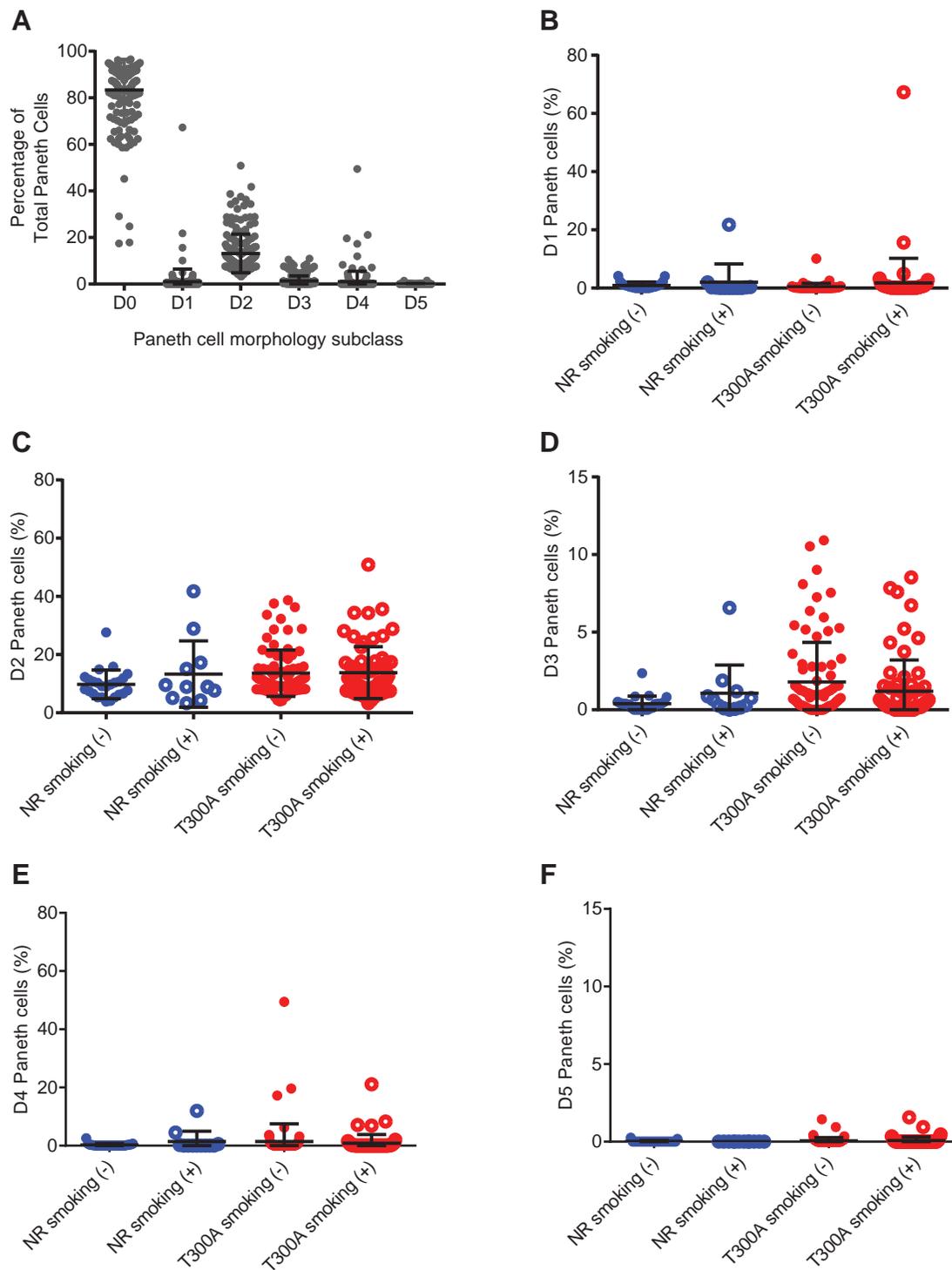
Fig. S3

Fig. S3. Distribution of Paneth cell morphology categories in CD subjects shown in Fig. 1C (n=126). The different categories of Paneth cell morphology subclasses (56) were used. In brief, normal morphology was designated as D0 (as in Fig. 1C). D1-D5 represent 5 different subclasses of abnormal morphology, and D0 + (D1+D2+D3+D4+D5) = 100%. (A) Distribution of different subclasses of abnormal Paneth cell subclasses in all CD subjects (including both genotypes and smoking status). (B-F) Percentages of abnormal Paneth cells by subclass in CD subjects stratified by genotype and smoking status as in Fig. 1 and 5 ($P > 0.05$ in B-F by Kruskal-Wallis tests). Data represent mean ± SEM.
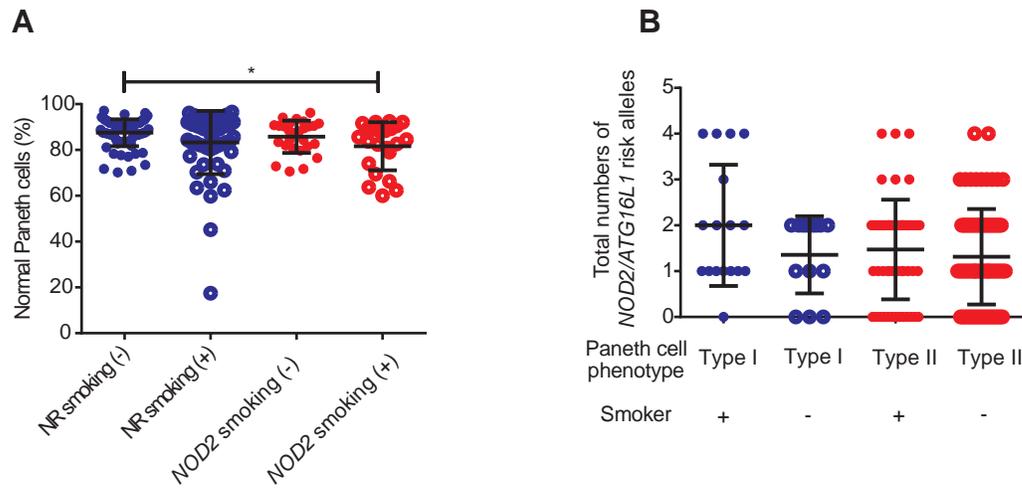
Fig. S4

Fig. S4. Common *NOD2* variants did not correlate with Paneth cell defects in CD subjects. (**A**) No significant Paneth cell defects were seen between subjects carrying *NOD2* variants with or without smoking history ($P = 0.1055$), or in NR patients with or without smoking history ($P = 0.1919$). NR nonsmokers n=76, NR smokers n=55, *NOD2* nonsmokers n=27, *NOD2* smokers n=22. Data represent mean ± SEM. (**B**) Smokers with Type I Paneth cell phenotype did not have more total numbers of *ATG16L1*[T300A] or common *NOD2* risk alleles ($P = 0.3100$ by Kruskal-Wallis test).
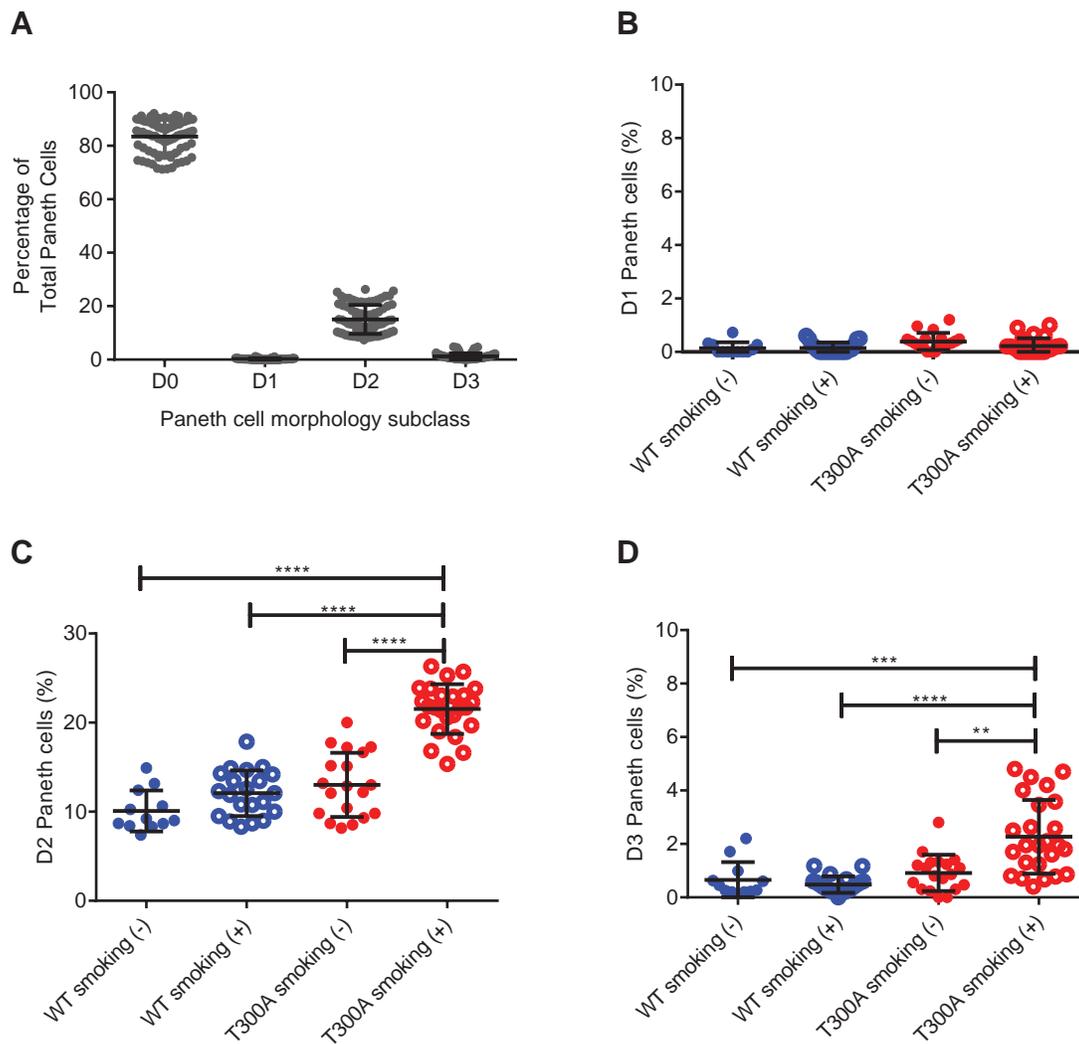
Fig. S5

Fig. S5. Distribution of Paneth cell morphology subclasses in mice shown in Fig. 2 and 5. Note that subclasses D4 and D5 were not observed in mice, and D0 + (D1+D2+D3) = 100%. (A) Distribution of different subclasses of abnormal Paneth cell subclasses in all mice (including both genotypes and smoking exposure). The majority of the abnormal Paneth cells were of the D2 (decreased granules) subclass. (B-F) Percentages of abnormal Paneth cells by subclass in mice stratified by genotype and smoking exposure as in Fig. 2 and 5. (B-D) Data were analyzed by Kruskal-Wallis test followed by Dunn's multiple comparison tests between groups. WT-nonsmoked: n = 12; WT-smoked: n = 21; T300A-nonsmoked: n = 19; T300A-smoked: n = 25. Results of 6 independent experiments. **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$. Data represent mean ± SEM.
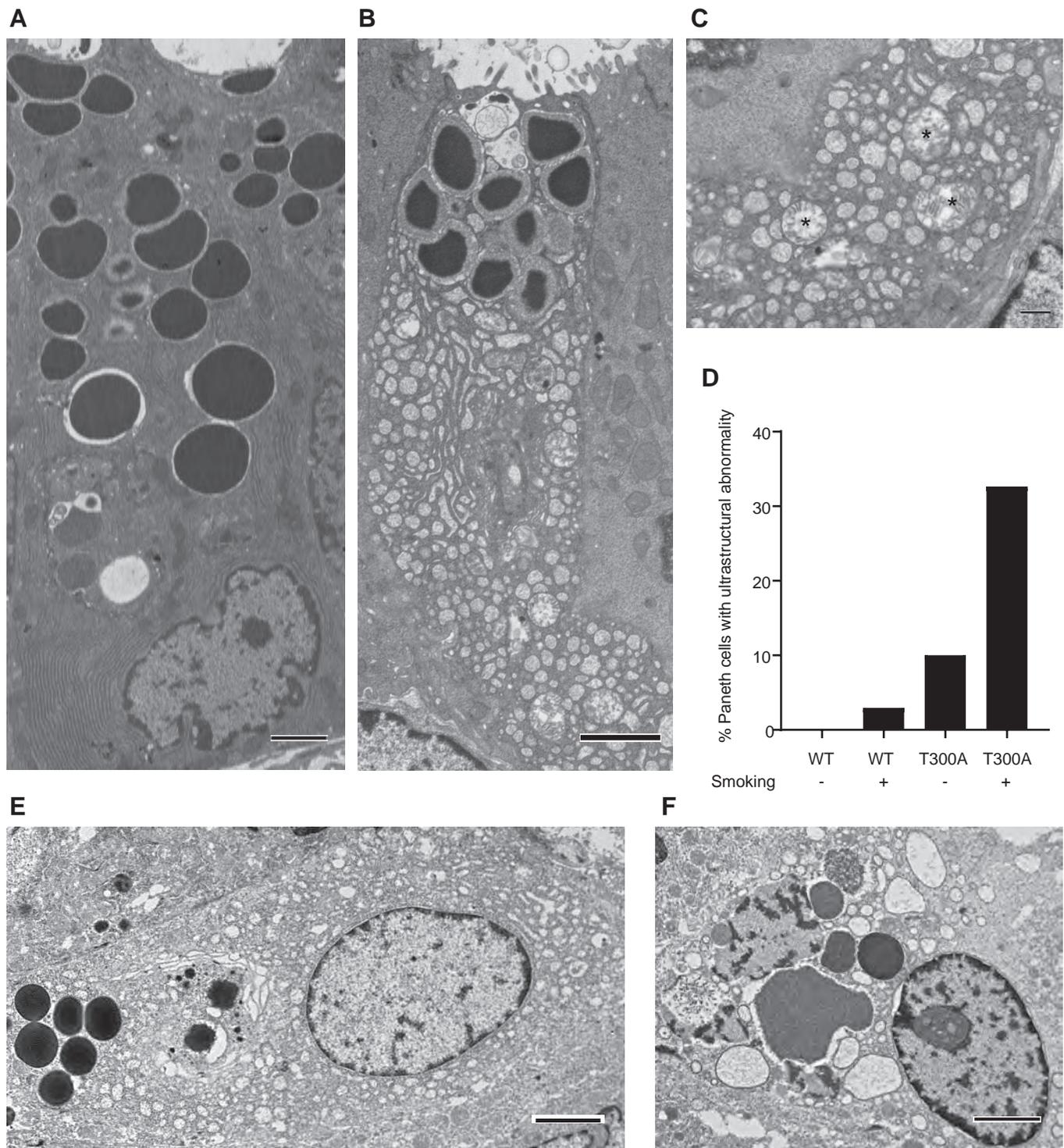
Fig. S6

Fig. S6. Transmission electron microscopy of Paneth cells in mice and CD patients. (A – D) represent findings in mice, and (E, F) represent findings in CD patients. (A) A representative normal Paneth cell with no overt ultrastructural changes. Scale bar: 2 μm. (B) A representative abnormal Paneth cell with characteristic ultrastructural changes. Scale bar: 2 μm. (C) Ultrastructural changes include cytoplasmic vesicles and degenerative mitochondria (asterisk). Scale bar: 500 nm. (D) Quantification of percentages of Paneth cells with cytoplasmic vesicles and degenerative mitochondria in each group. $Atg16l1^{T300A}$ (T300A) mice exposed to smoking contained the highest percentage of ultrastructural abnormality (Kruskal-Wallis test $P < 0.0001$). Numbers of Paneth cells analyzed: WT-nonsmoking: 32, WT-smoking: 34, T300A-nonsmoking: 30, T300A-smoking: 49. (E, F) Similar ultrastructural changes were also seen in Paneth cells from 2 CD patients who were both $ATG16L1^{T300A}$ genotype and smokers. Scale bars: 2 μm.
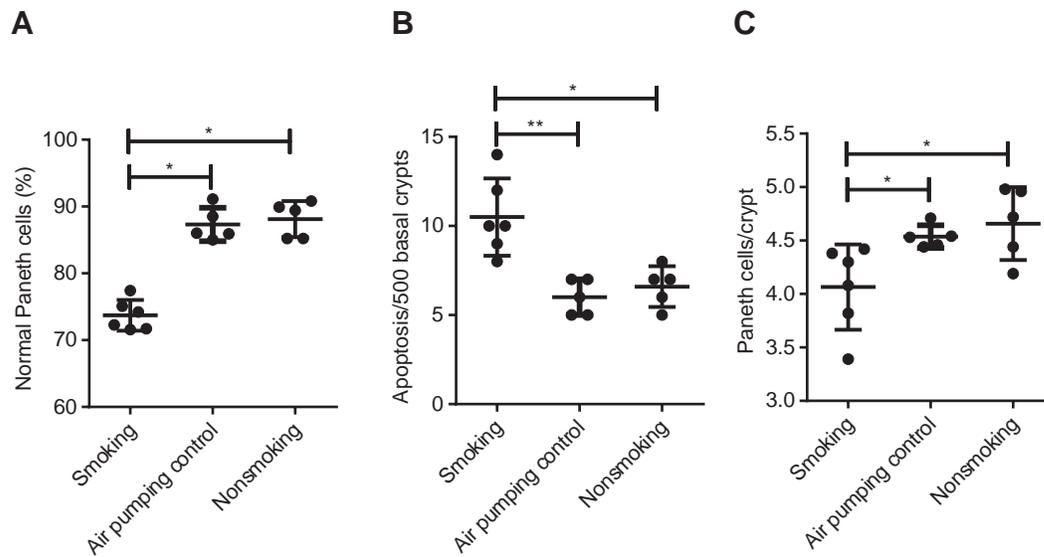
Fig. S7

Fig. S7. Physical presence in the smoking chamber with exposure to normal air did not induce Paneth cell defects in *Atg16l1*[T300A] mice. *Atg16l1*[T300A] mice were either not exposed to smoking (n=5), exposed to smoking (n=6), or placed in the smoking chamber and exposed only to normal air (i.e. no exposure to cigarette smoke) with the same schedule and duration as the mice exposed to cigarette smoke (n=5). Neither (A) percentage of normal Paneth cells, (B) crypt base apoptosis, nor (C) Paneth cells/crypt were affected in the T300A mice placed in the smoking chamber and exposed to air relative to the nonsmoking controls. $P > 0.9999$ between nonsmoking and air exposure by Kruskal-Wallis test. *: $P < 0.05$; **: $P < 0.01$. Data represent mean ± SEM.

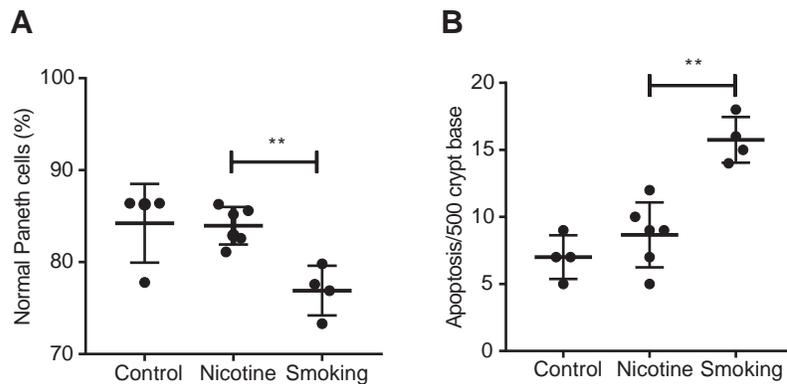**Fig. S8.** Oral nicotine administration did not induce Paneth cell defects in *Atg16l1^T300A* mice. Nicotine administration per oral did not induce (**A**) Paneth cell defects nor (**B**) crypt base apoptosis in *Atg16l1^T300A* mice ($P > 0.9999$ compared to control by one-way ANOVA in both groups). Control: n=4, nicotine: n=6, smoking: n=4. **: $P < 0.01$. Data represent mean ± SEM.
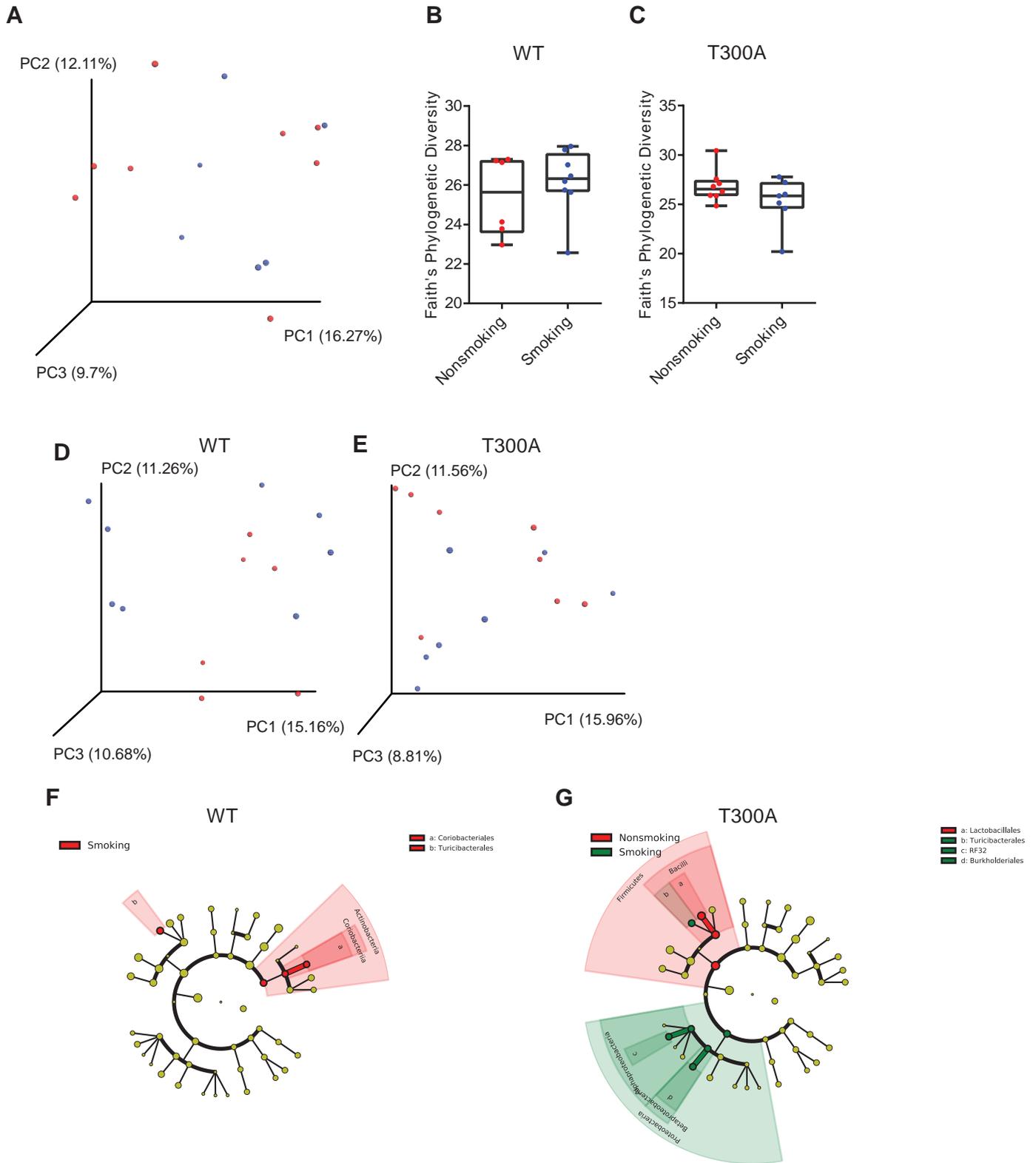
Fig. S9

A

PC2 (12.11%)

PC1 (16.27%)

PC3 (9.7%)

B

WT

Faith's Phylogenetic Diversity

Nonsmoking | Smoking

C

T300A

Faith's Phylogenetic Diversity

Nonsmoking | Smoking

D

WT

PC2 (11.26%)

PC1 (15.16%)

PC3 (10.68%)

E

T300A

PC2 (11.56%)

PC1 (15.96%)

PC3 (8.81%)

F

WT

Smoking

a: Coriobacteriales
b: Turicibacterales

G

T300A

Nonsmoking
Smoking

a: Lactobacillales
b: Turicibacterales
c: RF32
d: Burkholderiales

Fig. S9. No significant changes in microbiota compositions between *Atg16l1^T300A* (T300A) and WT littermates with and without cigarette smoking. (**A**) Beta diversity PCoA plots of nonsmoked *Atg16l1^T300A* mice and littermates ($P = 0.062$). Blue: WT; red: T300A. Alpha diversity (Faith's Phylogenetic) in the microbiota of (**B**) WT littermates ($P=0.7546$) and (**C**) T300A mice ($P=0.2810$) with and without smoking exposure. Beta diversity (Unweighted unifrac) PCoA plots of (**D**) WT littermates ($P = 0.051$) and (**E**) T300A mice ($P = 0.169$) with and without smoking exposure. Red: nonsmoked. Blue: smoked. Cladograms of microbiota changes in (**F**) WT littermates and (**G**) T300A mice after cigarette smoking. (A-F): WT-nonsmoked n=6, WT-smoked n=8, T300A-nonsmoked n=8, T300A-smoked n=7. (B, C): Data represent mean ± SEM.

A WT-WT cohoused

PC2 (12.28%)

PC1 (15.2%)

PC3 (10.07%)

B T300A-T300A cohoused

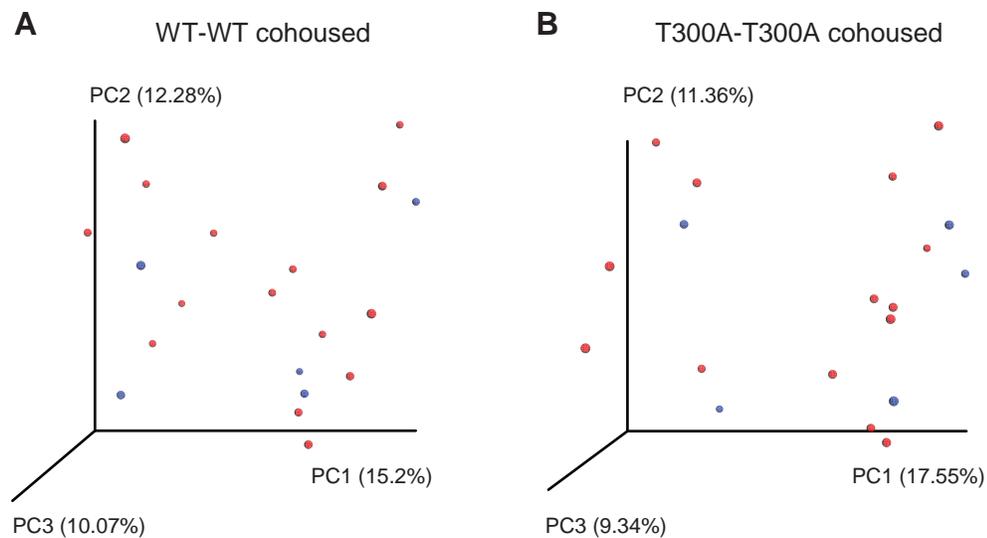PC2 (11.36%)

PC1 (17.55%)

PC3 (9.34%)

Fig. S10. Microbiota compositions in *Atg16l1^{T300A}* (T300A) and WT littermates in co-housing experiment. Beta diversity PCoA plots of (**A**) wild type microbiota donors (n=5) and recipients (n=15)($P = 0.542$) and (**B**) T300A microbiota donors (n=5) and recipients (n=17)($P = 0.382$). Red: recipients. Blue: donors.
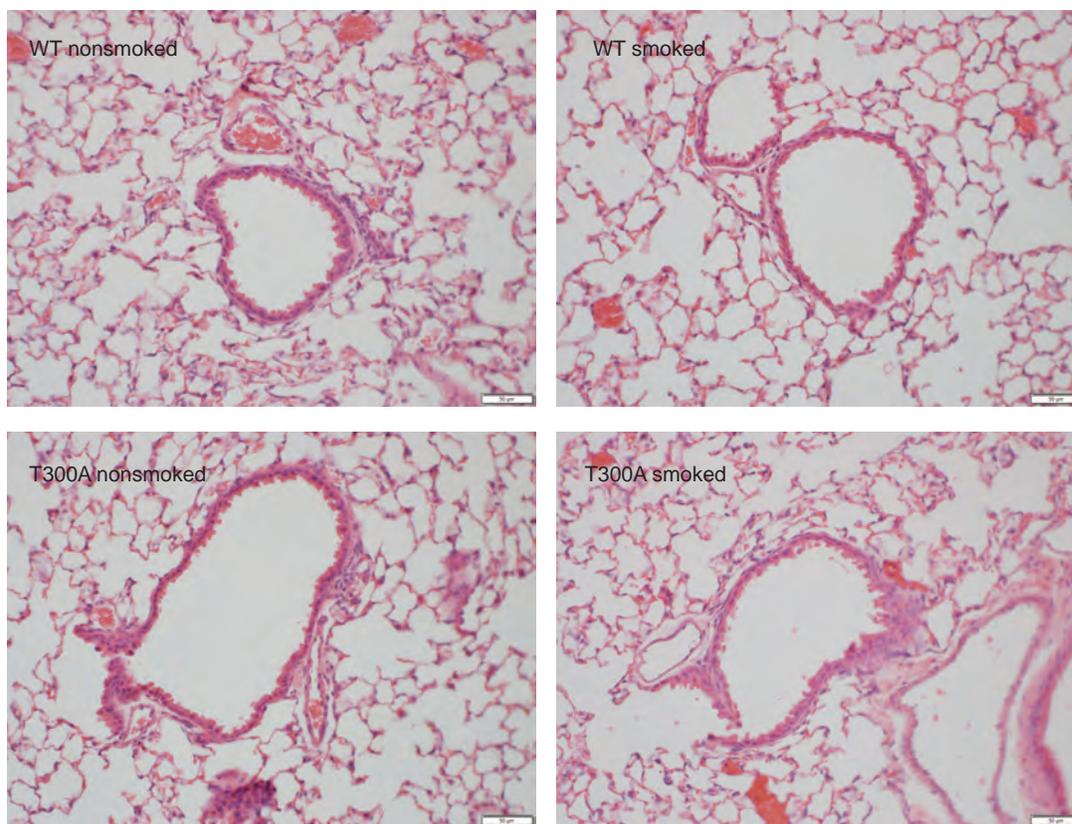
Fig. S11

Fig. S11. Lack of significant lung pathology in *Atg16l1^T300A* (T300A) and WT littermates exposed to cigarette smoking. Histology from mice after 6 weeks exposure to cigarette smoking did not show significant inflammation on routine H&E. WT-nonsmoked: n = 7, WT-smoked: n = 9, *Atg16l1^T300A*-nonsmoked: n = 7, *Atg16l1^T300A*-smoked: n = 8. Scale bars: 50 µm.
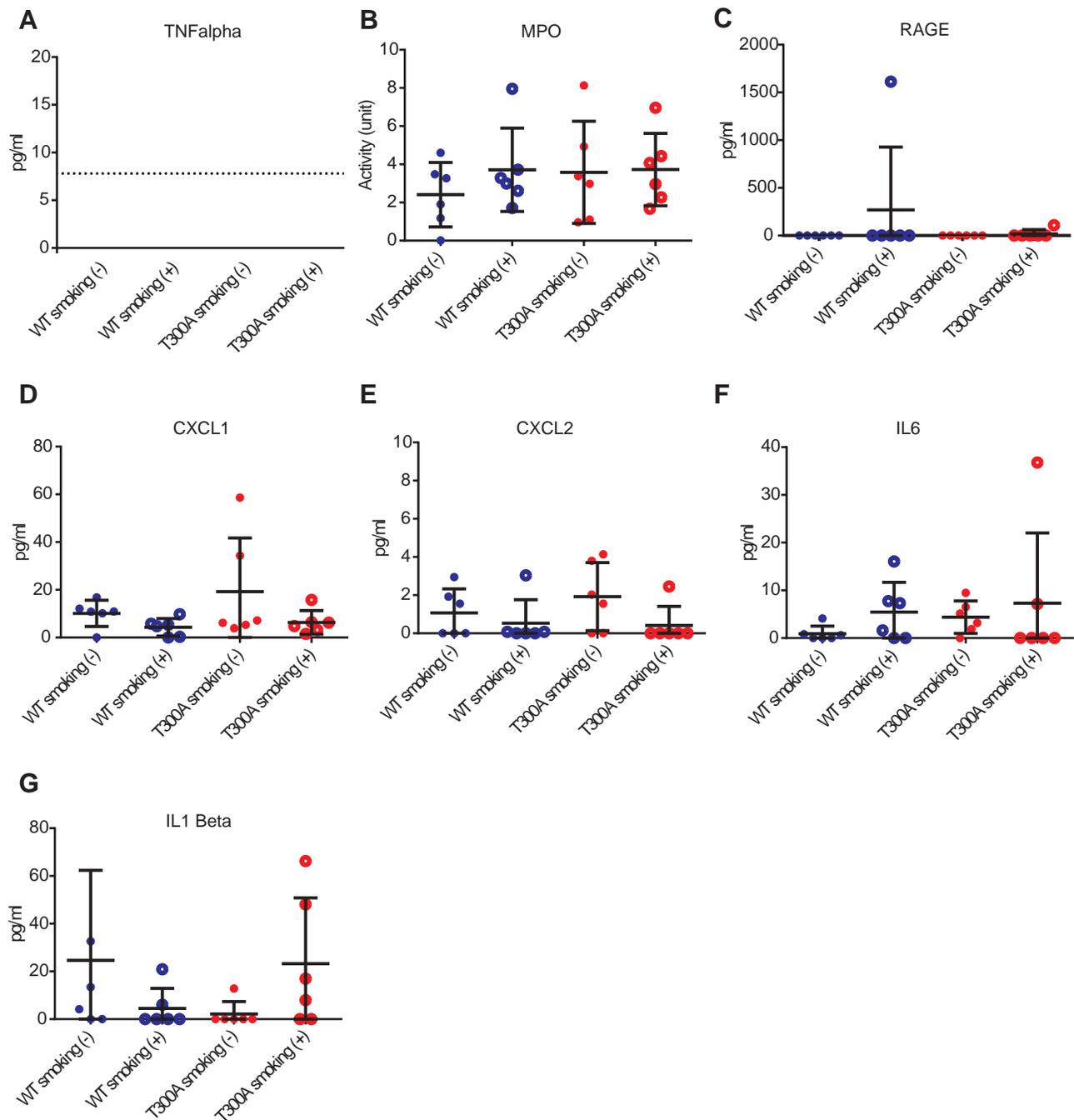
Fig. S12

Fig. S12. Lack of elevated systemic inflammatory markers in *Atg16l1^T300A* (T300A) and WT littermates exposed to cigarette smoking. Serum samples from T300A and WT littermates with and without exposure to 6-week cigarette smoking (n=6/group) were tested for (**A**) TNFα, (**B**) MPO, (**C**) RAGE, (**D**) CXCL1, (**E**) CXCL2, (**F**) IL6, and (**G**) IL1β. (B –G): *P* value between T300A-smoking (+) vs. all other groups > 0.05 (2-way ANOVA). Data represent mean ± SEM.
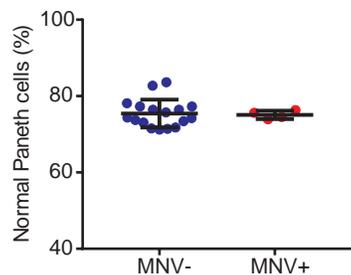
Fig. S13. MNV infection was detected in a subset of *Atg16l1*^{T300A} mice but MNV status did not correlate with percentage of normal Paneth cells. Approximately 16% of the mice shown in Fig. 2 had infection of MNV (across all four groups). Focusing on the *Atg16l1*^{T300A} smoked mice, the percentage of normal Paneth cells was not altered between the mice that were infected with MNV (n=4) and those that were not (n=17)(P = 0.9537 by Mann-Whitney test). Data represent mean ± SEM.
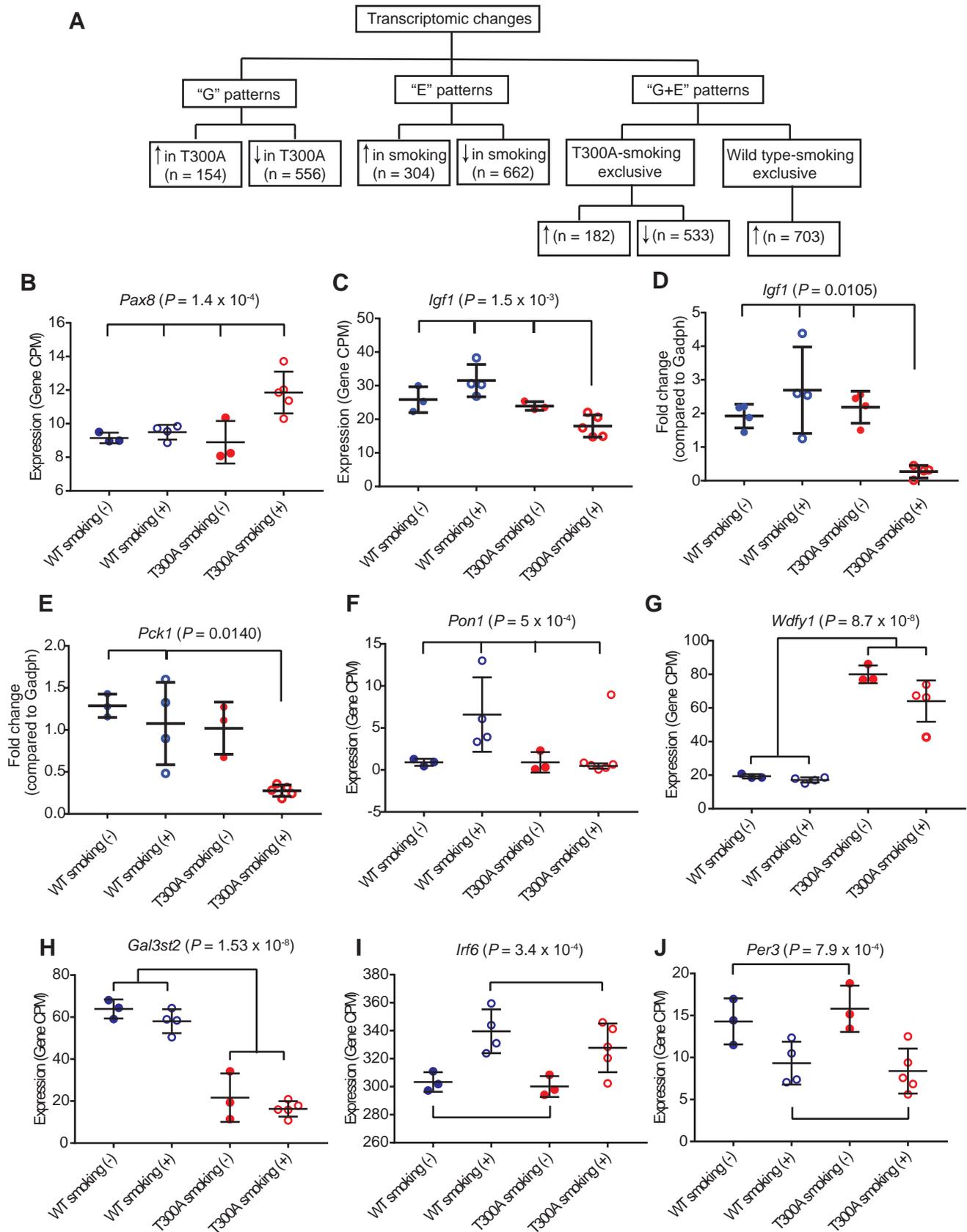
Fig. S14

Fig. S14. Combination of genetics and environmental trigger resulted in unpredicted and nonadditive transcriptomic changes. (**A**) The different categories of transcriptomic changes seen with genetics (G), environment (E), or G+E interactions. The numbers of involved genes are included. (**B-H**) Examples of different gene expression profiles. Examples of G+E specifically seen in *Atg16l1^T300A* (T300A) mice exposed to smoking include (**B**) *Pax8* and (**C**) *Igf1*. Representative RT-PCR performed on selected genes involved in the PPARα/γ pathway is shown for (**D**) *Igf1* and (**E**) *Pck1*. Examples of G+E specifically seen in WT mice exposed to smoking include (**F**) *Pon1*. Examples of G patterns include (**G**) *Wdfy1* and (**H**) *Gal3st2*. Examples of E patterns include (**I**) *Irf6* and (**J**) *Per3*. Data represent mean ± SEM.

Fig. S15

A

NR nonsmoker | NR smoker | T300A nonsmoker | T300A smoker

B

WT nonsmoked | WT smoked | T300A nonsmoked | T300A smoked

C

WT nonsmoked | WT smoked

T300A nonsmoked | T300A smoked
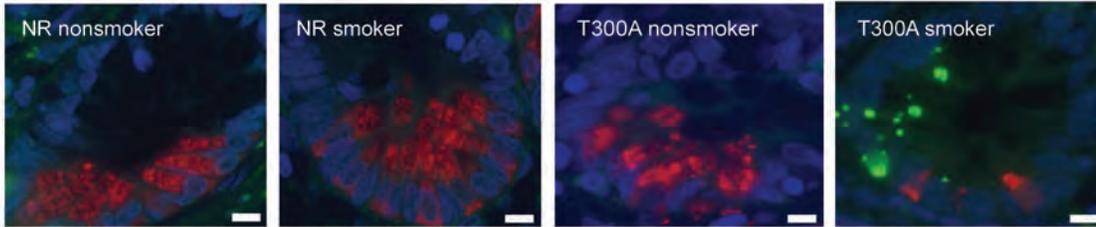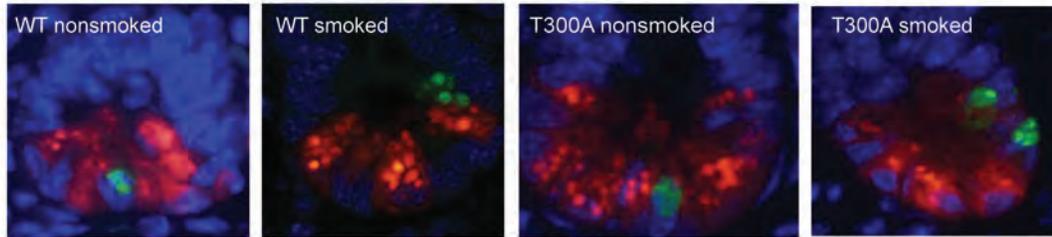
Fig. S15. Representative photomicrographs of immunofluorescence on ileum samples from CD subjects and mice. (**A**) Defensin 5 (red) and TUNEL (green) co-staining on CD samples. Blue: nuclei (Hoechst stain). (**B**) Lysozyme (red) and cleaved caspase-3 (green) co-staining on mouse samples. Blue: nuclei (Hoechst stain). (**C**) Cleaved caspase-3 (green) on villi (arrowheads) on mouse samples. Blue: nuclei (Hoechst stain).
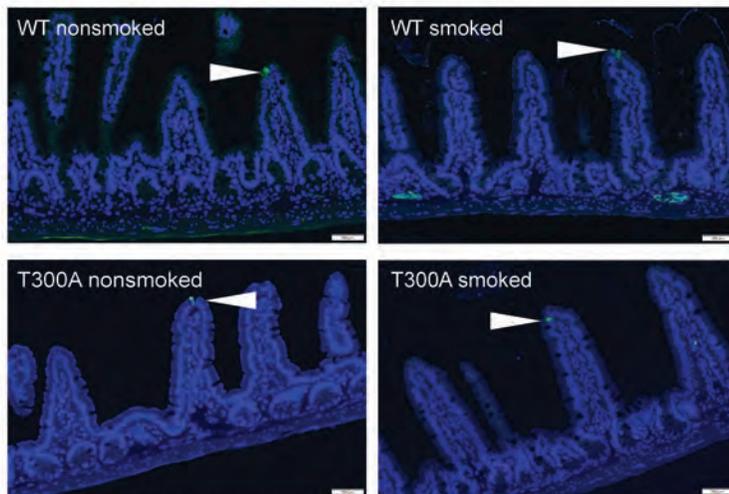
Fig. S16
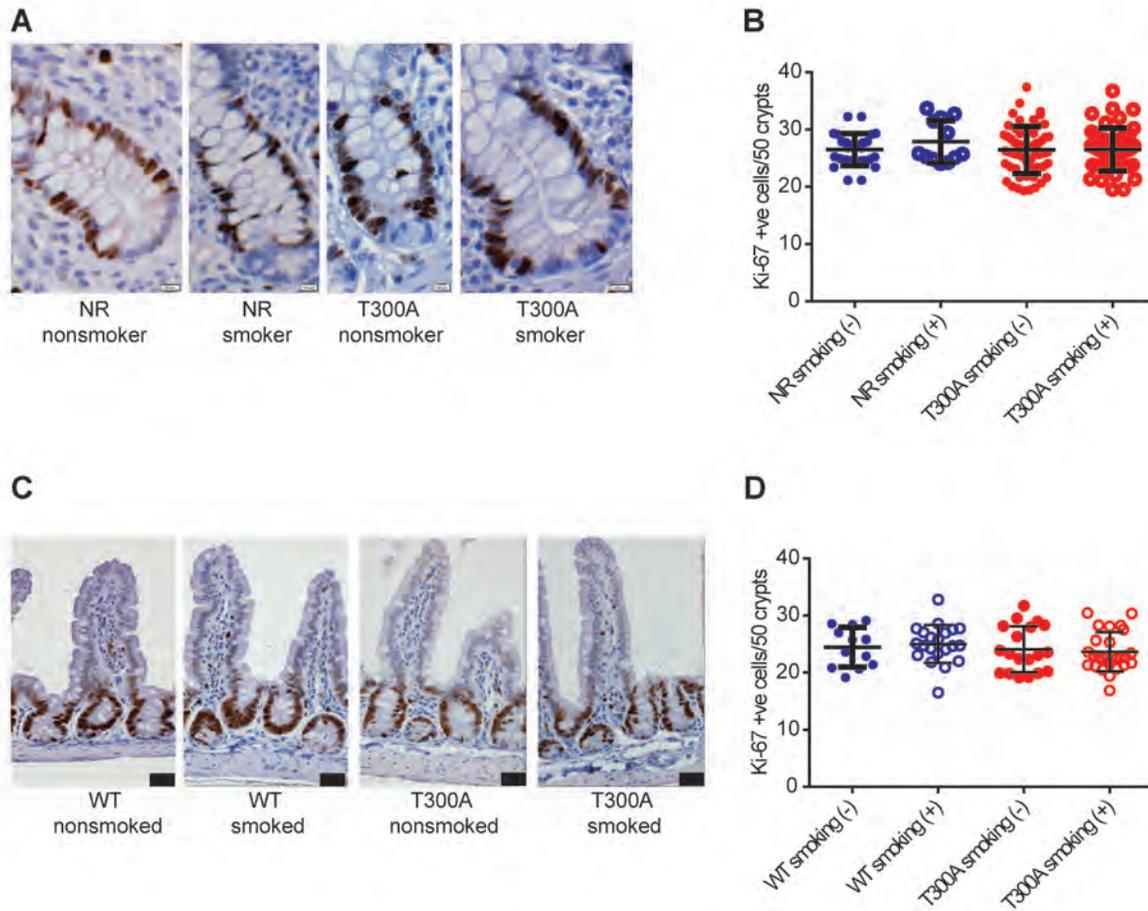
Fig. S16. Smoking-induced Paneth cell defects were not associated with changes in crypt proliferation. (**A**) Representative photomicrographs of Ki-67 immunohistochemistry on CD samples. (**B**) No significant increase in crypt proliferation ($P = 0.8036$) were seen in relation to *ATG16L1^{T300A}* (T300A) and cigarette smoking. NR nonsmokers n=27, NR smokers n=12, T300A nonsmokers n=59, T300A smokers n=47. Data were analyzed by Kruskal-Wallis test followed by Dunn's multiple comparison tests between groups. (**C**) Representative photomicrographs of Ki-67 immunohistochemistry on mouse samples. (**D**) No significant increase in crypt proliferation ($P = 0.4830$) were seen in relation to *Atg16l1^{T300A}* and cigarette smoking. WT-nonsmoked n = 12; WT-smoked n = 21; T300A-nonsmoked n = 19; T300A-smoked n = 25. Results of 6 independent experiments. Data were analyzed by Kruskal-Wallis test followed by Dunn's multiple comparison tests between groups. Data represent mean ± SEM.

Fig. S17
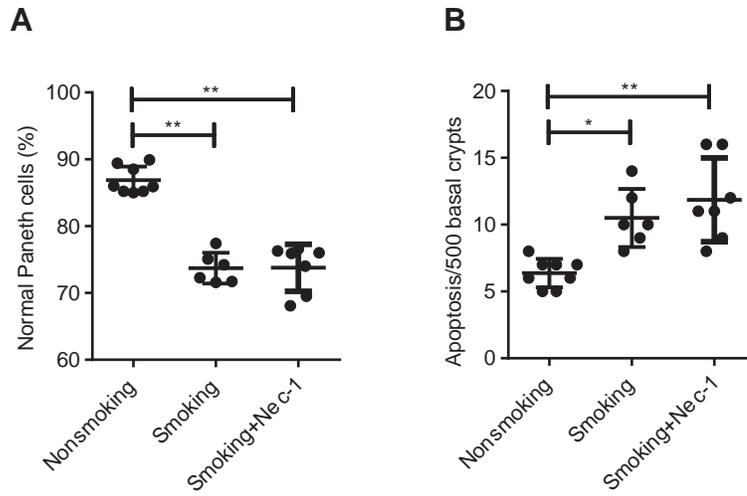
Fig. S17. Necroptosis was not associated with smoking-induced Paneth cell defects in *Atg16l1*^{T300A} mice. *Atg16l1*^{T300A} mice were either not exposed to smoking (n=8), exposed to smoking (n=6), or exposed to smoking in conjunction with Nec-1 administration (n=7). Neither (A) the percentage of normal Paneth cells nor (B) the amount of crypt base apoptosis were affected by Nec-1 administration. $P > 0.9999$ between smoking and smoking+Nec1 by Kruskal-Wallis tests for both. *: $P < 0.05$; **: $P < 0.01$. Data represent mean ± SEM.

Fig. S18

A

RNA-Seq

↓

Enrichr pathway analysis (GxE patterns)

↓

Enrichr GEO drug perturbation dataset analysis

↓

Potential drug target selection

B

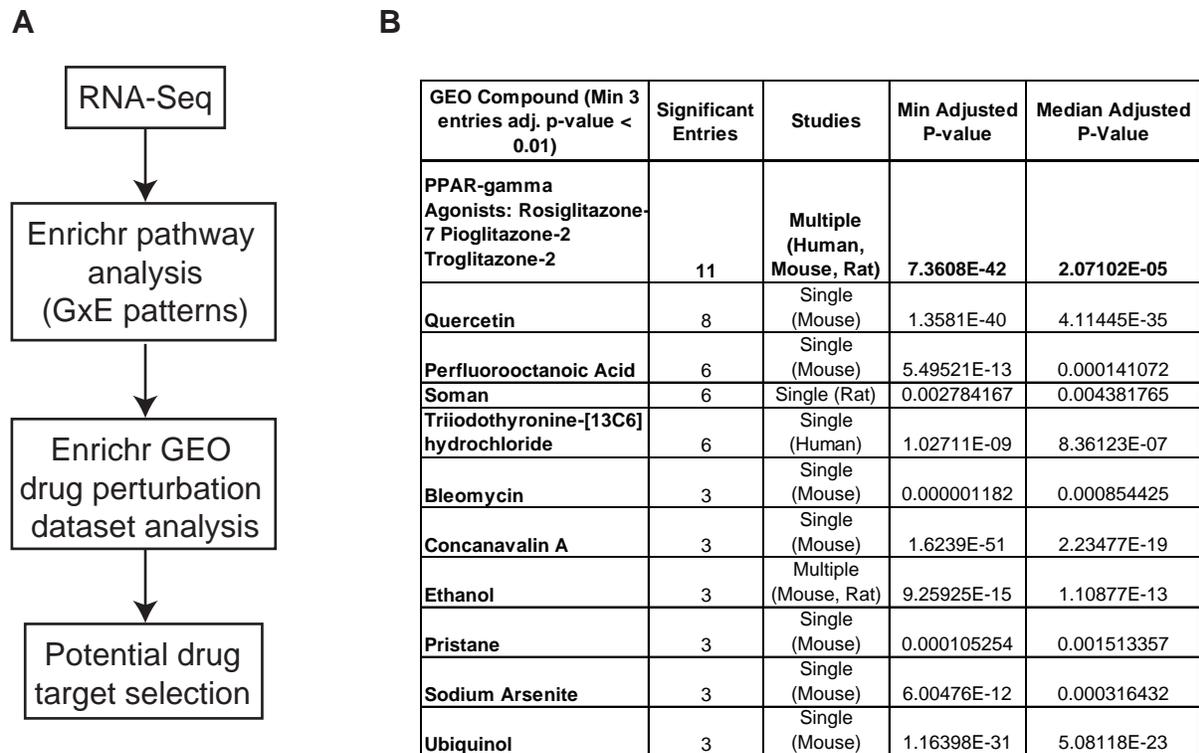| GEO Compound (Min 3 entries adj. p-value < 0.01)	Significant Entries	Studies	Min Adjusted P-value	Median Adjusted P-Value
PPAR-gamma Agonists: Rosiglitazone-7 Pioglitazone-2 Troglitazone-2	**11**	**Multiple (Human, Mouse, Rat)**	**7.3608E-42**	**2.07102E-05**
Quercetin	8	Single (Mouse)	1.3581E-40	4.11445E-35
Perfluorooctanoic Acid	6	Single (Mouse)	5.49521E-13	0.000141072
Soman	6	Single (Rat)	0.002784167	0.004381765
Triiodothyronine-[13C6] hydrochloride	6	Single (Human)	1.02711E-09	8.36123E-07
Bleomycin	3	Single (Mouse)	0.000001182	0.000854425
Concanavalin A	3	Single (Mouse)	1.6239E-51	2.23477E-19
Ethanol	3	Multiple (Mouse, Rat)	9.25925E-15	1.10877E-13
Pristane	3	Single (Mouse)	0.000105254	0.001513357
Sodium Arsenite	3	Single (Mouse)	6.00476E-12	0.000316432
Ubiquinol	3	Single (Mouse)	1.16398E-31	5.08118E-23

Fig. S18. Selection process for PPARγ agonist rosiglitazone testing**.** (A) Diagram describing the work flow from the full thickness transcriptomic dataset, *Ppara/g* regulation was identified using Enrichr pathway analysis as a major G+E pattern correlated with *Atg16l1^{T300A}*-smoking. This was followed by analyzing the subsets of WT-smoking vs. *Atg16l1^{T300A}*-smoking groups using the Enrichr GEO drug perturbation dataset analysis to identify potential compounds that could be used as therapeutic candidate for the phenotype. (B) The Enrichr drug perturbation dataset analysis identified PPARγ agonists as the strongest candidate molecule class. Detailed *PParg* associated genes are described in Table S9.

Fig. S19

Fig. S19. Mice with intestinal epithelium-specific knockout of *Pparg* (*Pparg/Villin-Cre*; n=9) showed (A) reduced percentages of normal Paneth cells; (B) reduced Paneth cells/crypt; and (C) increased crypt base apoptosis compared to *fl/fl* littermate control mice (n=8). Mann-Whitney tests were performed for all. *: $P < 0.05$, **: $P < 0.01$, ****: $P < 0.0001$. (D) Distribution of Paneth cell morphology subclasses in *Pparg/Villin-Cre* mice was similar to that of the *Atg16l1*[T300A]-smoking group. Data represent mean ± SEM.
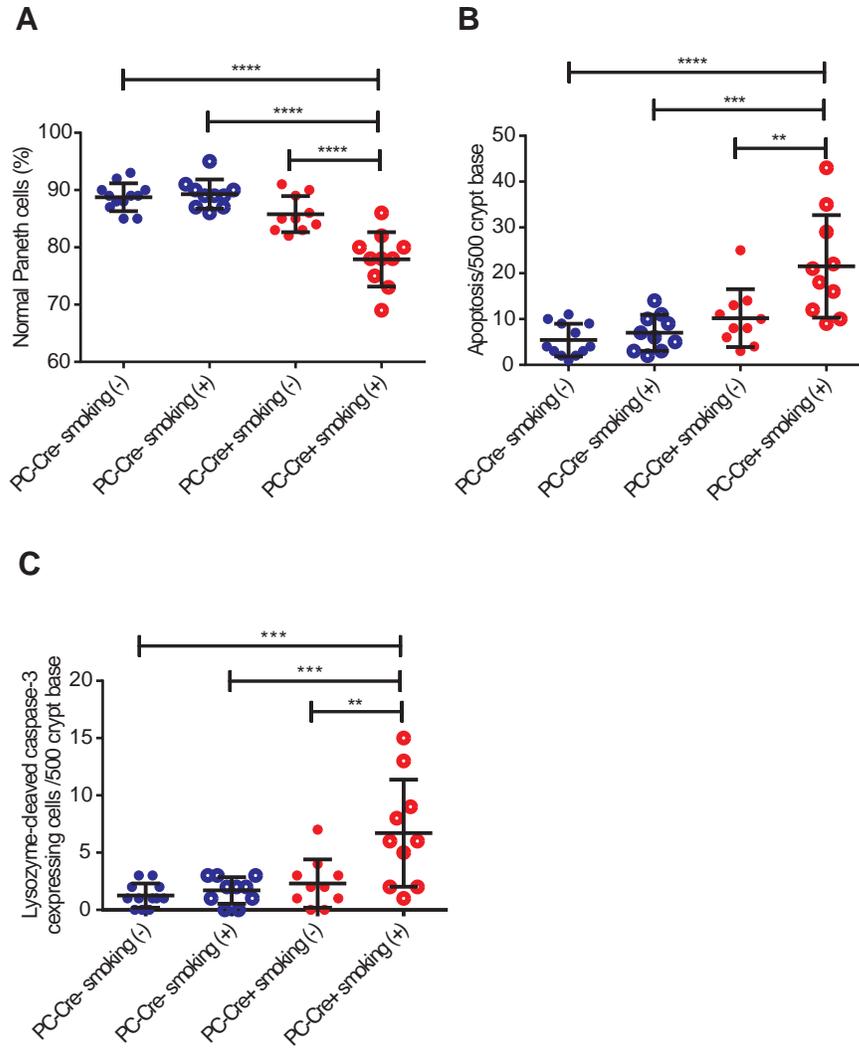
Fig. S20

Fig. S20. Paneth cell-specific *Atg16l1* whole gene knock out mice showed Paneth cell defects upon exposure to cigarette smoking. Mice with Paneth cell-specific *Atg16l1* deletion (PC-Cre+ mice) exposed to cigarette smoking showed more profound (**A**) Paneth cell defect ($P < 0.0001$). Cigarette smoking induced increased (**B**) crypt apoptosis ($P = 0.0035$) and (**C**) Paneth cell apoptosis ($P = 0.0031$) in PC-Cre+ mice compared to PC-Cre- mice. PC-Cre- non-smoked: n=12, PC-Cre- smoked: n=10, PC-Cre+ non-smoked: n=10, PC-Cre+ smoked: n=10. Statistical analysis for all panels was performed using 2-way ANOVA. **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$. Data represent mean ± SEM.
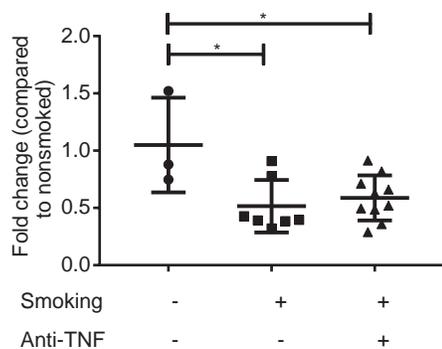
Fig. S21. Anti-TNFα did not restore *Igf1* gene expression level reduced by cigarette smoking in *Atg16l1*^{T300A} mice. *Igf1* gene expression status in *Atg16l1*^{T300A} mice that were smoked and treated with or without anti-TNFα was reduced compared to the nonsmoked *Atg16l1*^{T300A}. Nonsmoked n=3, smoked n=7, smoked + anti-TNFα n=10. *P* = 0.0154 by one-way ANOVA. Data represent mean ± SEM.

Table S1. Demographic data for the 186 CD subjects included in the analyses.

Characteristic	n (%)
Male sex	93 (50)
Caucasian	160 (86)
Never smoker	109 (59)
Age at diagnosis	
A1 ($\leq$16 yrs)	32 (17)
A2 (17-40 yrs)	127 (68)
A3 (>40 yrs)	27 (15)
Disease location	
L1 (Ileal)	74 (40)
L2 (Colonic)	10 (5)
L3 (Ileocolonic)	102 (55)
L4 (upper small bowel)	0 (0)
Disease behavior	
B1 (Non-stricturing, non-penetrating)	19 (10)
B2 (Stricturing)	104 (56)
B3 (Penetrating)	59 (32)
Presence of perianal disease	16 (9)
Treatment received at time of surgery	
5-aminosalicylic acid (ASA)	35/162 (22)
Steroids	62/162 (38)
Immunomodulators	66/162 (41)
Biologics	76/162 (47)
Average age at surgery (years)(range)	37.1 (10-82)
Prophylactic therapy after surgery	
Immunomodulators	83/126 (66)
Biologics	66/126 (52)

Table S2. Adjusted P values between each CD subject group in this study.

	Normal Paneth cell percentage	TUNEL-positive cells/500 crypt base	HD5-TUNEL coexpressing cells/500 crypt base	Paneth cells/crypt	Ki-67 expressing cells/crypt (50 crypts)
Figure or supplementary figure	Fig. 1C	Fig. 5A	Fig. 5B	Fig. 5C	Fig. S16B
NR smoking (-) vs NR smoking (+)	>0.9999	>0.9999	>0.9999	>0.9999	>0.9999
T300A smoking (-) vs T300A smoking (+)	0.0436	<0.0001	0.0073	0.0103	>0.9999
NR smoking (-) vs T300A smoking (-)	>0.9999	0.1349	>0.9999	0.8357	>0.9999
NR smoking (+) vs T300A smoking (+)	0.0027	0.0018	0.6328	>0.9999	>0.9999
NR smoking (-) vs T300A smoking (+)	0.0261	<0.0001	0.0100	>0.9999	>0.9999

Table S3. Adjusted P values between each group in *Atg16l1*[T300A] mouse experiments in this study.

	Normal Paneth cell percentage	Cleaved caspase-3 positive cells/500 crypt base	Cleaved caspase-3-lysozyme coexpressing cells/500 crypt base	Paneth cells/crypt	Cleaved caspase-3 positive cells /100 villi	Ki67 expressing cells/crypt (50 crypts)
Figure or supplementary figure	Fig. 2B	Fig. 5D	Fig. 5E	Fig. 5F	Fig. 5G	Fig. S16D
WT smoking (-) vs T300A smoking (-)	0.2057	>0.9999	>0.9999	>0.9999	0.7642	>0.9999
WT smoking (-) vs WT smoking (+)	>0.9999	> 0.9999	> 0.9999	>0.9999	>0.9999	>0.9999
WT smoking (-) vs T300A smoking (+)	< 0.0001	< 0.0001	0.0098	>0.9999	>0.9999	>0.9999
WT smoking (+) vs T300A smoking (-)	>0.9999	>0.9999	>0.9999	0.2809	>0.9999	>0.9999
T300A smoking (-) vs T300A smoking (+)	< 0.0001	<0.0001	0.0018	>0.9999	>0.9999	>0.9999
WT smoking (+) vs T300A smoking (+)	< 0.0001	<0.0001	0.0015	0.1850	>0.9999	0.7932