**Supplementary Methods**

*Genotyping of the Milwaukee cohort*

DNA samples were genotyped for 196,524 markers using the Human Immuno_BeadChip_1149691 (Illumina Inc., San Diego, CA, USA) according to the manufacture's protocol. Briefly, 200ng of DNA (4uL at 50ng/uL) was independently amplified, labeled, and hybridized to BeadChip microarrays then scanned with default settings using the Illumina iScan. Analysis was performed using Illumina's GenomeStudio Genotyping Module software v.2011. Genotype calls were initially generated using the Illumina-provided genotype cluster definitions file (ImmunoChip_Gentrain_June2010, generated using HapMap project DNA samples) with a Gencall cutoff of 0.15. This was followed by manual inspection of approximately 5,000 low call SNPs and SNPs with AB frequency greater than 0.55. Genotype calls for six specific SNPs were examined for correlation with Paneth cell morphology: *ATG16L1* SNPs rs12994997 and rs2241880 (T300A); and *NOD2* SNPs rs2066844 (R702W), rs2066845 (G908R), rs5743289, and rs5743293 (L1007x, SNP13).

*cDNA library construction*

Total RNA was isolated from ileal biopsy tissue using Qiagen RNeasy Minikit, according to the kit protocol. Total RNA was quantified using the Quant-iT™ RiboGreen® RNA Assay Kit and normalized to 4ng/ul. An aliquot of 200ng for each sample was transferred into library preparation, which was an automated variant of the Illumina TruSeq™ mRNA Sample

Preparation Kit. This method preserves strand orientation of the RNA transcript. It uses oligo dT beads to select mRNA from the total RNA sample. It is followed by heat fragmentation and cDNA synthesis from the RNA template. The resultant cDNA then goes through library preparation (end repair, base 'A' addition, adapter ligation, and enrichment) using Broad Institute designed indexed adapters substituted in for multiplexing. After enrichment the libraries were quantified with qPCR using the KAPA Library Quantification Kit for Illumina Sequencing Platforms and then pooled equimolarly. The entire process is in 96-well format and all pipetting is done by either Agilent Bravo or Hamilton Starlet.

*Illumina Sequencing*

Pooled libraries were normalized to 2nM and denatured using 0.2 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using either the HiSeq 2000 v3 or HiSeq 2500. Each run was a 76bp paired-end with an eight-base index barcode read. Data was analyzed using the Broad Picard Pipeline, which includes de-multiplexing and data aggregation.

**Supplementary Table 1. Mitochondrial oxidative phosphorylation gene cluster that is associated with Paneth cell phenotype.**

| Gene | Description |
|---|---|
| *UQCRHL* | ubiquinol-cytochrome c reductase hinge protein-like |
| *NDUFB3* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa |
| *NDUFS5* | NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase) |
| *NDUFB1* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa |
| *NDUFA1* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa |
| *COA3* | cytochrome c oxidase assembly factor 3 |
| *NDUFS6* | NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase) |
| *ATP5G1* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C1 (subunit 9) |
| *NDUFA9* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa |
| *COX5A* | cytochrome c oxidase subunit Va |
| *COX6A1* | cytochrome c oxidase subunit VIa polypeptide 1 |
| *UQCRH* | ubiquinol-cytochrome c reductase hinge protein |
| *UQCRQ* | ubiquinol-cytochrome c reductase, complex III subunit VII, 9.5kDa |
| *UQCR10* | ubiquinol-cytochrome c reductase, complex III subunit X |
| *ATP5I* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit E |
| *NDUFAF2* | NADH dehydrogenase (ubiquinone) complex I, assembly factor 2 |
| *ATP5G3* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C3 (subunit 9) |
| *ATP5J2* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F2 |
| *COX7B* | cytochrome c oxidase subunit VIIb |
| *ATP5H* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit d |
| *COX6B1* | cytochrome c oxidase subunit VIb polypeptide 1 (ubiquitous) |
| *NDUFA8* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa |
| *NDUFB9* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa |
| *UQCRFS1* | ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1 |
| *ATP5J* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F6 |
| *NDUFB7* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa |
| *COQ5* | coenzyme Q5 homolog, methyltransferase (S. cerevisiae) |
| *COX7A2* | cytochrome c oxidase subunit VIIa polypeptide 2 (liver) |
| *NDUFB2* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa |
| *NDUFV3* | NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa |
| *COX7C* | cytochrome c oxidase subunit VIIc |
| *COX5B* | cytochrome c oxidase subunit Vb |
| *COX6C* | cytochrome c oxidase subunit VIc |

| | |
|---|---|
| *NDUFA12* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 12 |
| *NDUFB10* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa |
| *COX8A* | cytochrome c oxidase subunit VIIIA (ubiquitous) |
| *ATP5O* | ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit |
| *COX4I1* | cytochrome c oxidase subunit IV isoform 1 |
| *NDUFS4* | NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase) |
| *ATP5L* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit G |
| *UQCRC1* | ubiquinol-cytochrome c reductase core protein I |
| *NDUFS3* | NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase) |
| *ATP5F1* | ATP synthase, H+ transporting, mitochondrial Fo complex, subunit B1 |
| *NDUFA2* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa |
| *COX7A1* | cytochrome c oxidase subunit VIIa polypeptide 1 (muscle) |
| *NDUFB5* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa |
| *ATP5E* | ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit |
| *ATP5C1* | ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1 |
| *NDUFB11* | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 11, 17.3kDa |
| *ATP5B* | ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide |

**Supplementary Table 2. Paneth cell gene cluster that is associated with Paneth cell phenotype.**

| Gene name | Description |
|---|---|
| *MSI1* | musashi RNA-binding protein 1 |
| *PLA2G2A* | phospholipase A2, group IIA (platelets, synovial fluid) |
| *CARD16* | caspase recruitment domain family, member 16 |
| *REG3A* | regenerating islet-derived 3 alpha |
| *DEFA6* | defensin, alpha 6, Paneth cell-specific |
| *EMP2* | epithelial membrane protein 2 |
| *EPHB2* | EPH receptor B2 |
| *DEFA5* | defensin, alpha 5, Paneth cell-specific |
| *LYZ* | lysozyme |
| *SPINK1* | serine peptidase inhibitor, Kazal type 1 |
| *LCN2* | lipocalin 2 |
| *REG1A* | regenerating islet-derived 1 alpha |
| *PIGR* | polymeric immunoglobulin receptor |

**Supplementary figure legends**

**Supplementary Figure 1**. Detailed Paneth cell morphologic classifications of 4 independent Crohn's disease (CD) cohorts. (A) Saint Louis adult CD (n=170). (B) Los Angeles adult CD (n=361). (C) Saint Louis pediatric CD (n=73). (D) Milwaukee pediatric CD (n=44).

**Supplementary Figure 2**. *ATG16L1* T300A and *NOD2* risk alleles did not correlate with Paneth cell phenotype in pediatric Crohn's disease patients (n=44). (A) No significant difference was seen between the numbers of *ATG16L1* T300A risk alleles and the actual percentage of normal Paneth cells. P = 0.4284 by T test. (B) No significant difference was seen between the numbers of *NOD2* risk alleles and the actual percentage of normal Paneth cells. P = 0.8982 by ANOVA. (C) No significant difference was seen between the total sum numbers of *ATG16L1* T300A and *NOD2* risk alleles and the actual percentage of normal Paneth cells. P = 0.0888 by ANOVA.

**Supplementary Figure 3**. Different microbial compositions between Crohn's disease (CD; n=36) and non-inflammatory bowel disease (IBD; n=47) patients. (A) Differential feature analysis for CD patients versus non-IBD patients by LEfSe. Red bars represent taxa with a significantly higher relative abundance in CD patients. Blue bars represent taxa with a significantly higher relative abundance in non-IBD patients. (B) Stacked bar plots of phylum-level compositions of mucosal microbiome between CD and non-IBD patients. Each bar represents one patient. (C) Cladogram of differential taxa between CD and non-IBD patients analyzed by LEfSe.

**Supplementary Figure 4.** The mucosal microbiome of pediatric Crohn's disease (CD) patients stratified by Paneth cell phenotype. (A) Unweighted beta-diversity of microbiome between CD (n=36) and non-inflammatory bowel disease (IBD) patients (n=47) by principal coordinate analysis (P = 0.137). Red: CD patients with Type I Paneth cell phenotype. Blue: CD patients with Type II Paneth cell phenotype. (B) Stacked bar plots of phylum-level compositions of mucosal microbiome between Type I and Type II Paneth cell phenotypes. Each bar represents one patient.

**Supplementary Figure 5**. Correlation of microbiota alpha diversity and degree of Paneth cell defect by Pearson's correlation test. (A) Shannon index in Crohn's disease patients (n=36). (B) Shannon index in non-inflammatory bowel disease patients (n=47). (C) Faith's Phylogenetic index in Crohn's disease patients (n=36). (D) Faith's Phylogenetic index in non-inflammatory bowel disease patients (n=47).

**Supplementary Figure 6.** Comparison of beta-diversity metrics and taxonomic differences between Type I and II Paneth cell phenotypes in non-inflammatory bowel disease (IBD) patients (n=47). (A, B) Unweighted and weighted beta-diversity comparison within and between non-IBD patients with Type I and II Paneth cell phenotypes. *: $P < 0.05$; ****: $P < 0.0001$ by one-way ANOVA. (C) The cladogram of mucosal microbiome in non-IBD patients stratified by Paneth cell phenotype.

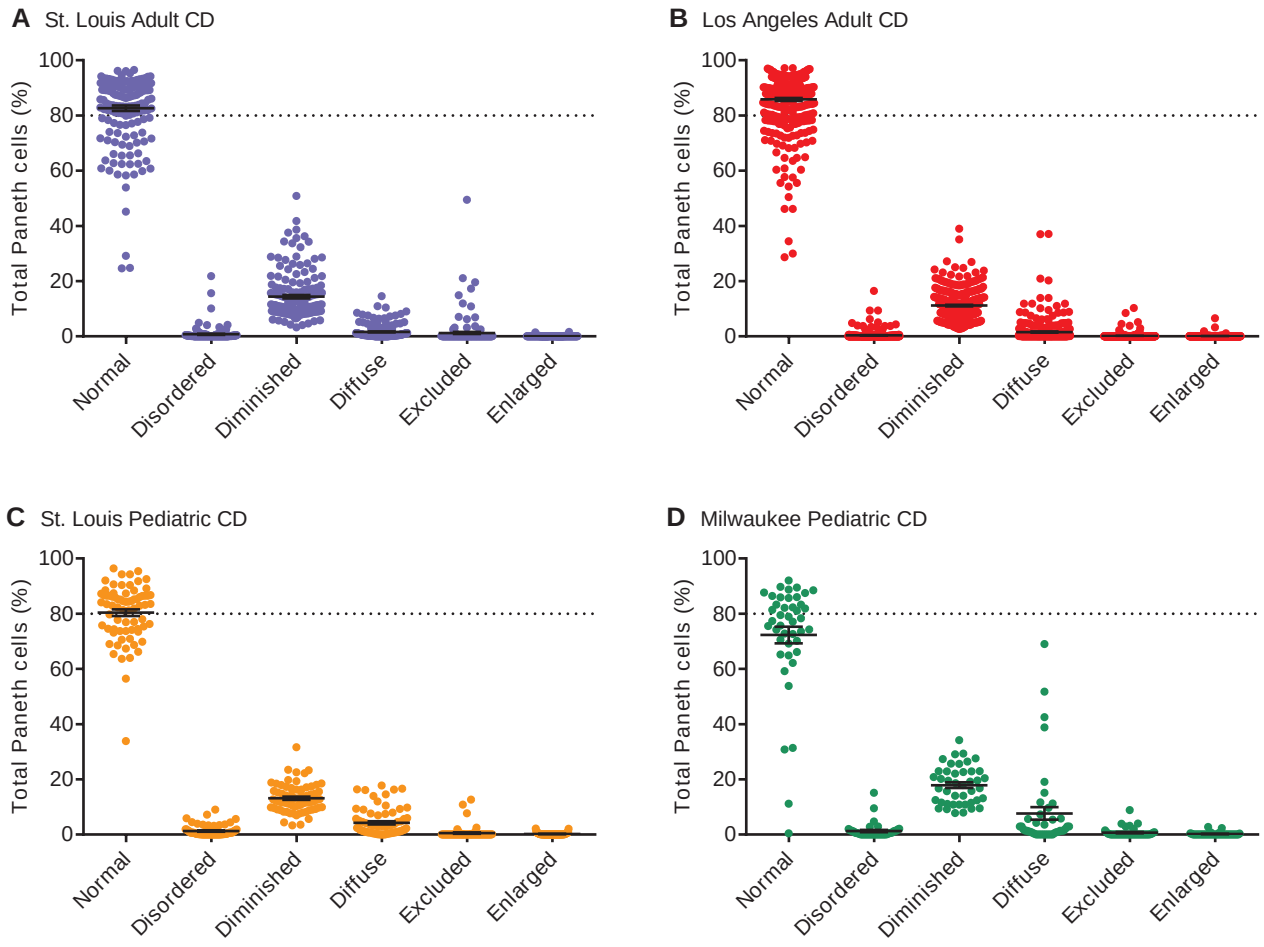
**Supplementary Figure 7.** Paneth cell phenotype correlates with different microbial taxa in Crohn's disease (CD) patients (n=36). (A) Differential feature analysis for CD patients with Type

I versus Type II Paneth cell phenotypes by LEfSe. Red bars represent taxa with a significantly higher relative abundance in CD patients with Type I Paneth cell phenotype. Blue bars represent taxa with a significantly higher relative abundance in CD patients with Type II Paneth cell phenotype. (B) Cladogram of differential taxa between CD patients with Type I and Type II Paneth cell phenotype analyzed by LEfSe.

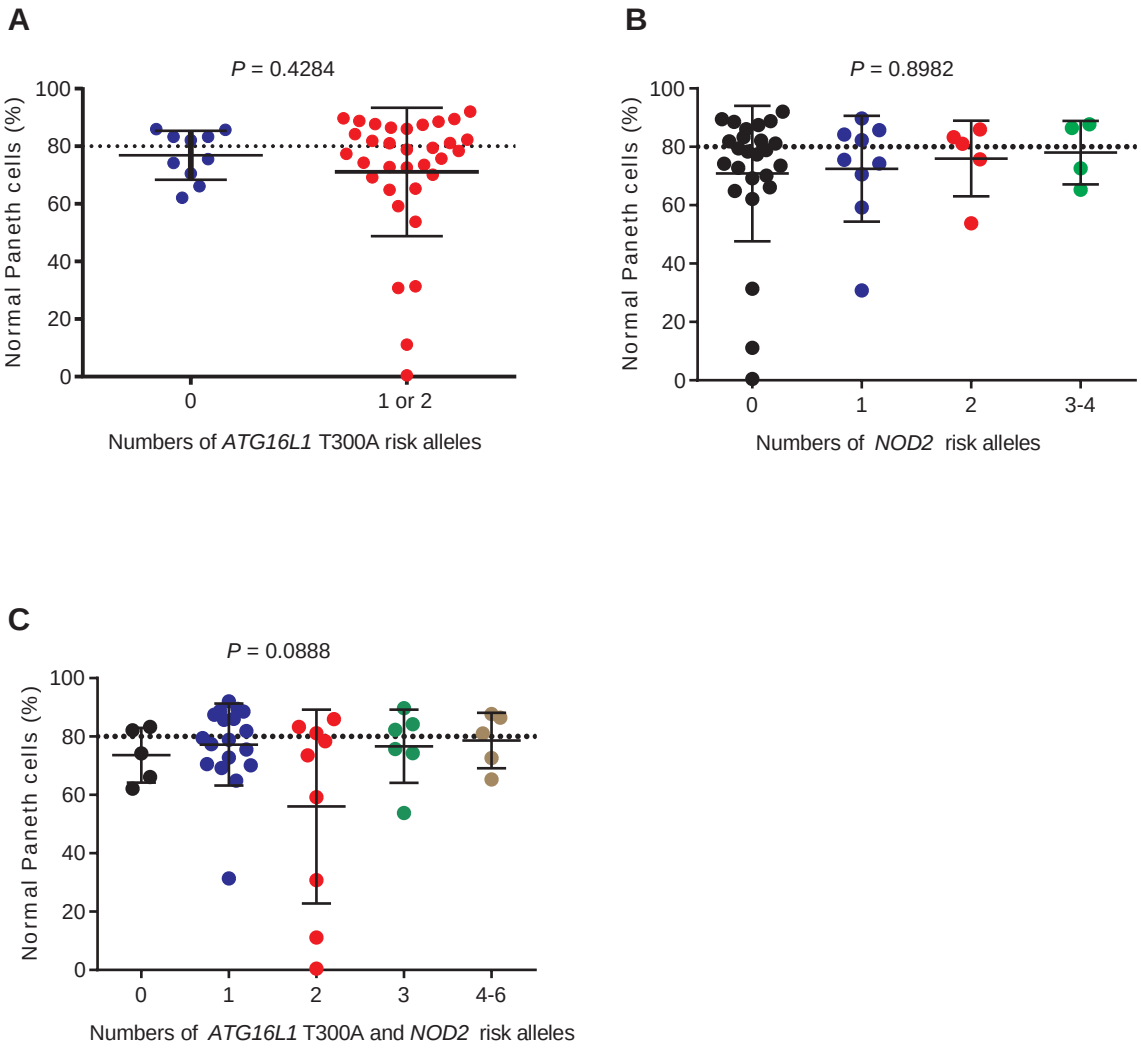
**Supplementary Figure 8.** Transcriptome profiles between Type I and II Paneth cell phenotypes in pediatric CD patients (n=38 with sufficient RNA). Principle Coordinate Analysis (PCA) of CD patients with either Type I (Red) or Type II (Blue) Paneth cell phenotypes based on gene expression data.

**Supplemental Figure 9.** Expression of selected Paneth cell-specific genes is associated with Paneth cell phenotype in Crohn's disease (CD) patients (n=38 with sufficient RNA). (A) *DEFA6*; (B) *PLA2G2A*; (C) *REG3A*. (D) Selected bacterial taxa that are significantly more abundant in CD patients with high expression level of Paneth cell genes identified by Differential Feature analysis (LEfSe). **: $P < 0.01$ by one-way ANOVA.

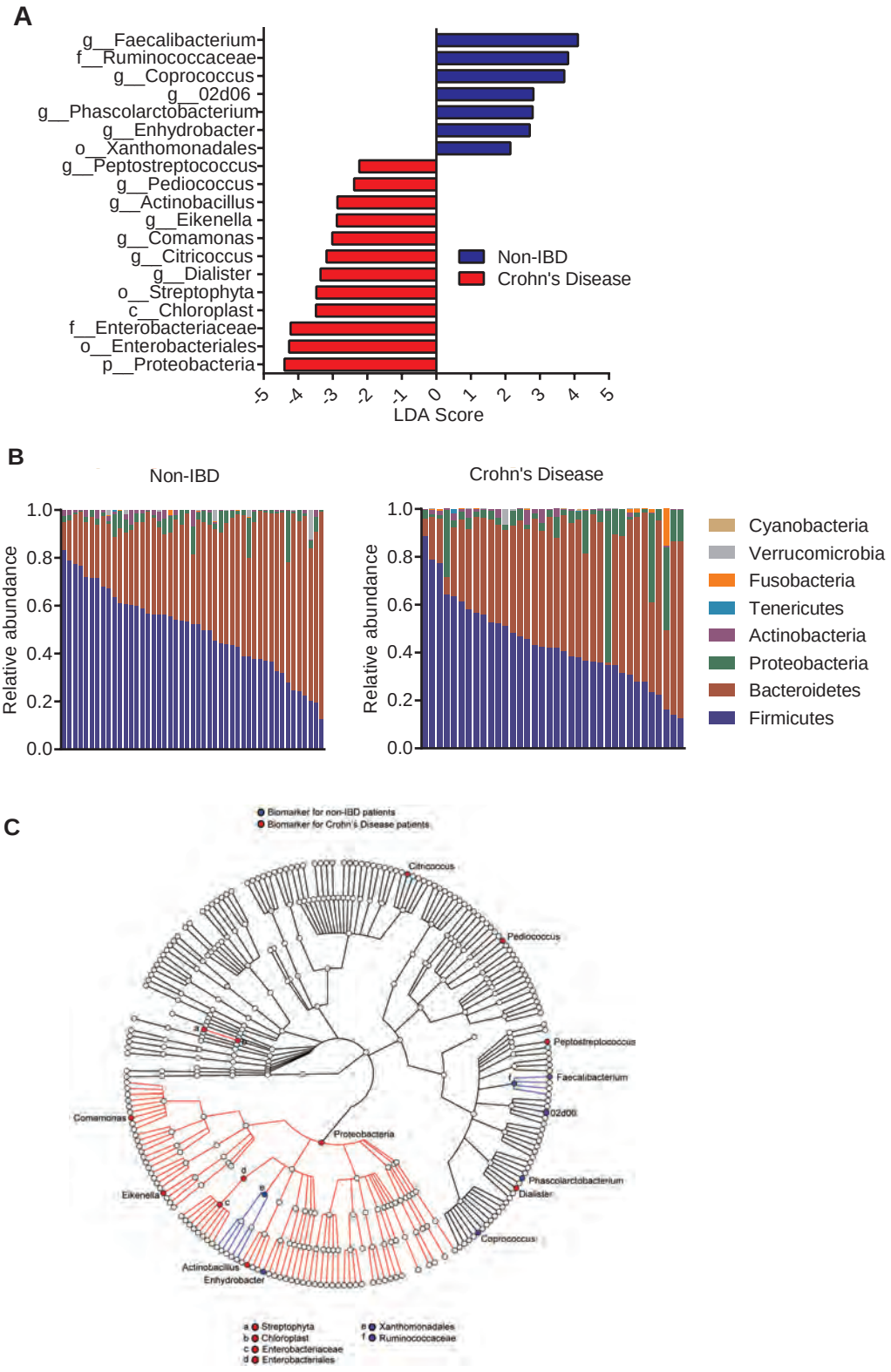
**Supplementary Figure 1**



## Supplementary Figure 2

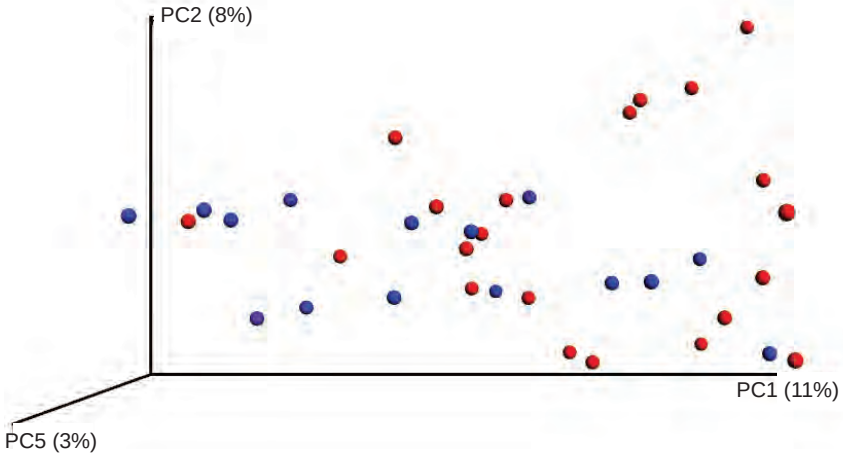


## Supplementary Figure 3

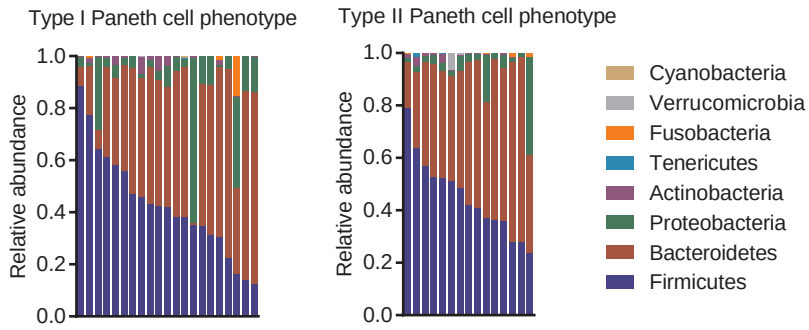


## Supplementary Figure 4

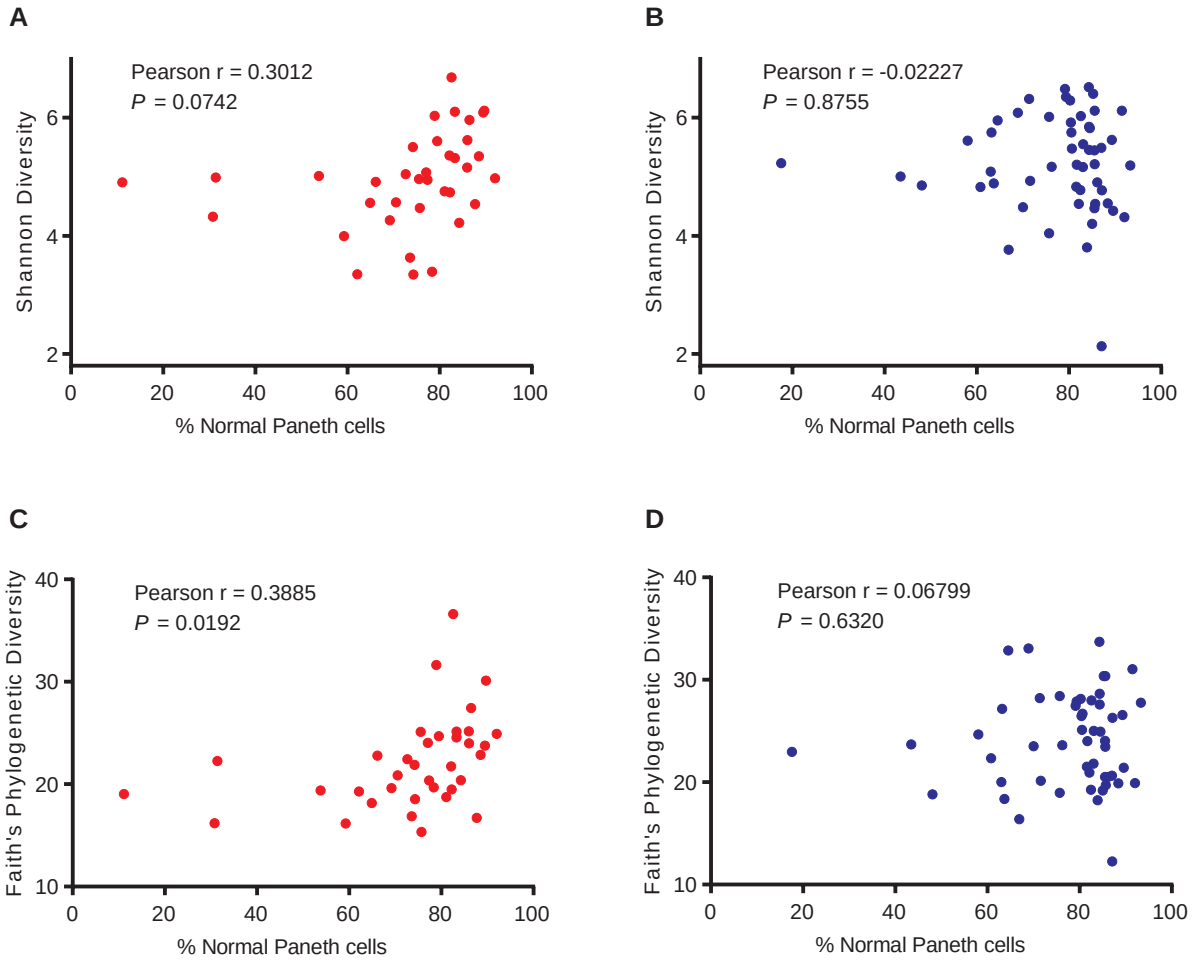
**A**



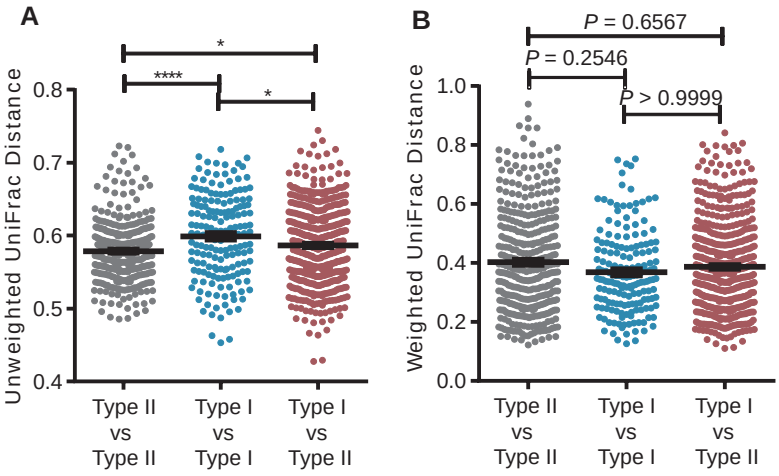

**B**

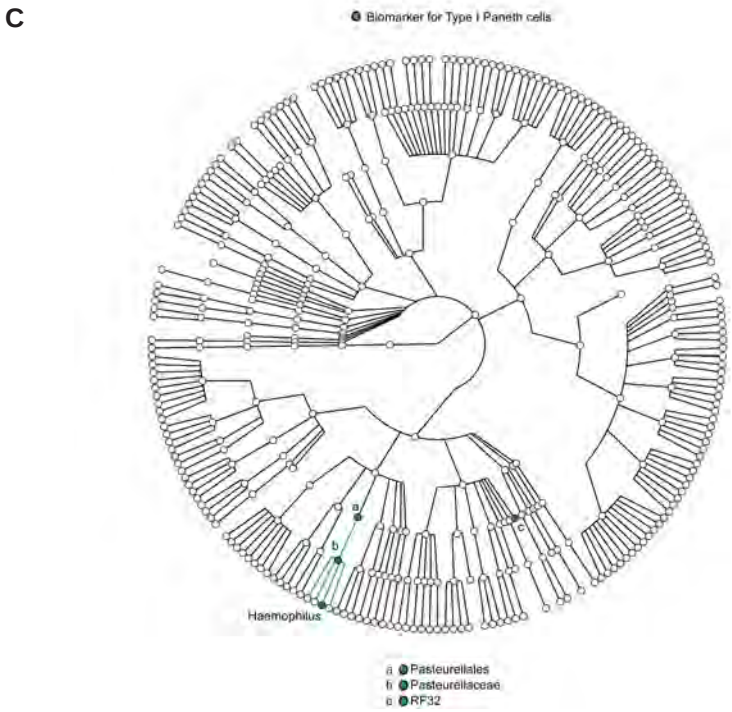


## Supplementary Figure 5



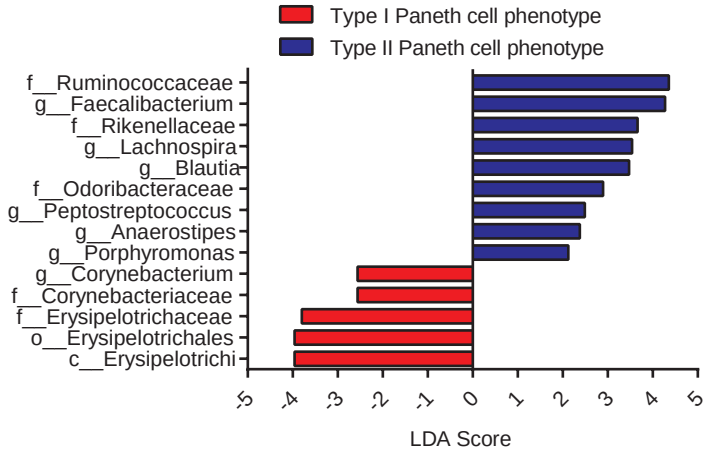
**Supplementary Figure 6**



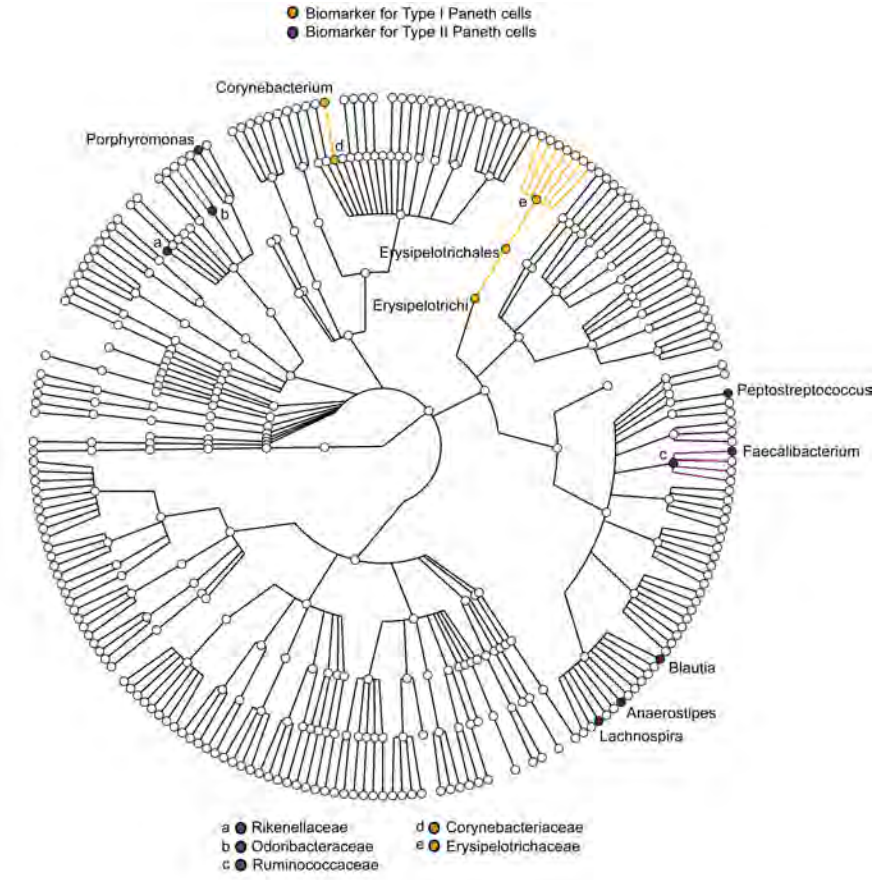



## Supplementary Figure 7

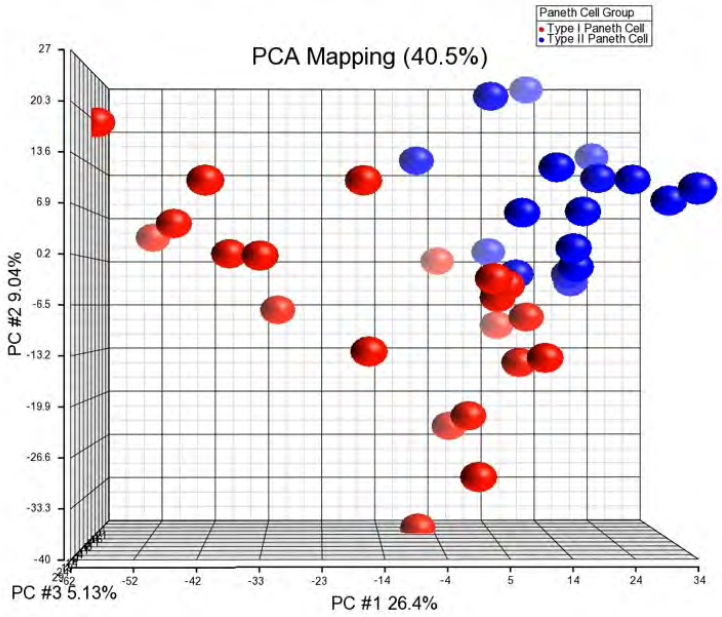
**A**




**B**



## Supplementary Figure 8



## Supplementary Figure 9