Supporting Information

Chehoud et al. 10.1073/pnas.1307855110

SI Materials and Methods

Sequences were quality-trimmed and assigned to their respective sample based on their barcodes. Sequences were removed if they contained ambiguous bases, more than eight homopolymers, primer and/or barcode mismatches, or were <200 or >800 nt long. Sequences were binned into de novo OTUs using CDHIT (1) with a 97% minimum sequence identity threshold. The most abundant sequence from each OTU was selected as the representative sequence for that OTU. Taxonomy was assigned to the representative sequences by using Ribosomal Database Project (2) with a minimum support threshold of 80%. Sequences assigned as Streptophyta taxa, found in mouse chow, were removed from the dataset. Relative abundances of taxa were compared by using a paired Student *t* test. *P* values were adjusted for multiple comparisons by using a FDR adjustment in the R statistics package. The sequences were aligned with PyNAST (3) to the Greengenes reference database (4). Because α- and β-diversity metrics are sensitive to sampling effort, the number of sequences per sample was standardized by subsampling. After filtering out hypervariable regions from the alignment using lanemask, the remaining sequences were then used to build a phylogenetic tree by using the FastTree algorithm (5). Microbial community distances were calculated between all pairs of samples using the weighted and unweighted UniFrac metric (6), Bray–Curtis distance, and the binary Jaccard distance. PCoA was performed on the resultant distance matrices to visualize relationships between samples. The nonparametric Adonis test, implemented in QIIME, was used to assess the statistical significance of groupings of distance matrices. *P* values were calculated by using 999 permutations.

The eight RNA-seq libraries were sequenced on five lanes of the HiSeq, and we obtained a total of 1,230,593,054 paired end reads. The Tuxedo software tool suite (includes TopHat, Bowtie, Cufflinks, CummeRbund) (7) was used to align, assemble, and calculate gene expression values from the RNA-seq data using default parameters. Transcripts were mapped to mouse reference genome NCBI 37/mm9. Gene expression values were calculated as fragments per kilobase of transcript per million mapped fragments (FPKM) (8). Those transcripts that did not meet the expression threshold of FPKM >1 in at least two of the eight skin samples were filtered out of the dataset.

1. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
2. Cole JR, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37:D141–D145.
3. Caporaso JG, et al. (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26(2):266–267.
4. DeSantis TZ, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7):5069–5072.
5. Price MN, Dehal PS, Arkin AP (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
6. Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235.
7. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
8. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628.

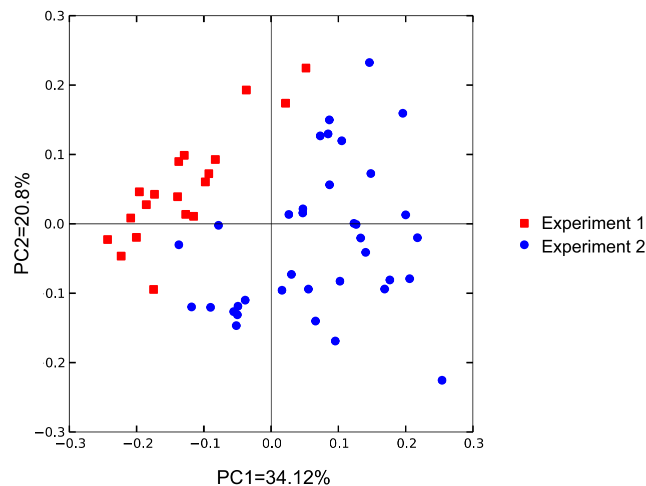

Fig. S1. Principal coordinate analysis of the weighted UniFrac illustrating the experiment-specific differences in overall skin microbiomes. Each circle or square represents an individual irrespective of time point of treatment. Red squares represent the first independent experiment; blue circles represent the second independent experiment. Percent variation explained by each principal component axis is indicated in parentheses.

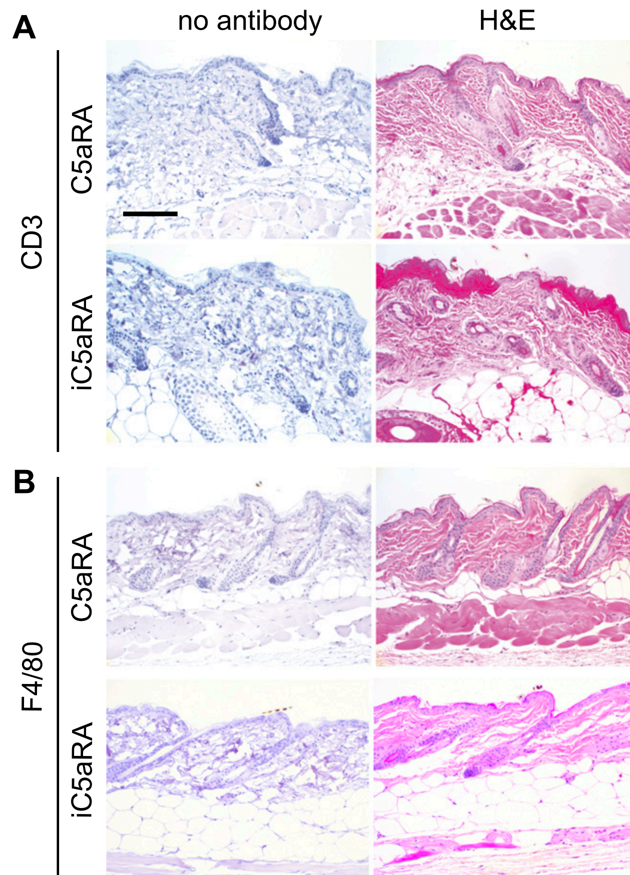

Fig. S2. Control staining for immunohistochemistry (Fig. 4) against CD3 (*A*) and F4/80 (*B*). (*Left*) Control sections with no primary antibody added before staining. (*Right*) H&E staining. (Scale bar: 100 μm.)

Table S1. R^2 and *P* values of Adonis test performed on each independent experiment comparing skin microbiota from complement C5aR-positive mice vs. C5aR-negative mice

Parameter	Experiment 1	Experiment 2
Weighted UniFrac		
R^2	0.184	0.063
P value	0.006	0.039
Unweighted UniFrac		
R^2	0.082	0.045
P value	0.012	0.010
Bray–Curtis distance		
R^2	0.117	0.060
P value	0.005	0.006
Jaccard distance		
R^2	0.082	0.042
P value	0.004	0.001

Table S2. Quantitative Real Time-PCR TaqMan primer and probe sets

Assay ID	Gene symbol
Mm00833184_s1	Ang
Mm00475988_m1	Arg1
Mm02620006_s1	C3ar1
Mm00500292_s1	C5ar1
Mm00438285_m1	Camp
Mm01143935_g1	Cfd
Mm00438186_m1	Cfh
Mm00432803_m1	Defb1
Mm04214158_s1	Defb3
Mm00651498_m1	Defb6
Mm01168928_g1	Elane
Mm00514794_m1	Ereg
Mm01267981_s1	Gpr77
Mm00434513_m1	Itgb2
Mm01203811_m1	Klk5
Mm00434787_m1	Ltf
Mm00727183_s1	Lyz2
Mm00440338_m1	Myd88
Mm00805062_m1	Nod1
Mm00467543_m1	Nod2
Mm00726747_s1	Rnase1
Mm00491347_m1	Rnase4
Mm01218201_m1	S100a7a
Mm00496696_g1	S100a8
Mm00656925_m1	S100a9
Mm00656927_g1	Saa1
Mm00441530_g1	Slpi
Mm00446095_m1	Tlr1
Mm01233819_m1	Tlr13
Mm00442346_m1	Tlr2
Mm01207404_m1	Tlr3
Mm00445273_m1	Tlr4
Mm00546288_s1	Tlr5
Mm02529782_s1	Tlr6
Mm00446590_m1	Tlr7
Mm04209873_m1	Tlr8
Mm00446193_m1	Tlr9
Mm00454744_g1	Wfdc12

This table displays TaqMan primer and probe sets (Life Technologies) used in Fig. 4*A*.