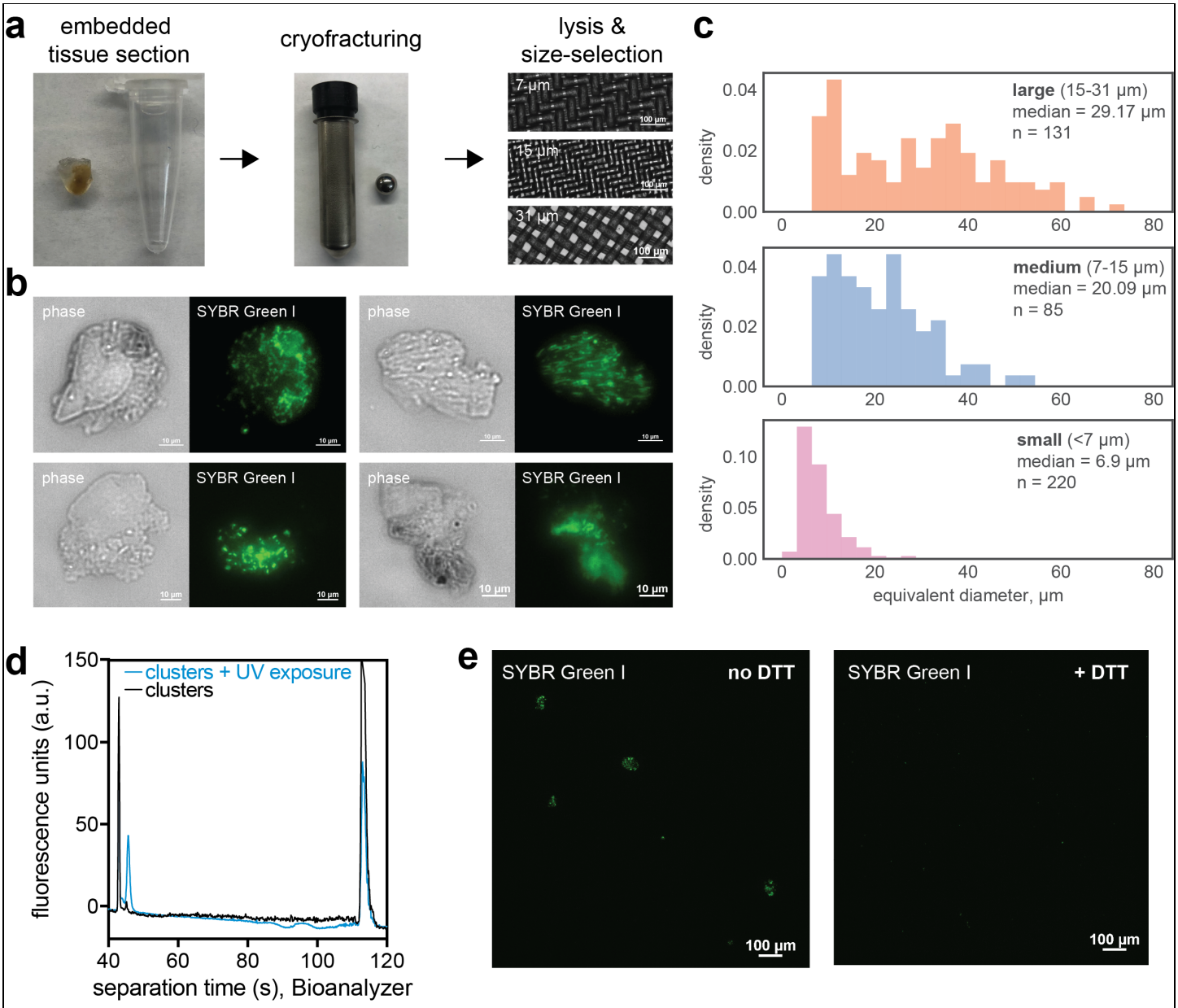
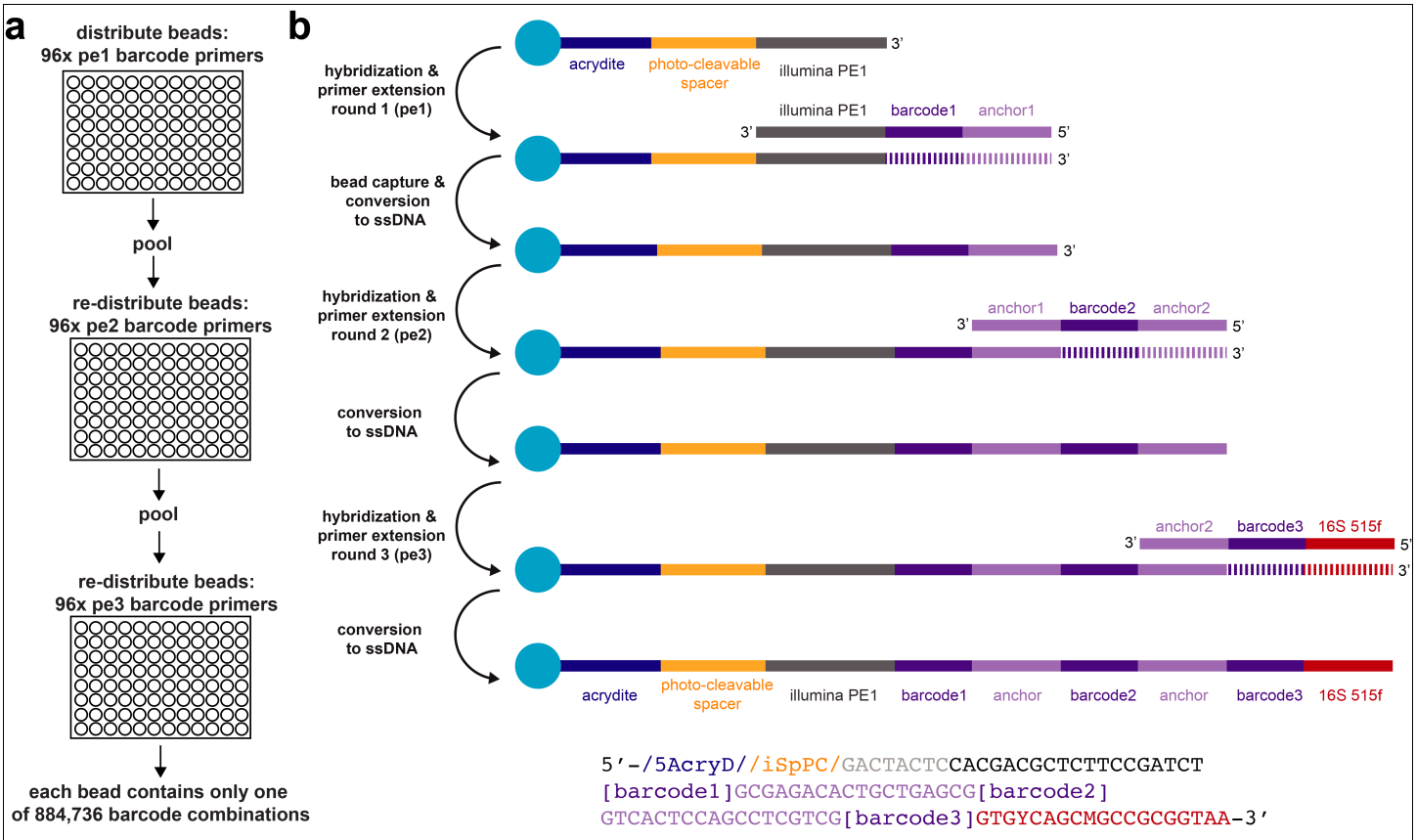


Supplementary Figure 1

Cluster generation and quality control.

a) Schematic of cluster generation process. A tissue section is fixed and embedded in a gel matrix by *in situ* acrylamide perfusion and polymerization. Shown is a murine intestinal section within a set gel as an example (excess gel is untrimmed at this step); a PCR tube placed to the right for scale. The gel-embedded sample is then subjected to cryofracturing, lysis preparation steps, and finally size-selection by passing clusters through nylon mesh filters of various sizes. **b)** Microscopy of four resulting clusters generated from murine colonic samples (size-selected for "large" clusters) visualized with phase-contrast or stained with SYBR Green I targeting genomic DNA; individual cells fixed in their original spatial orientation can be observed as punctate dots within the clusters. **c)** Resulting size distributions of clusters after size-selection to three size scales (small, medium and large); size-selected clusters were stained with SYBR Green I and imaged, clusters were identified by a fluorescence threshold, and the equivalent diameter of identified clusters was calculated using Nikon Elements AR. **d)** Photorelease of reverse amplification primer from clusters; clusters were subjected to no UV exposure or UV exposure for 10 minutes and supernatant was collected and analyzed via Agilent Bioanalyzer dsDNA HS assay; peaks at ~40s and ~110s are gel migration markers. A short primer product is observed to be released in a UV exposure dependent fashion. **e)** Degradation of cluster polyacrylamide gel matrices by exposure to reducing conditions; clusters were incubated in PCR encapsulation mix with and without 1mM DTT (i.e., final concentration of DTT in droplets) for 2 hours; without DTT clusters remain

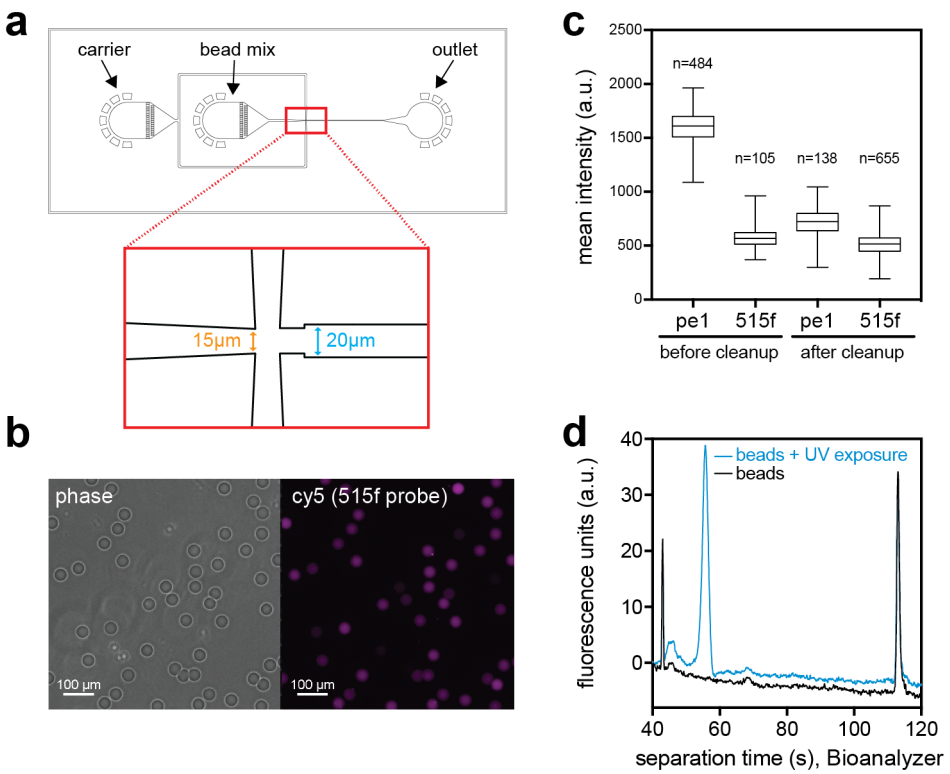
stable and retain their structure; with DTT reducing conditions, the gel matrix degrades resulting in dispersion of individual cells observable as stained puncta. Clusters were subjected to quality control (panels b-e) in at least 3 independent experiments with similar results.



Supplementary Figure 2

Barcoded bead synthesis schematic.

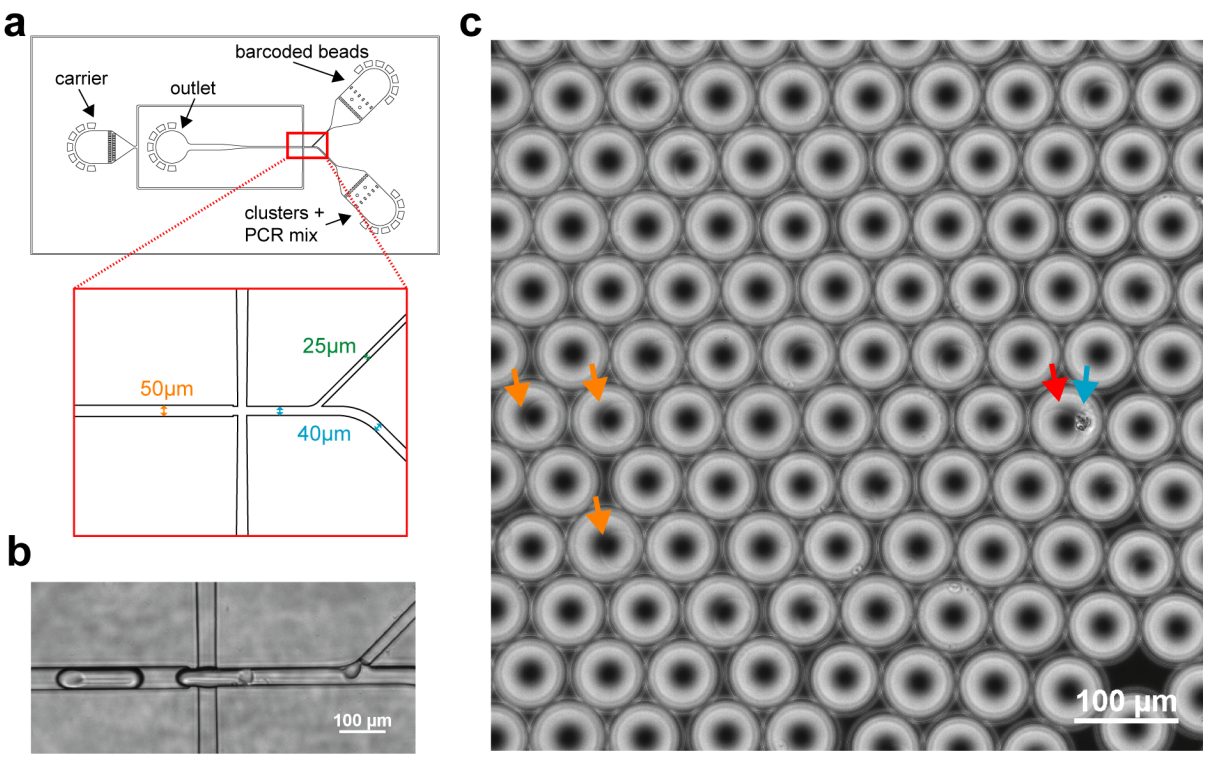
a) Beads are synthesized via a three-step split-and-pool synthesis approach, resulting in 96^3 or 884,736 possible unique barcodes. The three sets of primers are denoted primer extension sets 1-3 (i.e. pe1, pe2, and pe3). **b)** Extension strategy utilized for bead synthesis. A primer is linked to the gel bead via an acrydite linker and also contains a photocleavable linker group. Barcoded primers are hybridized to this linked primer and serve as an extension template for adding barcodes to the bead-linked primers. After each round, the extension template primer is stripped, and the next round of extension is performed. The sequence of the final primer product is indicated at the bottom.



Supplementary Figure 3

Barcoded bead quality control.

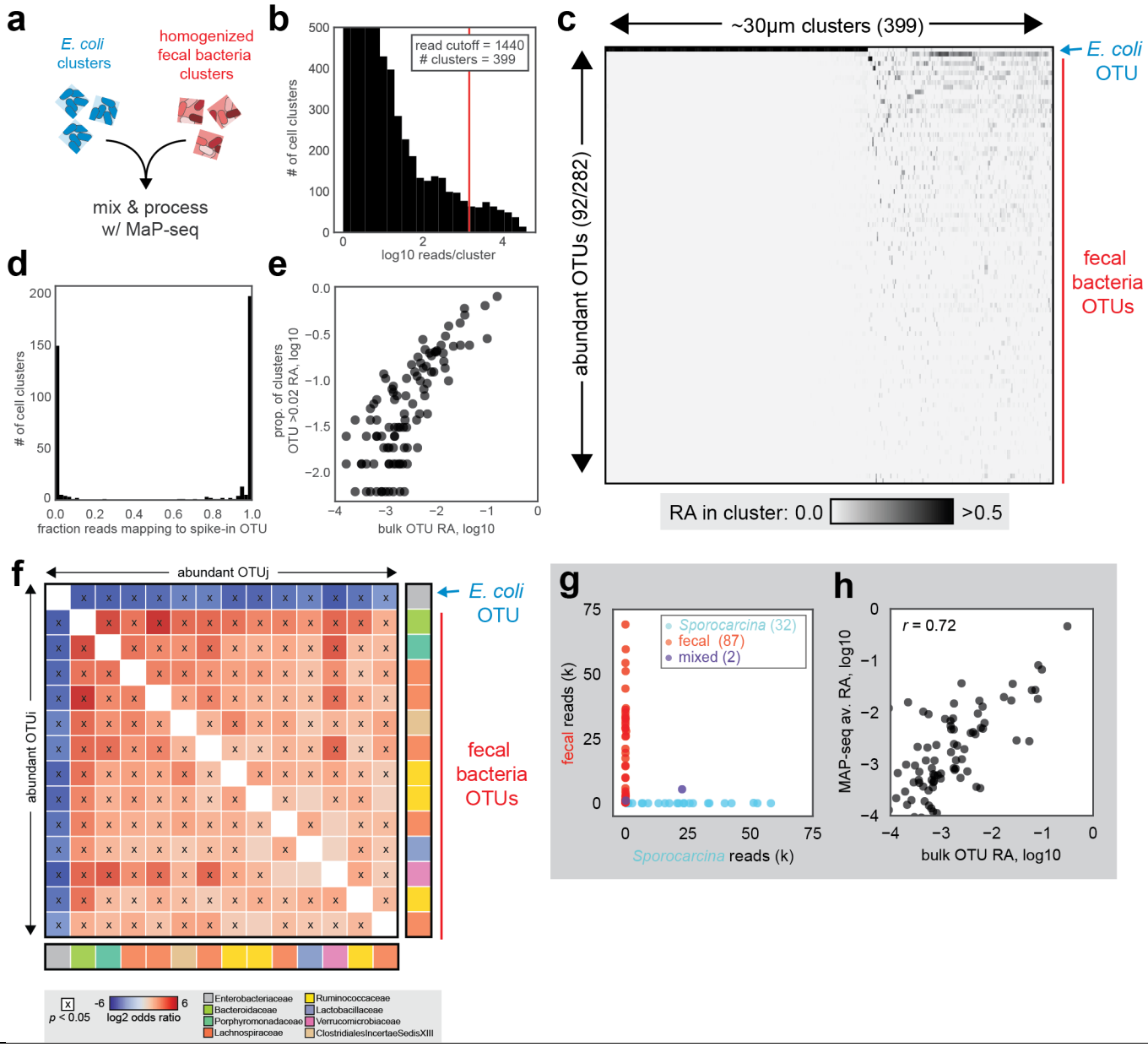
a) Schematic of the microfluidic droplet generation device utilized to fabricate barcoded beads. **b)** Image of resulting barcoded gel beads visualized by phase contrast and hybridized with a FISH probe targeted to the terminal 16S 515f primer region present in fully extended primer product (bead_515f_cy5, see Supplementary Table 4). **c)** Quantification of cleanup of primer synthesis intermediates by Exo1 cleanup; the mean fluorescence intensity of beads was quantified (using Nikon Elements AR) when hybridized by a FISH probe targeted to the 515f site present on fully extended primer product (bead_515f_cy5) or a FISH probe targeted to the pe1 primer extension site (bead_pe1_cy5, see Supplementary Table 4) present in all synthesis intermediates. Before cleanup the amount of pe1 sites on beads are higher than 515f sites, while after cleanup the amount of pe1 and 515f sites on beads are roughly equal, implying removal of un-extended primer intermediates (which contain pe1 sites, but not the terminal 515f site). Box plot indicates median, first, and third quartiles; whiskers indicate minimum and maximum; number of beads for each condition is indicated in the plot. **d)** Photorelease of amplification primer from beads; beads were subjected to no UV exposure or UV exposure for 10 minutes and supernatant was collected and analyzed via Agilent Bioanalyzer dsDNA HS assay; peaks at ~40s and ~110s are gel migration markers. A short primer product is observed to be released in a UV exposure dependent fashion. Beads were subjected to quality control (panels b-d) in at least 3 independent experiments with similar results.



Supplementary Figure 4

Microfluidic encapsulation of barcoded beads and clusters.

a) Schematic of the microfluidic droplet generation device utilized to co-encapsulate barcoded beads and clusters. Beads are packed single file to enable loading that beats Poisson encapsulation statistics expected by random loading. **b)** Image of the microfluidic device during operation. **c)** Resulting emulsion after encapsulation; beads can be observed as a faint sphere within droplets; orange arrows indicate three example droplets (of many in the field of view) with a single barcoded bead (but no clusters). One droplet with a single barcoded bead (red arrow) and a single cluster (blue arrow) can be observed in this field of view. Emulsions were subjected to quality control in at least 3 independent experiments with similar results.

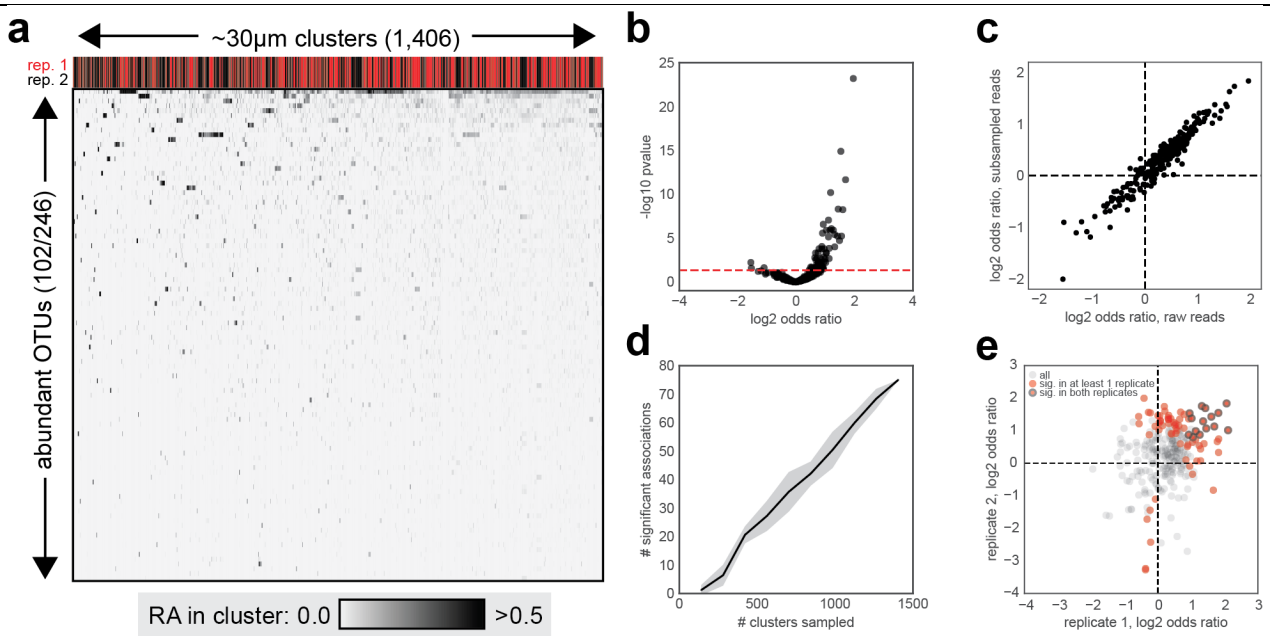


Supplementary Figure 5

Quality control of the MaP-seq technique by cluster mixing experiment.

a) Schematic of the cluster mixing experiment; clusters containing either *E. coli* or homogenized fecal bacteria are prepared, mixed, and processed through the MaP-seq pipeline. **b)** Example of resulting distribution of read counts per identified unique barcode for the mixing experiment. A conservative threshold cutoff for considering real clusters is set as the total number reads divided by 2,500 (i.e., the number of clusters that were utilized as input during microfluidic encapsulation, and assuming an equal read distribution for each cluster). The calculated read cutoff (1,440 reads) is indicated by the red line, which results in 399 clusters for downstream analysis. The Y-axis is set to a maximum of 500 clusters for visualization purposes. **c)** Resulting raw data for the mixing experiment displayed as a cluster map; columns indicate the 399 clusters passing the read cutoff and rows indicate prevalent and abundant OTUs (OTUs present >2% relative abundance in >1% of all clusters). The *E. coli* OTU is the first row, while other rows represent fecal bacterial OTUs. The plot is arranged as in Fig. 2a, RA denotes relative abundance. **d)** An alternative visualization of Fig. 1b, plotting the fraction of reads in each cluster mapping to the spike-in *E. coli* OTU; most clusters show either entirely spike-in mapping reads or no spike-in mapping reads as expected. A small number of clusters show low levels of residual contamination; for this reason, a conservative relative abundance cutoff is used throughout downstream analysis (>2% relative abundance) to classify an OTU as present within a given cluster. **e)** Detection sensitivity of MaP-seq; the relative abundance of OTUs is compared to the proportion of clusters an OTU is

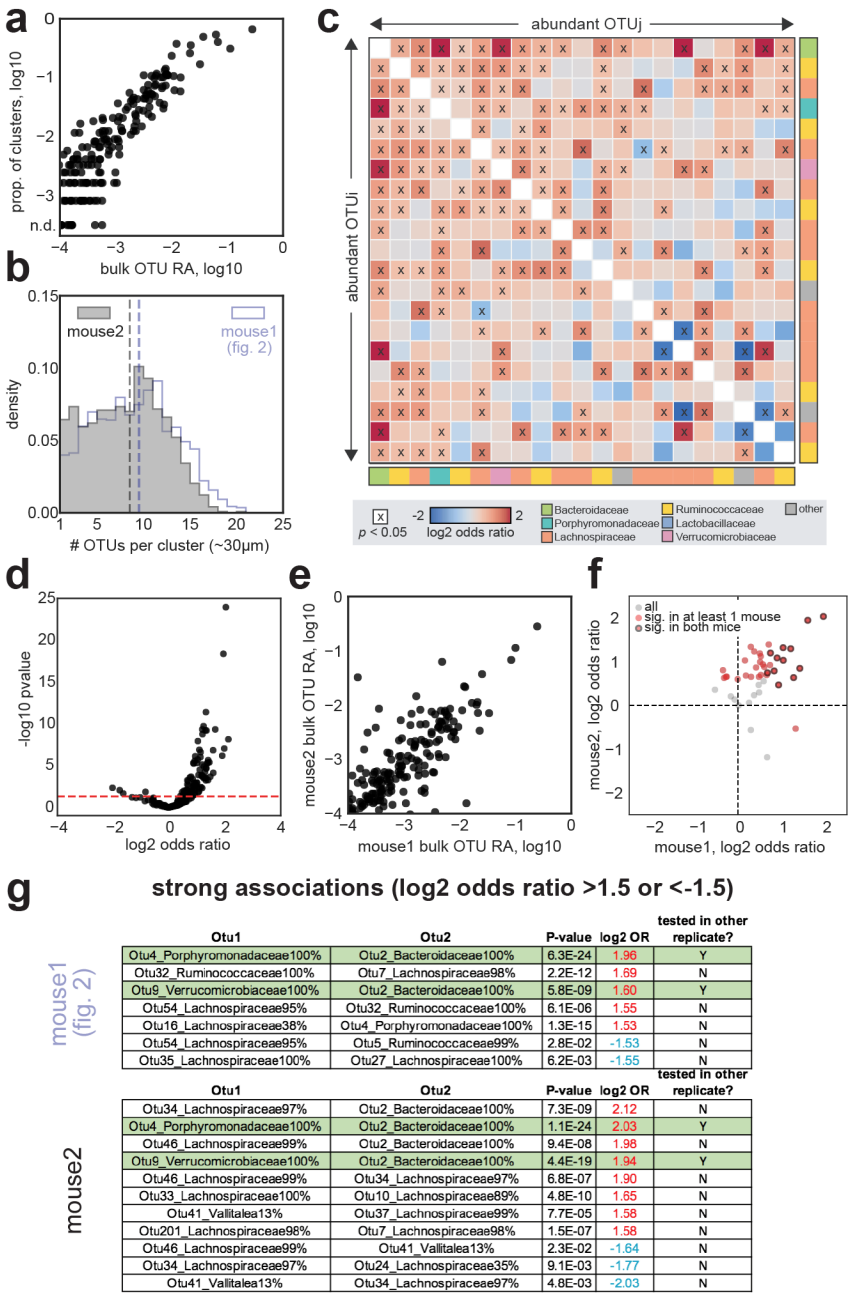
detected in (with >2% relative abundance cutoff). Higher abundance OTUs display higher detection sensitivity as expected. **f)** Detection of significant pairwise associations in the cluster mixing experiment. The two communities contain defined spatial associations; the fecal bacteria are expected to be positively associated with each other, whereas the fecal bacteria should be negatively associated with *E. coli*. Association analysis was conducted in the same manner as Fig. 2d (for n=14 abundant/prevalent OTUs across n=399 clusters, Fisher's exact test, two-sided, $p < 0.05$, FDR = 0.05); the fecal bacteria are found to be strongly associated and negatively associated with *E. coli* as expected. The associations are much stronger than observed in the murine gut (i.e. note that the color map scale spans a larger range this plot). **g-h)** To confirm technical reproducibility across different experiments and particles sizes, the cluster mixing experiment was repeated but with particles of ~20 μm median size. Fecal bacteria constituted one community and *Sporocarcina pasteurii*, an environmental taxa constituted a second community. g) is analyzed as in Fig. 1b and h) as in Fig. 1c. This revealed low mixing rates (1.65% mixed), negligible contamination (<0.003% of reads) and good correlation to bulk 16S sequencing (Pearson correlation $r = 0.72$, n=77 OTUs with greater than 0.01% RA are displayed, MaP-seq average RA calculated over n=87 fecal clusters), confirming technical reproducibility of the technique across different experiments and particle sizes.



Supplementary Figure 6

Spatial association detection and technical reproducibility.

a) Expanded view of Fig. 2a; abundant and prevalent OTUs (>2% abundance in >1% of clusters) are displayed; the cluster map is organized as in Fig. 2a, RA denotes relative abundance. The column indicators on top indicate the technical replicate each cluster originated from (red, replicate 1; black, replicate 2). **b)** A volcano plot visualization of data from Fig. 2d (associations across n=1,406 clusters and n=24 prevalent/abundant OTUs); red dotted line indicates threshold for statistically significant associations (Fisher's exact test, two-sided, $p<0.05$, FDR = 0.05). **c)** Correlation between association detection utilizing raw or subsampled reads. Reads were subsampled for all clusters to the minimum read cutoff (717 reads) and association detection was performed as before. The resulting odds ratios of pairwise associations (n=276 unique pairwise associations between the 24 prevalent/abundant OTUs) were highly correlated to those calculated from the raw reads (Pearson r = 0.96). These results suggest that variable read counts for each cluster do not significantly alter detected associations (i.e. due to use of a 2% abundance threshold). **d)** Dependence of association detection on cluster sampling depth. Clusters from the full dataset (n=1,406 clusters) were subsampled, and the same association detection was performed (Fisher's exact test, two-sided, $p<0.05$, FDR = 0.05). The number of significant associations detected is plotted; the line indicates the mean and error band indicates the standard deviation of ten iterations of subsampling. The number of significant associations detected linearly increases with the number of clusters sampled, implying even deeper cluster data collection could enable characterization of weaker associations between less abundant taxa. **e)** Technical reproducibility of association detection between the two technical replicate datasets of the same sample. Association detection was performed on each technical replicate, and the calculated odds ratio of association is plotted for all pairwise associations. For pairwise associations (n=276 total associations) detected as significant in at least one of the two replicates (n=74 associations), the sign of association is the same between both replicates for the majority of cases (64/74 associations). For associations detected as significant in both replicates (n=15 associations) the sign is the same in all cases. These results indicate good correspondence of detected microscopic spatial associations between technical replicates.



g

strong associations (log2 odds ratio >1.5 or <-1.5)

mouse1 (fig. 2)

Otu1	Otu2	P-value	log2 OR	tested in other replicate?
Otu4_Porphyromonadaceae100%	Otu2_Bacteroidaceae100%	6.3E-24	1.96	Y
Otu32_Ruminococcaceae100%	Otu7_Lachnospiraceae98%	2.2E-12	1.69	N
Otu9_Verrucomicrobiaceae100%	Otu2_Bacteroidaceae100%	5.8E-09	1.60	Y
Otu54_Lachnospiraceae95%	Otu32_Ruminococcaceae100%	6.1E-06	1.55	N
Otu16_Lachnospiraceae38%	Otu4_Porphyromonadaceae100%	1.3E-15	1.53	N
Otu54_Lachnospiraceae95%	Otu5_Ruminococcaceae99%	2.8E-02	-1.53	N
Otu35_Lachnospiraceae100%	Otu27_Lachnospiraceae100%	6.2E-03	-1.55	N

mouse2

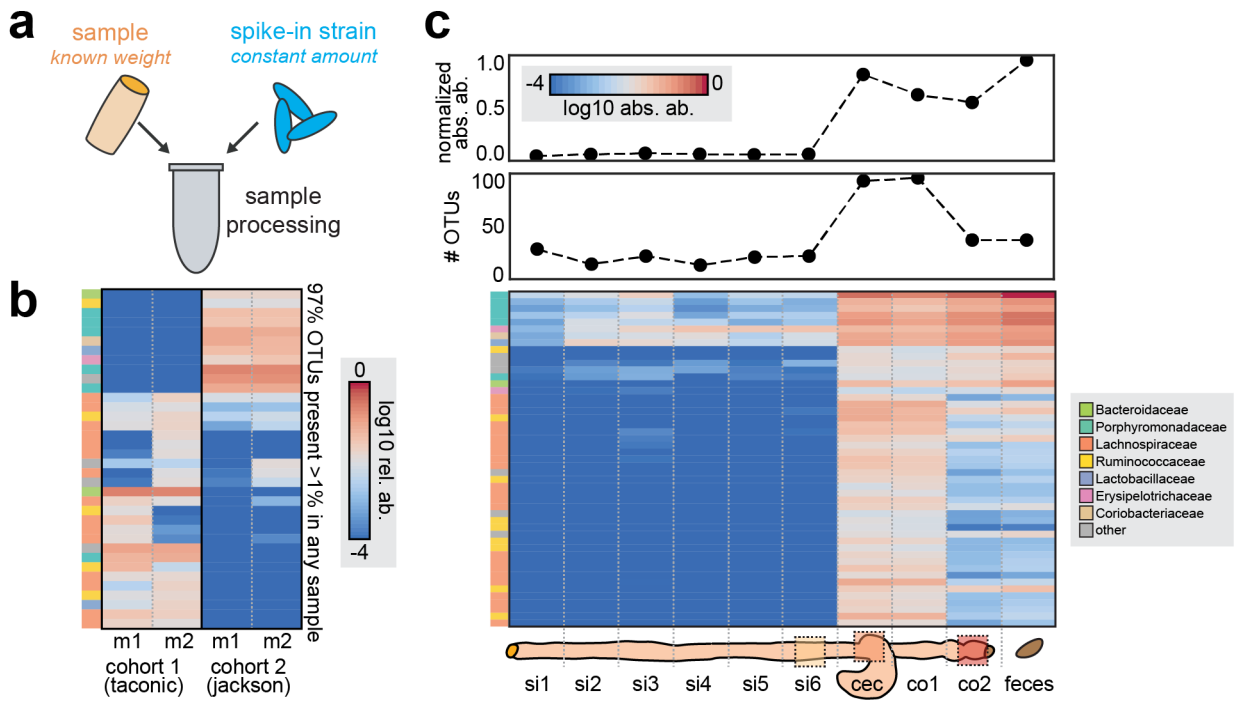
Otu1	Otu2	P-value	log2 OR	tested in other replicate?
Otu34_Lachnospiraceae97%	Otu2_Bacteroidaceae100%	7.3E-09	2.12	N
Otu4_Porphyromonadaceae100%	Otu2_Bacteroidaceae100%	1.1E-24	2.03	Y
Otu46_Lachnospiraceae99%	Otu2_Bacteroidaceae100%	9.4E-08	1.98	N
Otu9_Verrucomicrobiaceae100%	Otu2_Bacteroidaceae100%	4.4E-19	1.94	Y
Otu46_Lachnospiraceae99%	Otu34_Lachnospiraceae97%	6.8E-07	1.90	N
Otu33_Lachnospiraceae100%	Otu10_Lachnospiraceae89%	4.8E-10	1.65	N
Otu41_Vallitalea13%	Otu37_Lachnospiraceae99%	7.7E-05	1.58	N
Otu201_Lachnospiraceae98%	Otu7_Lachnospiraceae98%	1.5E-07	1.58	N
Otu46_Lachnospiraceae99%	Otu41_Vallitalea13%	2.3E-02	-1.64	N
Otu34_Lachnospiraceae97%	Otu24_Lachnospiraceae35%	9.1E-03	-1.77	N
Otu41_Vallitalea13%	Otu34_Lachnospiraceae97%	4.8E-03	-2.03	N

Supplementary Figure 7

Spatial organization in the distal colon of a second co-housed mouse.

a) Correlation between bulk 16S sequencing measurements to OTU prevalence across clusters as obtained by MaP-seq (n=1,228 clusters total); plotted in the same manner as Fig. 2b; a high correlation is observed as with the first replicate (r=0.9, Pearson correlation coefficient, n=182 OTUs greater than 0.01% RA as measured by bulk 16S sequencing). n.d. indicates not detected. **b)** Histogram of the number of OTUs per cluster, plotted in the same manner as Fig. 2c. The replicate mouse data (i.e., mouse 2) is plotted with grey shaded bars (median 8 OTUs/cluster), while data from the first mouse (i.e., mouse 1, see Fig. 2) is plotted with purple open outline (median 9 OTUs/cluster). **c)** Detection of pairwise associations in the second replicate mouse, plotted as in Fig. 2d. n=1,228 detected clusters are aggregated across two technical replicates, and for n=21 prevalent/abundant OTUs, 94 statistically significant pairwise associations were detected (Fisher's exact test, two-sided, $p < 0.05$, FDR = 0.05). **d)** Volcano plot of detected associations (Fisher's exact test, two-sided, $p<0.05$, FDR = 0.05), sample sizes as in c), plotted as in Supplementary Fig. 6b. **e)** Correlation of OTU relative abundance as obtained by bulk 16S sequencing of colonic samples between the two mice; a low correlation

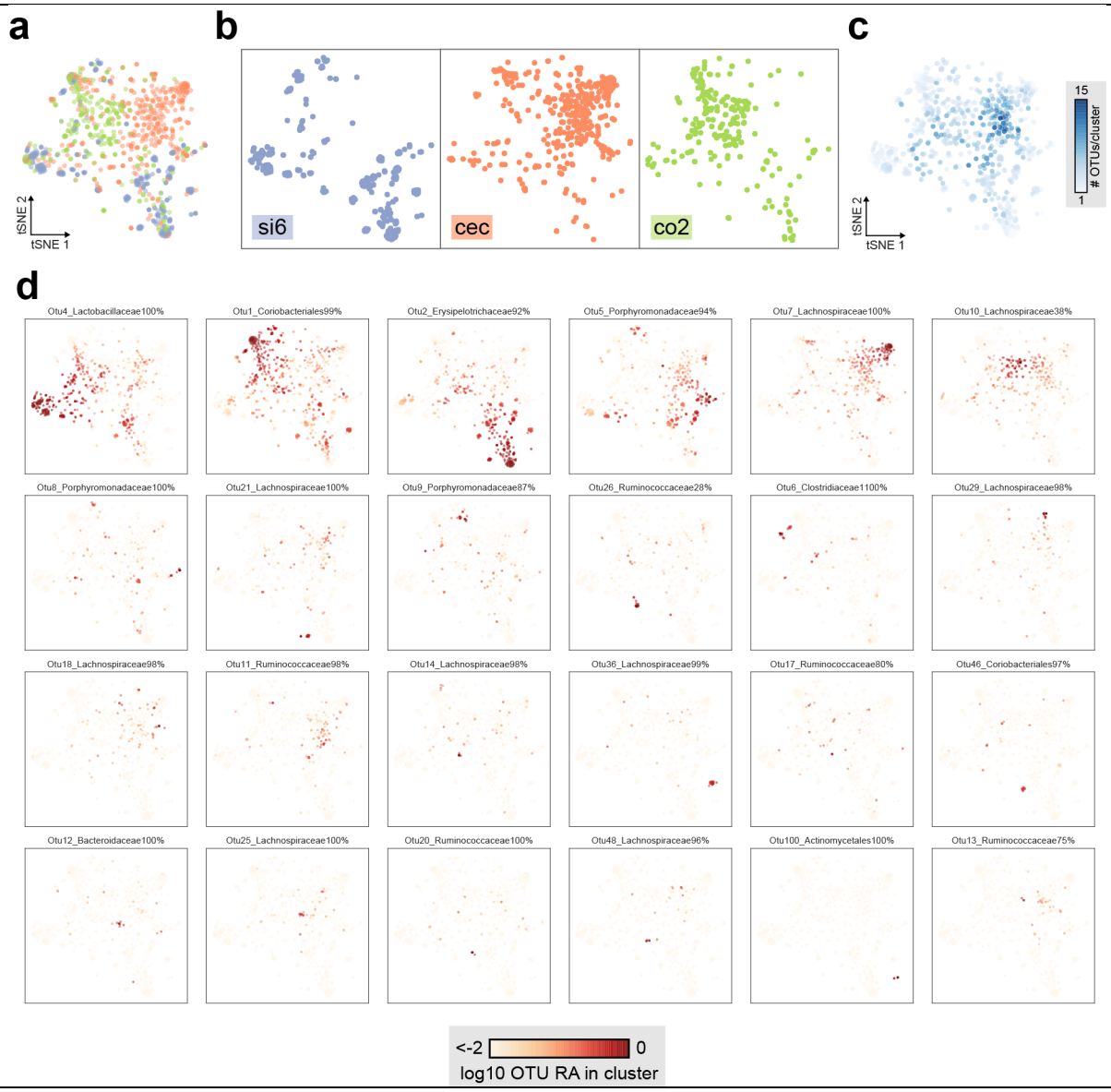
is observed (r=0.76, Pearson correlation coefficient, n = 168 OTUs greater than 0.01% RA in both samples) indicating microbiota community heterogeneity even between co-housed mice. **f)** Correlation of associations across biological replicates. Based on filtering criteria for testing for pairwise associations, 24 OTUs were tested for pairwise associations in mouse 1 and 21 OTUs were tested for pairwise associations in mouse 2. Between these 24 and 21 OTUs, only 10 were shared, again highlighting community heterogeneity. For these 10 OTUs, the odds ratio of resulting pairwise associations (n=45 total possible pairwise associations) is plotted as in Supplementary Fig. 6d. For statistically significant associations detected in both replicates, the sign is the same (positive) in both replicates, indicating that some strong associations are conserved across biological replicates. **g)** The strongest associations (log2 odds ratio >1.5 or <-1.5, as calculated in c) and Fig. 2d) with statistical significance in both replicate datasets; top shows mouse 1 data and bottom shows mouse 2 data. Despite the heterogeneity between colonic communities, two strong associations are reproducibility observed in both datasets (highlighted in green). Other strong associations include OTUs that were present at low abundance in the corresponding biological replicate and thus pairwise associations were not tested (i.e. indicated in the last column of table, “tested in other replicate?”).



Supplementary Figure 8

Spike in sequencing and characterization of microbiota across the murine gastrointestinal tract.

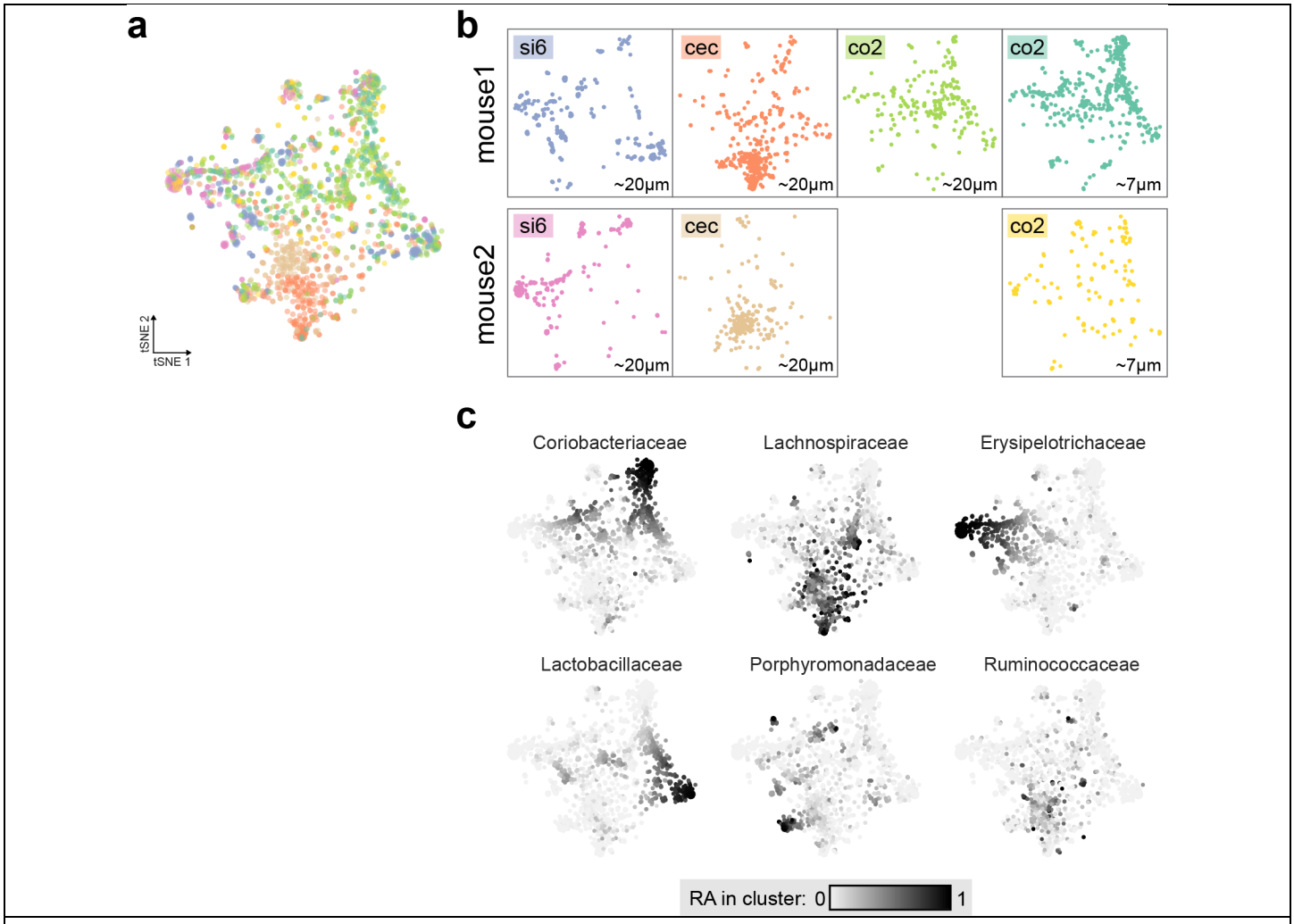
a) Schematic of spike-in sequencing approach; a known weight of a sample is added to a constant amount of spike-in allowing for estimation of relative bacterial densities on an absolute scale. **b)** The second cohort of mice characterized (Fig. 3) displays high community heterogeneity as compared to the first cohort (Fig. 2). Both cohorts consisted of two co-housed mice (m1, m2). Plotted is bulk 16S sequencing of whole distal colon sections; columns indicate OTUs detected in any sample >1% relative abundance, and shading indicates relative abundance in log10 scale. Across all the samples, only ~20% of 97% OTUs are shared (i.e., OTUs >0.01% relative abundance in both cohorts). This heterogeneity highlights challenges with microbiota profiling techniques which are not unbiased, i.e., in which probes must be designed in advance. **c)** Microbiota composition and community properties across the GI tract for a second co-housed mouse plotted in the same manner as Fig. 3a. As in the other mouse, increasing absolute abundance (~34 fold higher bacterial density in large vs. small intestine) and higher species diversity are observed passing down through the gastrointestinal tract.



Supplementary Figure 9

Additional information for tSNE analysis of clusters across the GI tract.

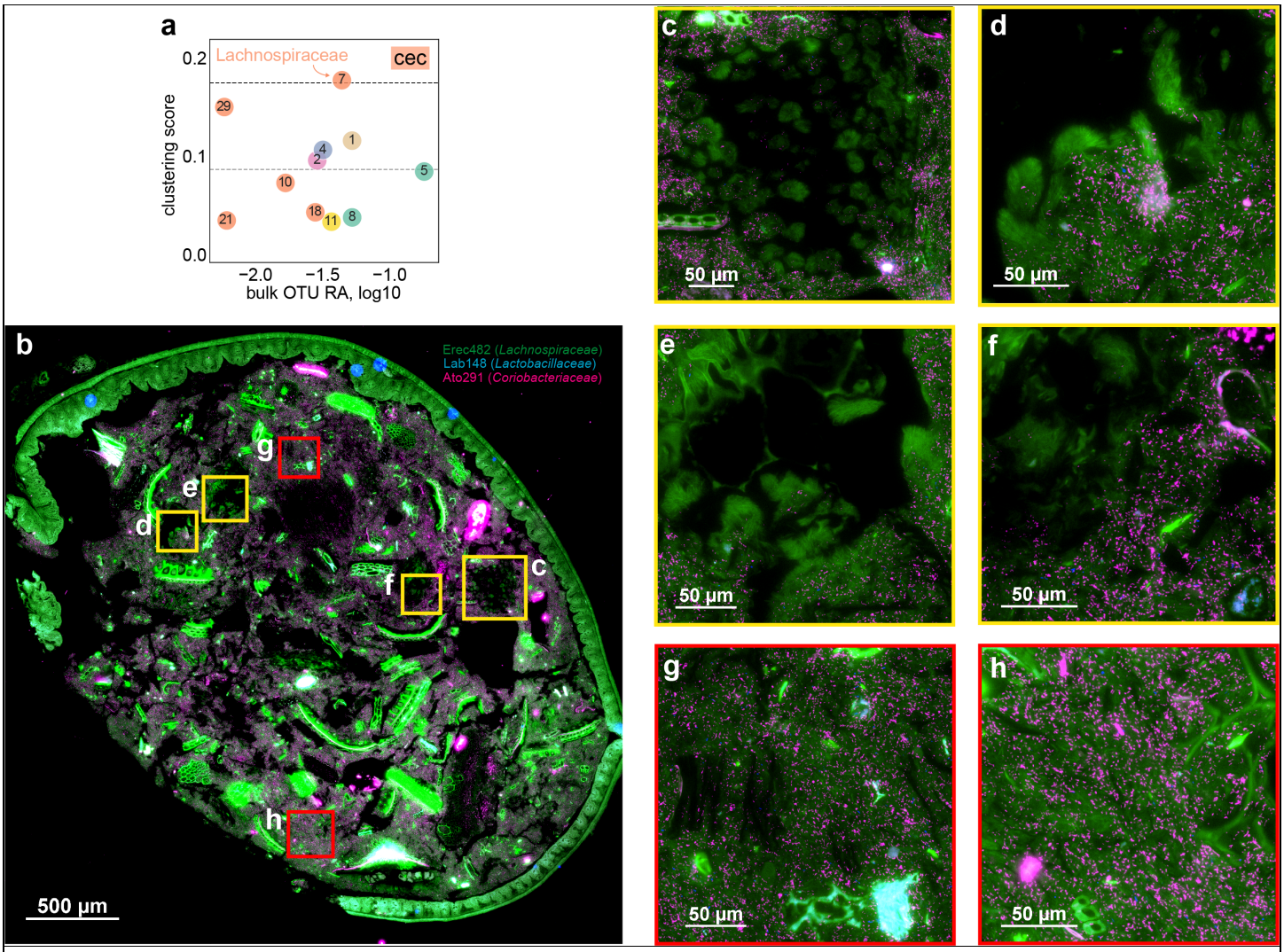
a) Same figure as Fig. 3c for reference; n=1,050 clusters are shown. **b)** Clusters from each source (si6, cec, co2) plotted separately on the same tSNE manifold for visualization purposes. **c)** Clusters are shaded by the number of OTUs per cluster (OTUs >2% RA in the subsampled dataset utilized for tSNE analysis). **d)** Clusters are shaded by the log10 relative abundance of individual OTUs within each cluster. The cecum-specific clusters, enriched for *Lachnospiraceae* taxa, show high diversity. The 24 OTUs with the highest average relative abundance across all clusters are displayed.



Supplementary Figure 10

tSNE analysis of clusters across the GI tract of a second co-housed mouse.

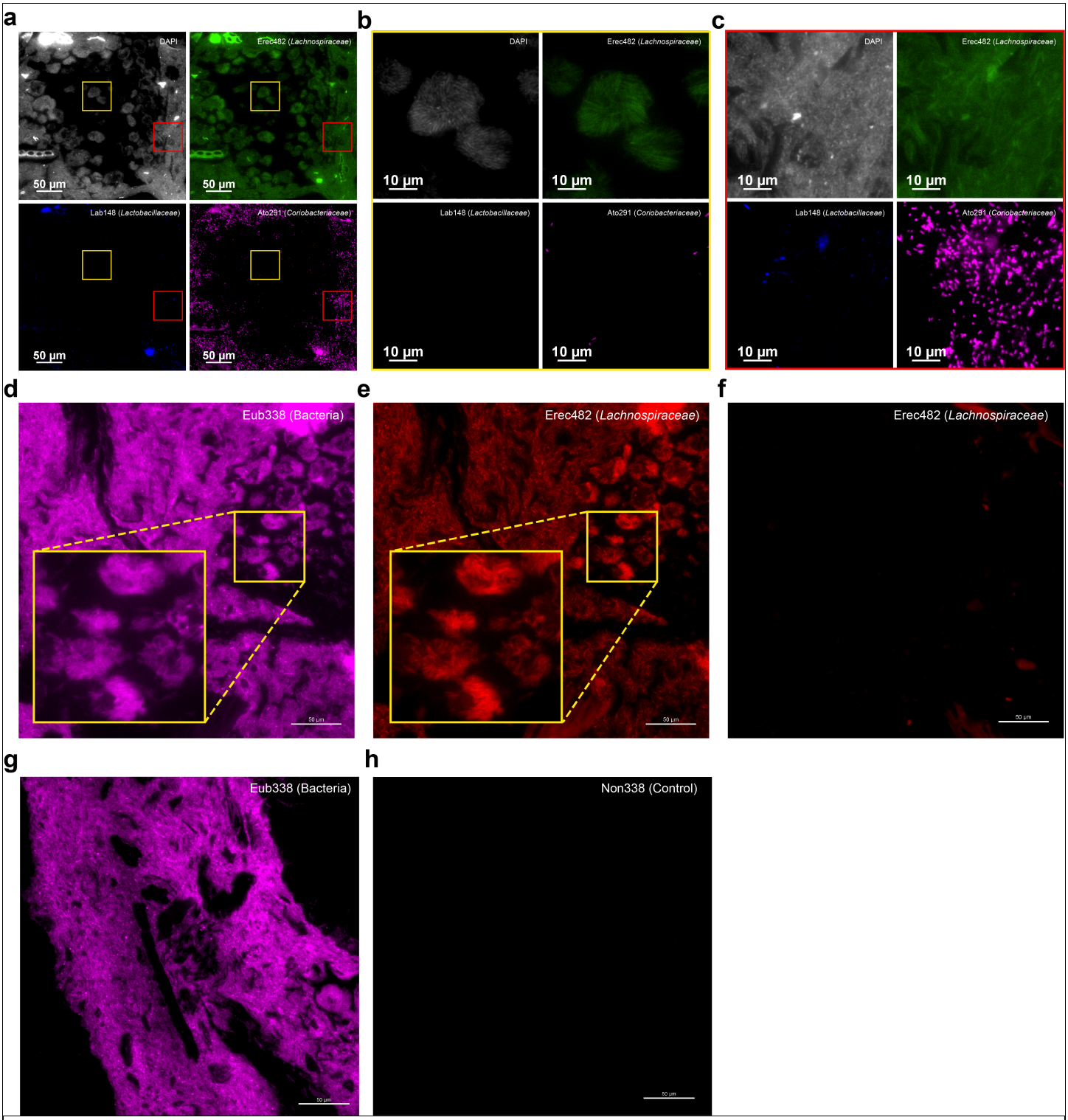
a) tSNE visualization of clusters from three GI tract regions across two co-housed mice replicates utilizing Bray-Curtis dissimilarity of OTU relative abundances (subsampled to 239 reads across all clusters). mouse1 si6, cec, co2 at 20 µm is the same dataset as in Fig. 3c. The number of clusters analyzed is: mouse1 si6: n=386, mouse1 cec: n=405, mouse1 co2: n=259, mouse1 co2 7 µm: n=529, mouse2 si6: n=261, mouse2 cec: n=223, mouse2 co2 7 µm: n=151. Clusters are colored by their source, as per the colors in b). mouse2 co2 at 20 µm clusters were not obtained due to technical failure of the reaction. Note that due to the relatively low cluster coverage for mouse2 samples, pairwise association analysis was not possible. **b)** Clusters from each source are plotted separately on the same tSNE manifold for visualization purposes. Clusters from each site in second co-housed mouse co-clustered with clusters corresponding sites in the first mouse, showing good correspondence between biological replicates. Clusters obtained for the colon at two different size scales also displayed co-clustering. **c)** Each cluster is colored by the relative abundance of the six most abundant families within each cluster (linear scale). As in mouse1, cecum clusters of mouse2 show dense groups dominated by *Lachnospiraceae* not found in other sites.



Supplementary Figure 11

Analysis of taxa with altered spatial structuring in the cecum.

a) OTU clustering or self-aggregation in the murine cecum; for prevalent OTUs (>2% RA in >10% all clusters) the proportion of times an OTU is observed as the majority of the cluster (>50% relative abundance) is plotted. Grey dotted line indicates the average clustering value, and black dotted line indicates two times the average clustering value. **b)** FISH imaging of a cecum section from the same sample profiled by MaP-seq; green is Erec482 probe targeting *Lachnospiraceae*, blue is Lab148 probe targeting *Lactobacillaceae*, and magenta is Ato291 probe targeting *Coriobacteriales*. **c-f)** Four representative regions showing Erec482 targeted *Lachnospiraceae* displaying self-aggregating clusters. The source of each of the four regions is indicated by a yellow outline in b). **g-h)** Two representative regions showing areas with no *Lachnospiraceae* self-aggregation. The source of the two regions is indicated by a red outline in b). FISH staining was performed on two independent sections from the same sample with similar results; regions shown (with and without Lachnospiraceae self-aggregation) are representative of multiple regions from the full section.

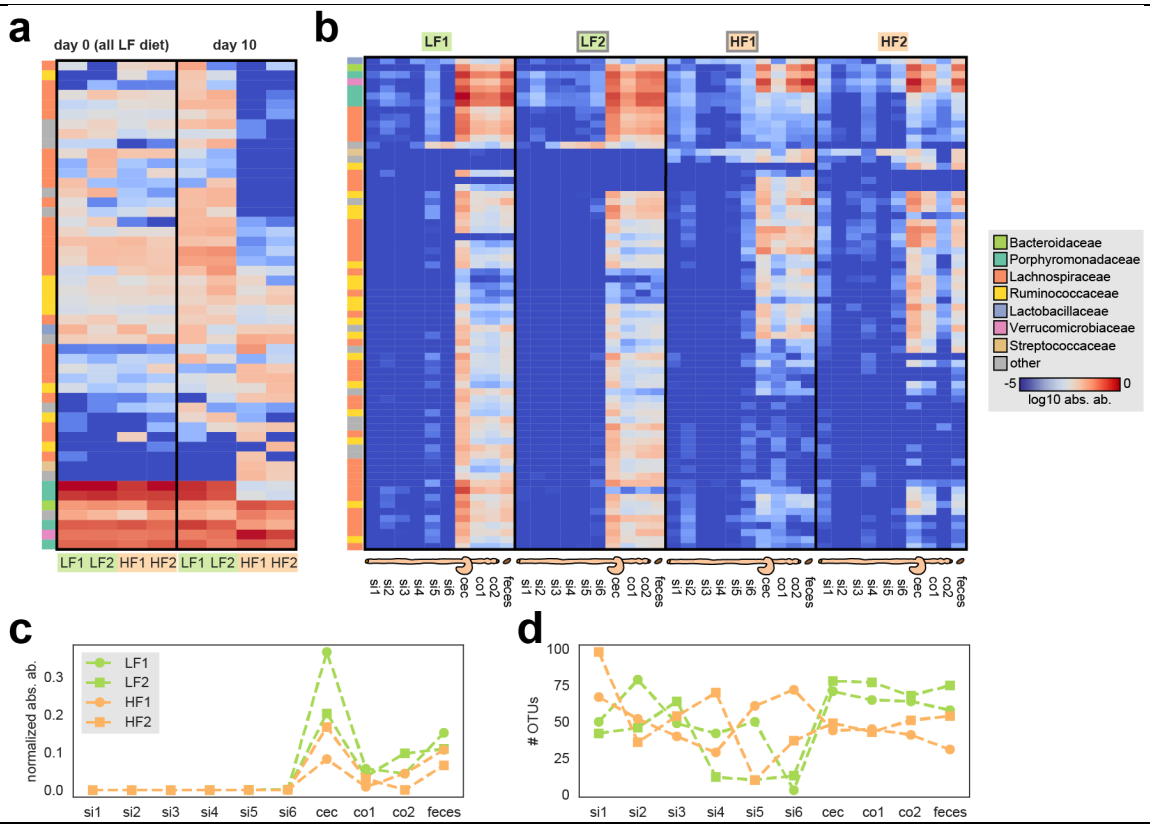


Supplementary Figure 12

Erec482-stained bacterial aggregations appear to exclude other bacteria and additional imaging controls.

a) The same region shown in fig. S11c is displayed, but the four channels are displayed independently. **b)** To investigate if other bacteria not targeted by the utilized FISH probes (Lab148 and Ato291 probes) may be present in the apparent Erec482 targeted

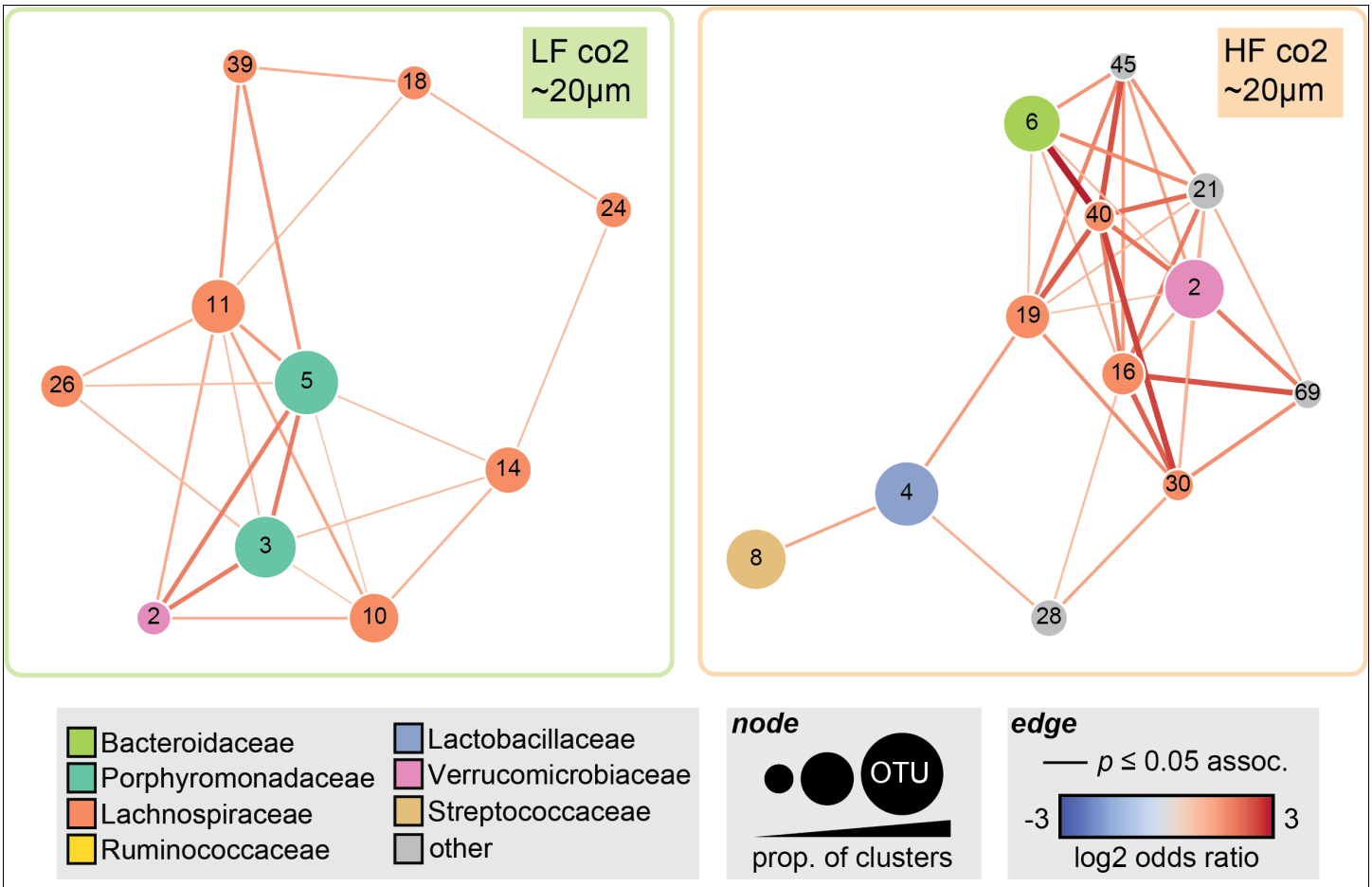
Lachnospiraceae clusters, DAPI counterstaining (targeted to cell gDNA) was also investigated. A bacterial aggregation is displayed from the image in a); the region is indicated by a yellow outline. Apparent Erec482 aggregations display a single bacterial morphology under DAPI staining, and the DAPI staining co-localizes with Erec482 probe fluorescence. These results imply that the apparent Erec482 *Lachnospiraceae* clusters exclude other bacteria in the cecum. **c)** A representative region not displaying Erec482 targeted *Lachnospiraceae* clusters; a variety of cell morphologies are observable with DAPI staining and Erec482, Lab148 and Ato291 stained bacteria are present. The region displayed is indicated by a red outline in a). **d-f)** To validate the Erec482-stained structures, we performed two-color FISH utilizing the Erec482 probe (this time with a Cy3 fluorophore) and a Eub338 probe targeted to all bacteria. d) shows the Eub338 probe, e) shows the Erec482 probe, demonstrating that similar aggregations as observed previously (i.e. see inset zoom of specific structures, yellow outline) are co-stained in both channels, indicating they are bacteria. f) shows a different section not stained with a Cy3 probe but with same exposure settings, indicating that the Erec482 staining is specific and not due to autofluorescence. **g-h)** Additional controls showing Eub338 and Non338 (scrambled control probe) FISH with same exposure settings. g) shows Eub338 probe, h) shows Non338 probe. Lumenal bacteria are bound by the Eub338 and not Non338 probe validating the FISH staining conditions. For all experiments shown in this figure, FISH staining was performed on at least two independent sections from the same sample with similar results; regions shown are representative of multiple regions from the full section.



Supplementary Figure 13

Bulk characterization of microbiota across the murine gastrointestinal tract during dietary perturbation.

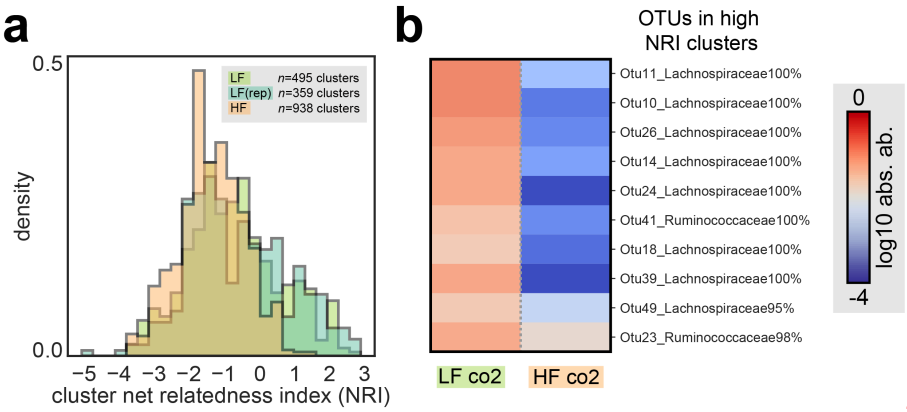
All panels are plotted in the same manner as Fig. 3a. **a)** Microbiota absolute abundances in fecal matter before dietary change (left, day 0, all mice co-housed and fed LF diet) and after dietary change (right, day 10, LF1/LF2 fed LF diet, HF1/HF2 fed HF diet). **b)** Microbiota absolute abundances across the GI tract for all mice at day 10. Distal colon samples from mice LF2 and HF1 (grey outline on label) were characterized by MaP-seq as shown in Fig. 4. In this cohort of mice, the distal small intestine is dominated by a taxa corresponding to SFB, which is lost on the HF diet. **c)** Normalized bacterial absolute abundance in different regions; as in the previous cohort increasing densities are observed across the GI tract. d) Number of OTUs (alpha diversity) in different regions; mice fed the HF diet show a marked decrease in bacterial diversity in the large intestine compared to mice fed the LF diet.



Supplementary Figure 14

Spatial associations in HF and LF fed mice.

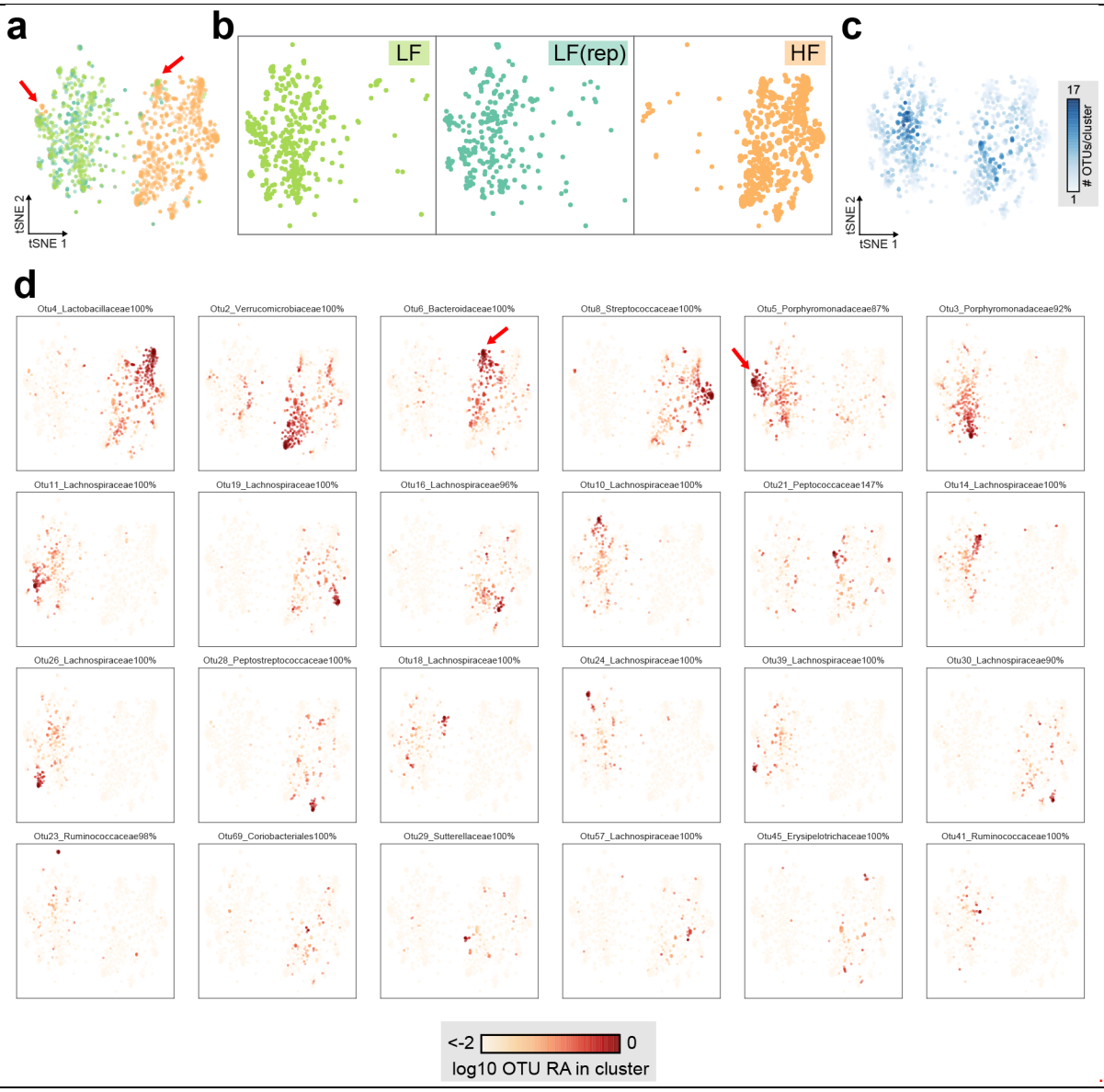
Pairwise spatial associations for prevalent and abundant OTUs visualized as a force directed graph. Nodes indicate OTUs, and sizes are proportional to prevalence of OTUs across clusters and coloring represent taxonomy at the family level. Edges represent statistically significant associations (Fisher's exact test, two-sided, $p < 0.05$, FDR = 0.05). n=495 clusters are utilized for calculation of pairwise associations for both samples (the HF dataset is downsampled to match the number of clusters in the LF dataset). Clusters from the LF and HF diets each show distinct sets of prevalent and abundant OTUs (OTU 2 is the only taxa shared between the datasets when the abundance/prevalence criteria is imposed). The LF clusters show strong co-association between *Porphyromonadaceae* taxa as observed previously, and positive associations between *Lachnospiraceae* taxa. The HF diet shows strong positive associations between the taxa of diverse bacterial families present. Note that the color scale for association odds ratio is increased (log2 odds ratio -3 to 3) compared to Fig. 2 and Fig. 3 (log 2 odds ratio -2 to 2) to accommodate stronger positive associations observed in the HF diet.



Supplementary Figure 15

Additional analysis of cluster net relatedness index (NRI) during dietary perturbation.

a) To confirm the robustness of the cluster NRI values (Fig. 4c) we calculated NRI values on the same subsampled read dataset utilized for tSNE analysis (121 reads for all clusters) and for an additional biological replicate from an adjacent colonic segment of the same LF mouse. Calculations were performed in the same manner as Fig. 4c. The LF clusters showed a subset of high NRI clusters, which is not observed in HF clusters. These results demonstrate the calculation of NRI is robust to subsampling the dataset and demonstrate replication by an independent dataset generated from an adjacent tissue segment. b) For clusters across the LF and LF(rep) datasets with NRI values > 1, the top ten OTUs with greatest average abundance in these clusters are shown (top to bottom, decreasing average abundance). These taxa constitute ~60% of the total abundance in the high NRI clusters. The absolute abundance of these taxa as assessed by bulk sequencing of colonic tissue is shown for the LF and HF colon sections for mice analyzed by MaP-seq (i.e. subset of heatmap shown in Fig. 4a). All the high NRI taxa are decreased or almost completely lost upon switch to the HF diet.



Supplementary Figure 16

Additional information for tSNE analysis of dietary perturbation clusters.

a) Same figure as Fig. 4d for reference; the number of clusters shown are LF co2 n=495, LF(rep) n=359, HF co2 n=938. **b)** Clusters from each source (LF, LF(rep), HF) plotted separately on the same tSNE manifold for visualization purposes. **c)** Clusters are shaded by the number of OTUs per cluster (OTUs >2% RA in the subsampled dataset utilized for tSNE analysis). **d)** Clusters are shaded by the log10 relative abundance of individual OTUs within each cluster. Red arrows on *Bacteroidaceae* OTU 6 and *Porphyromondaceae* OTU 5 plots indicate the same regions in Fig. 4d where clusters dominated by each of these taxa respectively are observed in both diets. The 24 OTUs with the highest average relative abundance across all clusters are displayed.