Science
AAAS

Supplementary Materials for

Noninvasive assessment of gut function using transcriptional recording sentinel cells

Florian Schmidt *et al*.

Corresponding authors: Andrew J. Macpherson, andrew.macpherson@dbmr.unibe.ch; Randall J. Platt, rplatt@ethz.ch

Science **376**, eabm6038 (2022)
DOI: 10.1126/science.abm6038

The PDF file includes:

Figs. S1 to S12
Tables S22 to S25

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S21
MDAR Reproducibility Checklist

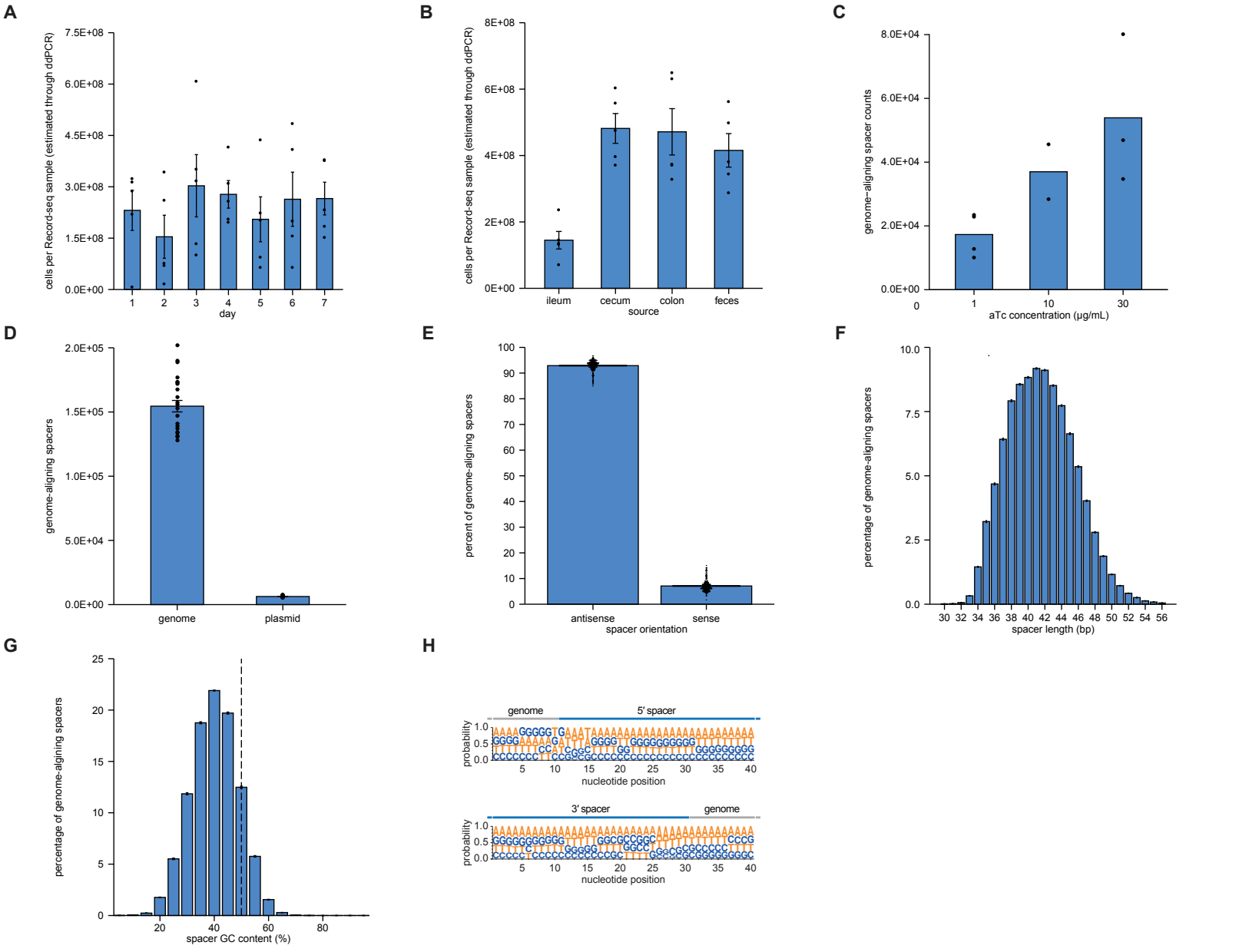

A
cells per Record-seq sample (estimated through ddPCR)
7.5E+08
6.0E+08
4.5E+08
3.0E+08
1.5E+08
0.0E+00
1
2
3
4
5
6
7
day
B
cells per Record-seq sample (estimated through ddPCR)
8E+08
6E+08
4E+08
2E+08
0E+00
ileum
cecum
colon
feces
source
C
genome−aligning spacer counts
8.0E+04
6.0E+04
4.0E+04
2.0E+04
0.0E+00
0
1
10
30
aTc concentration (µg/mL)
D
genome-aligning spacers
2.0E+05
1.5E+05
1.0E+05
5.0E+04
0.0E+00
genome
plasmid
E
percent of genome-aligning spacers
100
90
80
70
60
50
40
30
20
10
0
antisense
sense
spacer orientation
F
percentage of genome-aligning spacers
10.0
7.5
5.0
2.5
0.0
30 32 34 36 38 40 42 44 46 48 50 52 54 56
spacer length (bp)
G
percentage of genome-aligning spacers
25
20
15
10
5
0
20
40
60
80
spacer GC content (%)
H
genome
5′ spacer
probability
1.0
0.5
0.0
5 10 15 20 25 30 35 40
nucleotide position
3′ spacer
genome
probability
1.0
0.5
0.0
5 10 15 20 25 30 35 40
nucleotide position

Fig. S1. Transcriptional recording sentinel cells acquire transcriptional records within the mouse gut and preserve this information throughout time.

(A and B) Bar plots showing the cell number used per Record-seq input as estimated by droplet digital PCR (ddPCR) from (A) feces on the indicated days after gavage of *E. coli* sentinel cells and (B) different gut sections on day 20. The concentration of the recording plasmid (pFS_0453) was measured by ddPCR and the number of cells was calculated assuming 20 copies of pFS_0453 (pET30b+ origin of replication) per *E. coli* cell. Shown is the mean ± s.e.m. of n=5 independent biological replicates. (C) Bar plot showing the number of *E. coli* genome-aligning spacers obtained from colon or cecum contents of mice supplied with various concentrations of anhydrotetracycline (aTc) in the drinking water. Shown is the mean of n=2-4 independent biological replicates. (D) Bar plot showing the number of *E. coli* genome-aligning spacers and recording plasmid-derived spacers. Shown is the mean ± s.e.m. of n=20 independent biological replicates of chow-fed mice corresponding to a total of 3,249,165 spacers (3,123,056 genome-aligning, 126,109 plasmid-aligning). (E) Bar plot showing the percentage of spacers aligning to the sense or antisense strand of *E. coli* genes. (F and G) Histograms showing the (F) length and (G) GC content distribution of *E. coli* genome-aligning spacers. (H) Nucleotide probabilities (WebLogo) of the 5′ (top) and 3′ (bottom) end of the spacers, along with the corresponding sequence flanking the nucleotides in the *E. coli* genome. Spacer (blue) and flanking nucleotides (gray) are indicated. Nucleotide probabilities were computed from 10,000 genome-aligning spacers and their flanking regions. Panels A and B correspond to the chow samples on day 1 to 7 and 20, respectively, from **Fig. 1D**. Panels E, F, and G were computed using 19,479,559 spacers from n=270 samples corresponding to **Fig. 1D**.

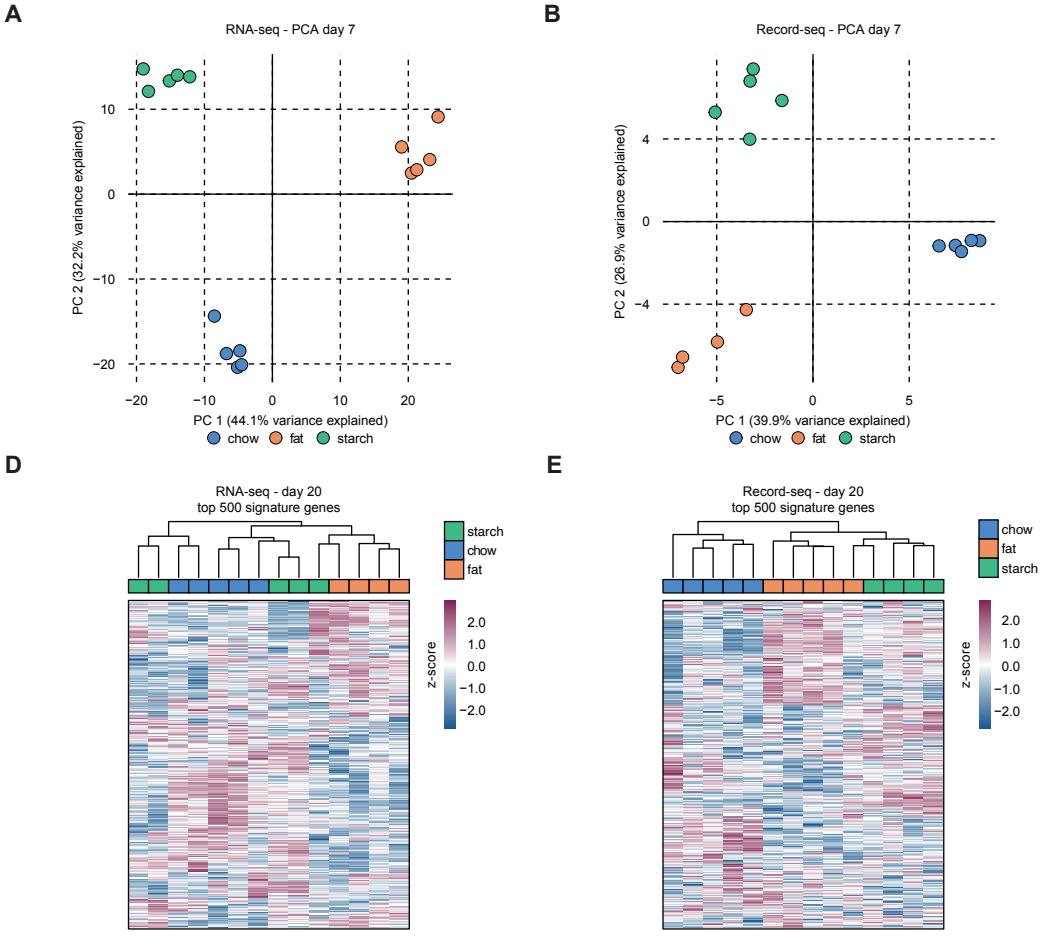

A
RNA-seq - PCA day 7
PC 2 (32.2% variance explained)
PC 1 (44.1% variance explained)
chow fat starch
B
Record-seq - PCA day 7
PC 2 (26.9% variance explained)
PC 1 (39.9% variance explained)
chow fat starch
C
Record-seq - UMAP day 2 to 20
UMAP 2
UMAP 1
day 2 20
chow fat starch
D
RNA-seq - day 20
top 500 signature genes
starch
chow
fat
z-score
E
Record-seq - day 20
top 500 signature genes
chow
fat
starch
z-score

Fig. S2. Record-seq reveals transcriptional changes describing the adaptation of *E. coli* to diet-dependent intraluminal environments.

(A and B) PCA-projected (A) RNA-seq and (B) Record-seq data from mice fed a chow (blue), fat (orange), or starch (green) diet on day 7. (C) UMAP embedding of Record-seq data from mice fed a chow (blue), fat (orange), or starch (green) diet on days 2 to 20. Dot sizes denote successive time points. (D and E) Heatmap showing hierarchical clustering of (D) RNA-seq and (E) Record-seq data on day 20 using the top 500 diet-specific signature genes identified prior to the diet switch on day 7. Z-score standardized gene-aligning spacer counts are shown. Panels A to E correspond to **Fig. 1D** with n=5 independent biological replicates for each diet. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

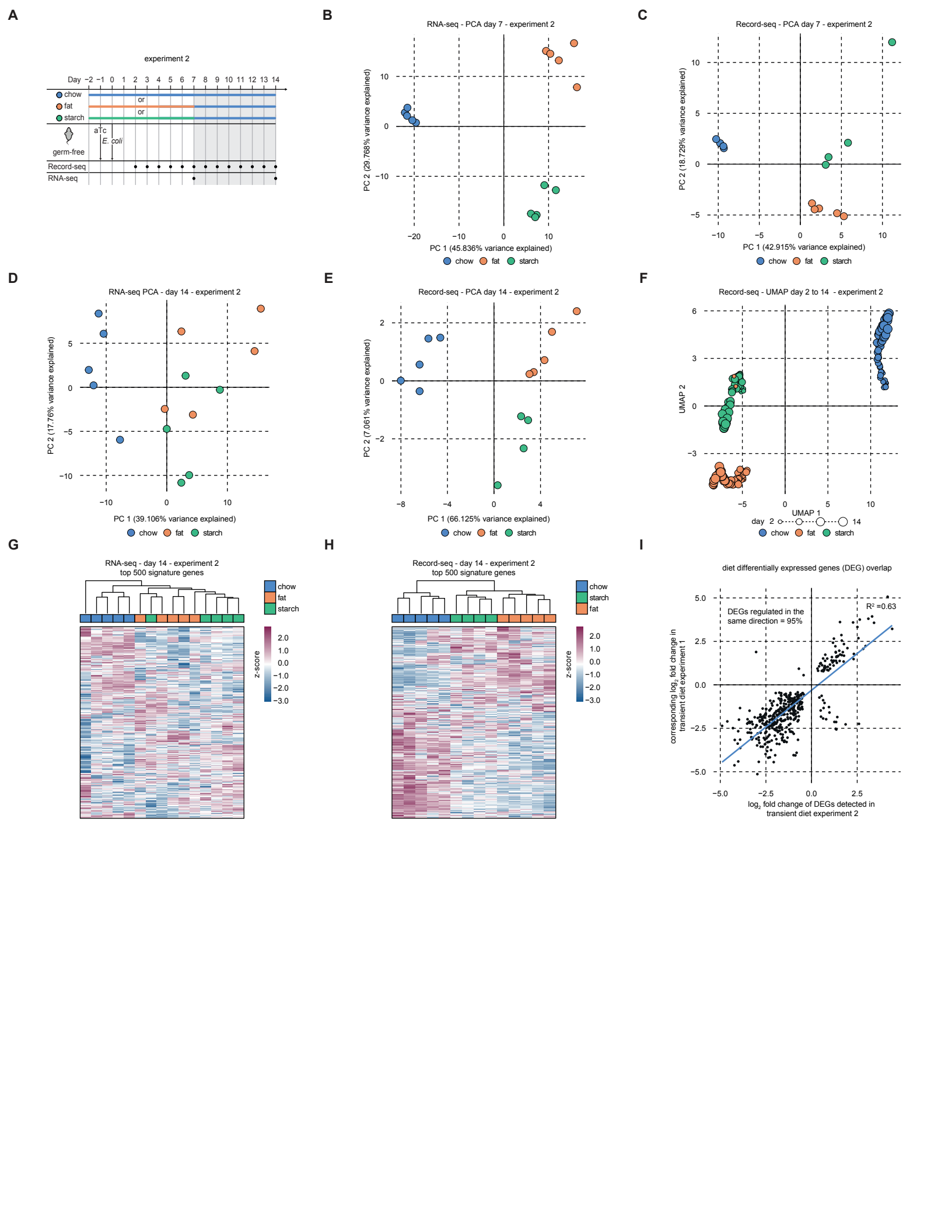

A
experiment 2
Day −2 −1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
chow
fat
starch
or
or
aTc
E. coli
germ-free
Record-seq
RNA-seq
B
RNA-seq - PCA day 7 - experiment 2
PC 2 (29.768% variance explained)
PC 1 (45.836% variance explained)
chow fat starch
C
Record-seq - PCA day 7 - experiment 2
PC 2 (18.729% variance explained)
PC 1 (42.915% variance explained)
chow fat starch
D
RNA-seq PCA - day 14 - experiment 2
PC 2 (17.76% variance explained)
PC 1 (39.106% variance explained)
chow fat starch
E
Record-seq - PCA day 14 - experiment 2
PC 2 (7.061% variance explained)
PC 1 (66.125% variance explained)
chow fat starch
F
Record-seq - UMAP day 2 to 14 - experiment 2
UMAP 2
UMAP 1
day 2 14
chow fat starch
G
RNA-seq - day 14 - experiment 2
top 500 signature genes
chow
fat
starch
z-score
H
Record-seq - day 14 - experiment 2
top 500 signature genes
chow
starch
fat
z-score
I
diet differentially expressed genes (DEG) overlap
DEGs regulated in the same direction = 95%
R² =0.63
corresponding log2 fold change in transient diet experiment 1
log2 fold change of DEGs detected in transient diet experiment 2

Fig. S3. Record-seq results are reproducible across independent experiments.

(A) Timeline of longitudinal in vivo recording experiment assessing the impact of diet on the *E. coli* transcriptome inside the gut. This was an independent replicate of the experiment in **Fig. 1D**. Germ-free mice were supplied with aTc in the drinking water and orally gavaged with *E. coli* sentinel cells. Mice were fed a chow, fat, or starch diet 2 days prior to gavage until day 7 of the experiment. From day 7 onwards, all groups received a chow diet. Fecal sampling for Record-seq and/or RNA-seq is indicated. (B and C) PCA-projected (B) RNA-seq or (C) Record-seq data on day 7. (D and E) PCA-projected (D) RNA-seq or (E) Record-seq data on day 14. (F) UMAP embedding of Record-seq data on days 2 to 14 from mice fed chow (blue), fat (orange), or starch (green) diet until day 7. Dot sizes denote successive time points. (G and H) Heatmap showing hierarchical clustering of (G) RNA-seq or (H) Record-seq data on day 14 using the top 500 diet-specific signature genes identified prior to the diet switch on day 7. Z-score standardized gene-aligning spacer counts are shown. (I) Scatter plot showing the correlation in log_2FC of DEGs and percentage of these genes regulated in the same direction for the two diet experiments outlined in **Fig. 1D** and **fig. S3A.** Genes detected as differentially expressed in Record-seq in chow versus starch groups on day 7 in the diet experiment outlined in **Fig. 1D** were used to perform this analysis. Panels B to H correspond to **fig. S3A** with n=5 independent biological replicates for each condition. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

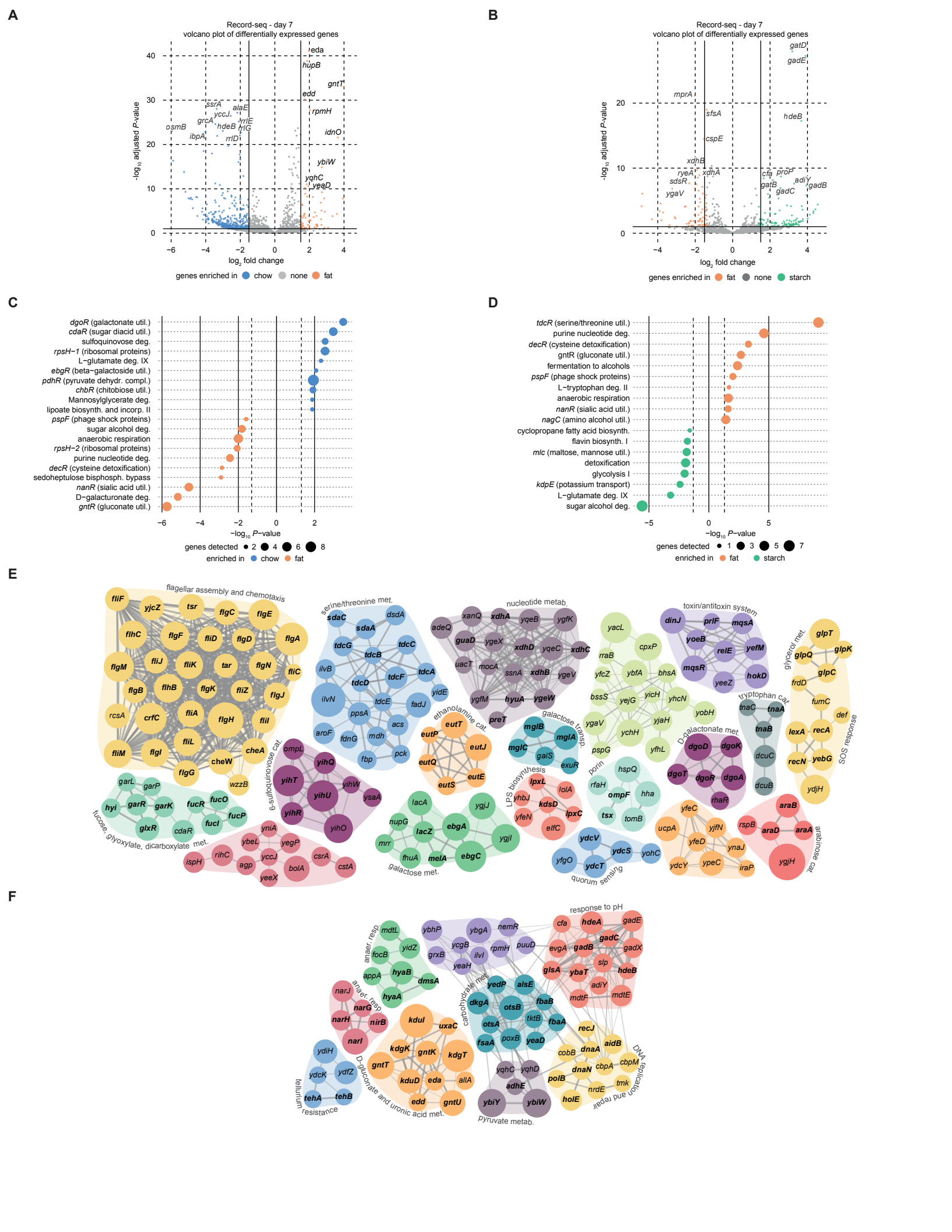

A
Record-seq - day 7
volcano plot of differentially expressed genes
eda
hupB
gntT
edd
rpmH
idnO
ybiW
yqhC
yeaD
ssrA
alaE
yccJ
grcA
hdeB
rrlE
rrlG
osmB
ibpA
rrlD
-log10 adjusted P-value
log2 fold change
genes enriched in chow none fat
B
Record-seq - day 7
volcano plot of differentially expressed genes
gatD
gadE
mprA
sfsA
hdeB
cspE
xdhB
ryeA
xdhA
sdsR
ygaV
cfa
proP
gatB
adiY
gadB
gadC
-log10 adjusted P-value
log2 fold change
genes enriched in fat none starch
C
dgoR (galactonate util.)
cdaR (sugar diacid util.)
sulfoquinovose deg.
rpsH−1 (ribosomal proteins)
L−glutamate deg. IX
ebgR (beta−galactoside util.)
pdhR (pyruvate dehydr. compl.)
chbR (chitobiose util.)
Mannosylglycerate deg.
lipoate biosynth. and incorp. II
pspF (phage shock proteins)
sugar alcohol deg.
anaerobic respiration
rpsH−2 (ribosomal proteins)
purine nucleotide deg.
decR (cysteine detoxification)
sedoheptulose bisphosph. bypass
nanR (sialic acid util.)
D−galacturonate deg.
gntR (gluconate util.)
−log10 P−value
genes detected 2 4 6 8
enriched in chow fat
D
tdcR (serine/threonine util.)
purine nucleotide deg.
decR (cysteine detoxification)
gntR (gluconate util.)
fermentation to alcohols
pspF (phage shock proteins)
L−tryptophan deg. II
anaerobic respiration
nanR (sialic acid util.)
nagC (amino alcohol util.)
cyclopropane fatty acid biosynth.
flavin biosynth. I
mlc (maltose, mannose util.)
detoxification
glycolysis I
kdpE (potassium transport)
L−glutamate deg. IX
sugar alcohol deg.
−log10 P−value
genes detected 1 3 5 7
enriched in fat starch
E
flagellar assembly and chemotaxis
serine/threonine met.
nucleotide metab.
toxin/antitoxin system
glycerol met.
tryptophan cat.
D-galactonate met.
SOS response
ethanolamine cat.
galactose transp.
LPS biosynthesis
porin
6-sulfoquinovose cat.
fucose, glyoxylate, dicarboxylate met.
galactose met.
quorum sensing
arabinose cat.
F
response to pH
anaer. resp.
carbohydrate met.
D-gluconate and uronic acid met.
pyruvate metab.
DNA replication and repair
tellurium resistance

Fig. S4. Record-seq reveals a wide-range of genes and pathways orchestrating the adaptation of *E. coli* to diet-dependent intraluminal environments.

(A and B) Volcano plots showing Record-seq differentially expressed genes (DEGs) from mice fed a (A) chow (blue) or fat (orange) diet or a (B) fat (orange) or starch (green) diet on day 7 as shown in **Fig. 1D** scheme *(*P_{adj}<0.1; $\log_2$-fold change>1.5). (C and D) Pathways and transcriptional/translational regulators identified as enriched *(P*<0.05) using EcoCyc based on Record-seq data on day 7 from mice fed a (C) chow (blue) or fat (orange) diet or a (D) fat (orange) or starch (green) diet. Dot sizes show gene numbers detected as significantly upregulated for the respective pathway. (E and F) STRING analysis of genes significantly upregulated in *E. coli* from mice fed a (E) chow or (F) starch diet. Node size corresponds to log_2FC of upregulation (E: 1.0-5.0, F: 1.0-4.4). Panels A-F correspond to **Fig. 1D** with n=5 independent biological replicates. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

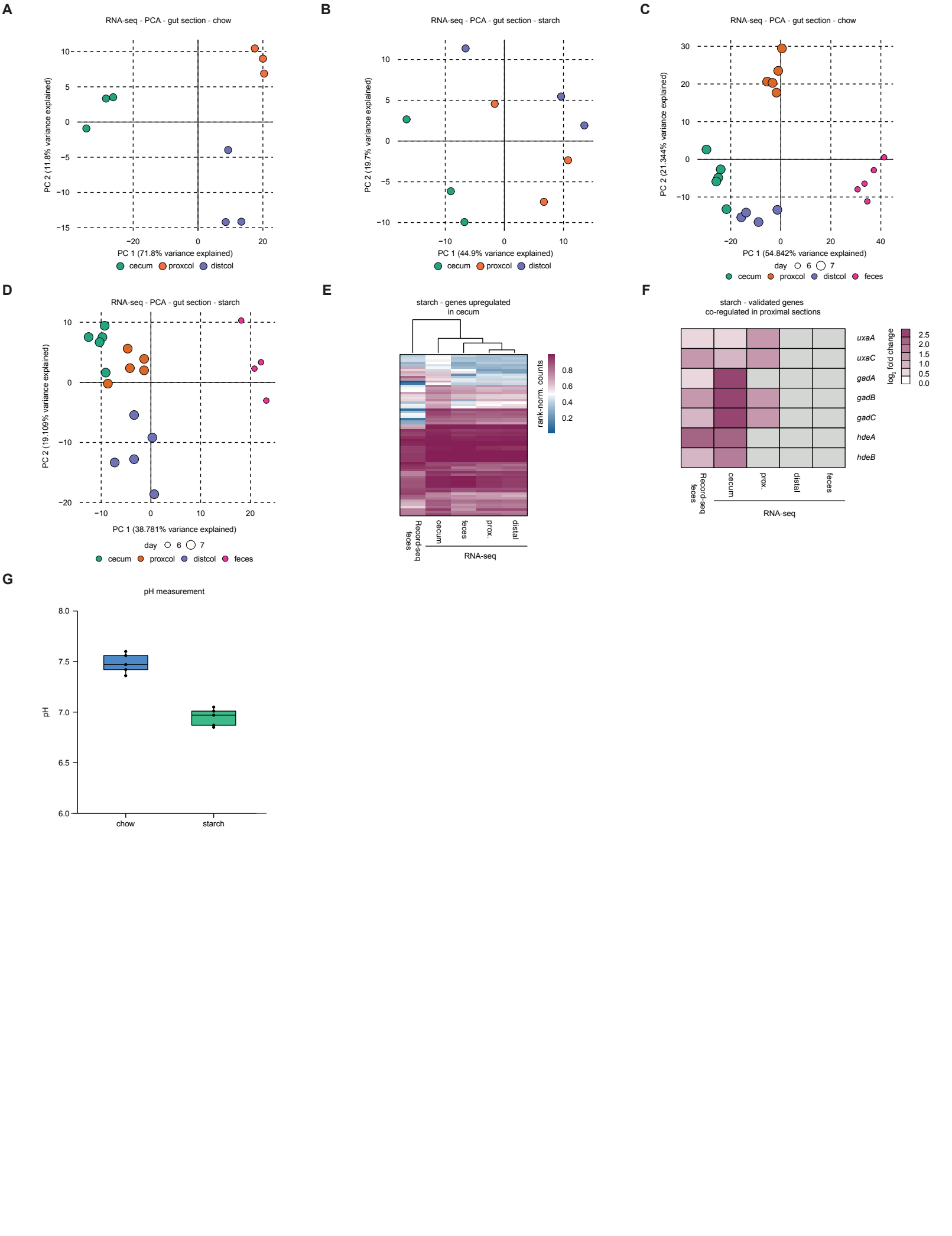

A
RNA-seq - PCA - gut section - chow
PC 2 (11.8% variance explained)
PC 1 (71.8% variance explained)
cecum proxcol distcol
B
RNA-seq - PCA - gut section - starch
PC 2 (19.7% variance explained)
PC 1 (44.9% variance explained)
cecum proxcol distcol
C
RNA-seq - PCA - gut section - chow
PC 2 (21.344% variance explained)
PC 1 (54.842% variance explained)
day 6 7
cecum proxcol distcol feces
D
RNA-seq - PCA - gut section - starch
PC 2 (19.109% variance explained)
PC 1 (38.781% variance explained)
day 6 7
cecum proxcol distcol feces
E
starch - genes upregulated in cecum
rank-norm. counts
Record-seq feces
cecum
feces
prox.
distal
RNA-seq
F
starch - validated genes co-regulated in proximal sections
log2 fold change
uxaA
uxaC
gadA
gadB
gadC
hdeA
hdeB
Record-seq feces
cecum
prox.
distal
feces
RNA-seq
G
pH measurement
pH
chow
starch

Fig. S5. Record-seq sentinel cells capture the milieu of proximal gut sections in a non-invasive fashion.

(A and B) PCA-projected *E. coli* RNA-seq data from cecum (green), proximal colon (orange), and distal colon (purple) of mice fed a (A) chow or (B) starch diet on day 7. (C and D) PCA-projected *E. coli* RNA-seq data from cecum (green), proximal colon (orange), distal colon (purple), and feces (pink) of mice fed a (C) chow or (D) starch diet on the indicated days. (E) Heatmap of cecum signature genes (213 genes overexpressed in the cecum) showing hierarchical clustering of rank-normalized RNA-seq and Record-seq data from the indicated intestinal sections from mice fed a starch diet. (F) Heatmap showing log_2FC as determined by Record-seq or RNA-seq from feces, cecum, proximal colon, and distal colon for genes that were experimentally validated (*uxaAC* in **Fig. 3D**, *gadABC*, *hdeAB* in **fig. S5G**) as a subset of genes identified as differentially regulated in the chow and starch diet groups by fecal Record-seq but not fecal RNA-seq. Grey boxes indicate no significant differential regulation. (G) Box plot showing cecal luminal pH under a chow or starch diet. Representative result from two independent experiments of n=5 mice per group. P=1.975·10^{-5} (T-test). Panels A, B, E and F with n=3 independent biological replicates each pooled from n=3 individual mice. Panels C and D with n=5 independent biological replicates each from an individual mouse. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

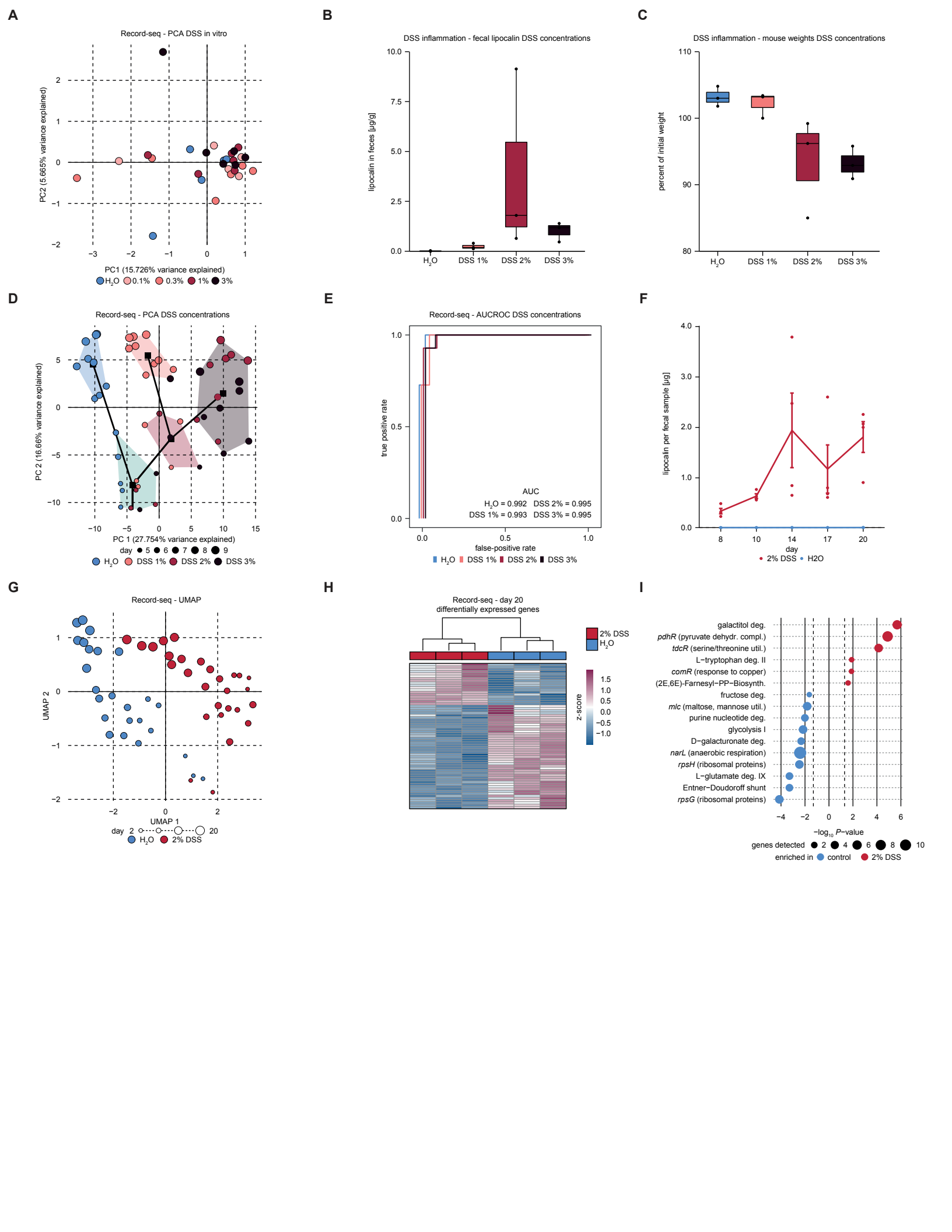

A
Record-seq - PCA DSS in vitro
PC2 (5.665% variance explained)
PC1 (15.726% variance explained)
H2O 0.1% 0.3% 1% 3%
B
DSS inflammation - fecal lipocalin DSS concentrations
lipocalin in feces [µg/g]
H2O DSS 1% DSS 2% DSS 3%
C
DSS inflammation - mouse weights DSS concentrations
percent of initial weight
H2O DSS 1% DSS 2% DSS 3%
D
Record-seq - PCA DSS concentrations
PC 2 (16.66% variance explained)
PC 1 (27.754% variance explained)
day 5 6 7 8 9
H2O DSS 1% DSS 2% DSS 3%
E
Record-seq - AUCROC DSS concentrations
true positive rate
false-positive rate
AUC
H2O = 0.992 DSS 2% = 0.995
DSS 1% = 0.993 DSS 3% = 0.995
H2O DSS 1% DSS 2% DSS 3%
F
lipocalin per fecal sample [µg]
day
2% DSS H2O
G
Record-seq - UMAP
UMAP 2
UMAP 1
day 2 20
H2O 2% DSS
H
Record-seq - day 20
differentially expressed genes
2% DSS
H2O
z-score
I
galactitol deg.
pdhR (pyruvate dehydr. compl.)
tdcR (serine/threonine util.)
L−tryptophan deg. II
comR (response to copper)
(2E,6E)-Farnesyl−PP−Biosynth.
fructose deg.
mlc (maltose, mannose util.)
purine nucleotide deg.
glycolysis I
D−galacturonate deg.
narL (anaerobic respiration)
rpsH (ribosomal proteins)
L−glutamate deg. IX
Entner−Doudoroff shunt
rpsG (ribosomal proteins)
−log10 P−value
genes detected 2 4 6 8 10
enriched in control 2% DSS

Fig. S6. Record-seq provides a non-invasive assessment of DSS-induced colitis.

(A) PCA-projected of Record-seq data of *E. coli* exposed in vitro to 0 (blue), 0.1% (light pink), 0.3% (salmon), 1% (red), or 3% (black) dextran sulfate sodium (DSS), n=5 independent biological replicates. (B and C) Box plots showing (B) fecal lipocalin levels and (C) percent of initial weight on day 10 in control mice (blue) or mice treated with 1% (salmon), 2% (red), or 3% (black) DSS, n=3 for each condition (D) PCA-projected trajectory plot of Record-seq data from control mice (blue) and mice treated with 1% (salmon), 2% (red), or 3% (black) DSS. Convex hulls represent k-medoids clusters (see **Methods**). Dot sizes denote successive time points. (E) Area under the receiver operating characteristic curve (AUCROC) for evaluating the performance of multi-class SVM classifiers for distinguishing Record-seq samples based on DSS treatment groups. (F) Line plot showing fecal lipocalin levels from control mice (blue) or mice treated with 2% DSS (red), shown is mean ± s.e.m. (G) UMAP embedding of Record-seq data from control mice (blue) or mice treated with 2% DSS (red). Dot sizes denote successive time points from day 2 to 20. (H) Heatmap showing hierarchical clustering of Record-seq data from control mice (blue) or mice treated with 2% DSS (red), using differentially expressed genes identified on day 20. Z-score standardized gene-aligning spacer counts are shown. (I) Pathways and transcriptional/translational regulators identified as enriched *(P*<0.05) using EcoCyc based on Record-seq data on days 2-20 for control mice (blue) or mice treated with 2% DSS (red). Dot size increases with number of significantly upregulated genes for the respective pathway. Panels B to E correspond to **Fig. 4A** with n=3 independent biological replicates. Panels F to I correspond to **Fig. 4D** with n=3-4 independent biological replicates. Count thresholds were 10^4 (panels A and G to I). $5{\cdot}10^3$ (panels D and E). Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

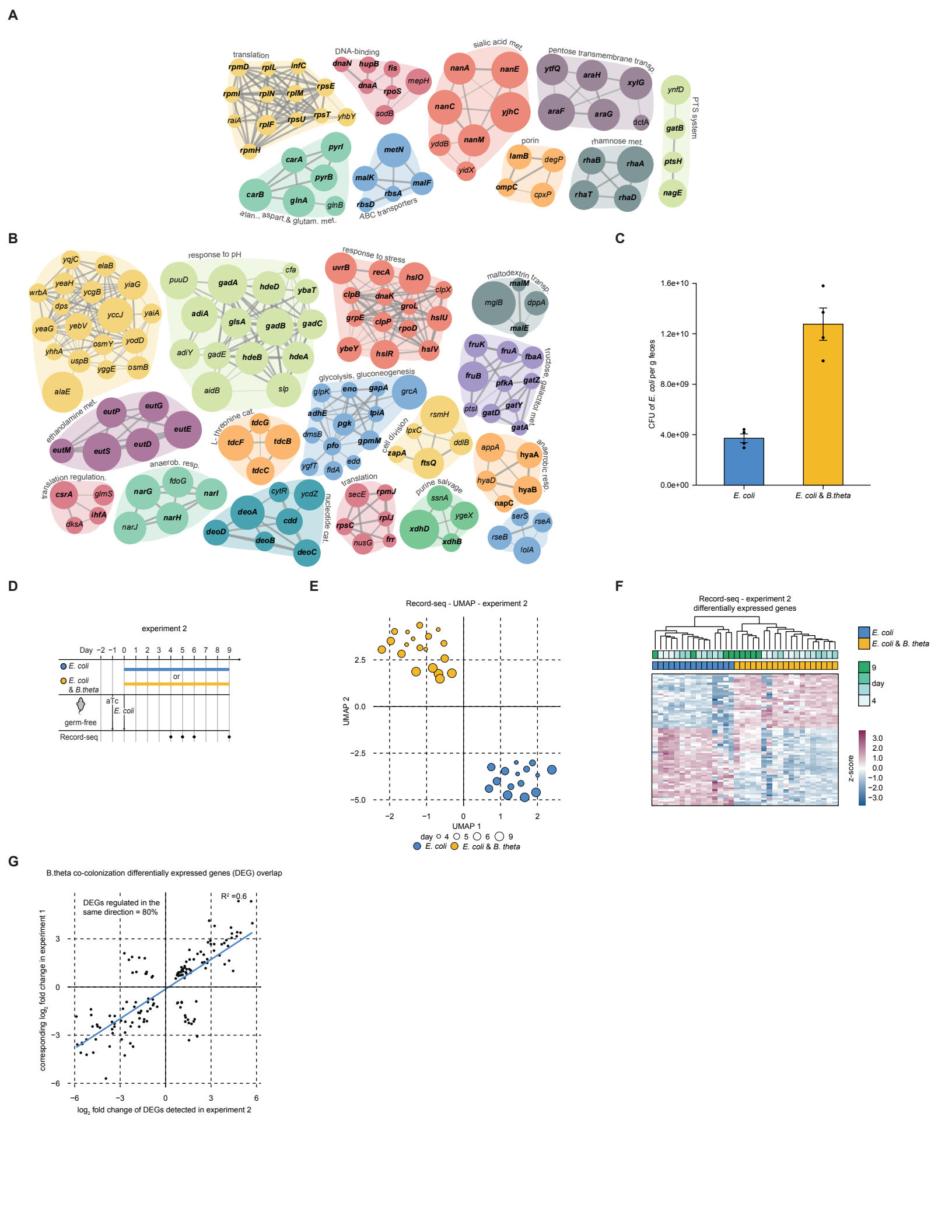

A
translation
rpmD rplL infC rpmI rplN rplM rpsE raiA rplF rpsU rpsT yhbY rpmH
DNA-binding
dnaN hupB fis dnaA rpoS mephH sodB
sialic acid met.
nanA nanE nanC yjhC yddB nanM yidX
pentose transmembrane transp.
ytfQ araH xylG araF araG dctA
ynfD gatB ptsH nagE
PTS system
porin
lamB degP ompC cpxP
rhamnose met.
rhaB rhaA rhaT rhaD
carA pyrI pyrB carB glnA glnB
alan., aspart.& glutam. met.
metN malK malF rbsA rbsD
ABC transporters
B
response to pH
yqjC elaB wrbA yeaH ycgB yiaG dps yccJ yaiA yeaG yebV osmY yodD yhhA uspB yggE osmB alaE
puuD gadA hdeD cfa ybaT adiA glsA gadB gadC adiY gadE hdeB hdeA aidB slp
response to stress
uvrB recA hslO clpX clpB dnaK groL grpE clpP hslU rpoD ybeY hslR hslV
maltodextrin transp.
malM mglB dppA malE
fruK fruA fbaA fruB pfkA gatZ ptsI gatD gatY gatA
fructose galactitol met.
glycolysis, gluconeogenesis
glpK eno gapA grcA adhE pgk tpiA dmsB pfo gpmM ygfT fldA edd
ethanolamine met.
eutP eutG eutE eutM eutS eutD
L- threonine cat.
tdcG tdcF tdcB tdcC
cell division
rsmH ipxC ddlB zapA ftsQ
anaerob. resp.
fdoG narG narI narJ narH
translation regulation.
csrA glmS dksA ihfA
translation
secE rpmJ rplJ rpsC nusG frr
purine salvage
ssnA ygeX xdhD xdhB
anaerobic resp.
appA hyaA hyaD hyaB napC
serS rseA rseB lolA
cytR ycdZ deoA cdd deoD deoB deoC
nucleotide cat.
C
CFU of E. coli per g feces
1.6e+10
1.2e+10
8.0e+09
4.0e+09
0.0e+00
E. coli
E. coli & B.theta
D
experiment 2
Day −2 −1 0 1 2 3 4 5 6 7 8 9
E. coli
E. coli & B.theta
or
aTc
E. coli
germ-free
Record-seq
E
Record-seq - UMAP - experiment 2
UMAP 2
2.5
0.0
−2.5
−5.0
−2 −1 0 1 2
UMAP 1
day 4 5 6 9
E. coli
E. coli & B. theta
F
Record-seq - experiment 2
differentially expressed genes
E. coli
E. coli & B. theta
9
day
4
z-score
3.0
2.0
1.0
0.0
−1.0
−2.0
−3.0
G
B.theta co-colonization differentially expressed genes (DEG) overlap
DEGs regulated in the same direction = 80%
R² =0.6
corresponding log2 fold change in experiment 1
3
0
−3
−6
−6 −3 0 3 6
log2 fold change of DEGs detected in experiment 2

Fig. S7. Record-seq illuminates both host-microbe and microbe-microbe interactions.

(A and B) STRING analysis of genes significantly (A) upregulated or (B) downregulated in *E. coli* in the presence of *B. theta* compared to *E. coli* in monocolonized mice. Node size indicates log_2FC (panel A: 0.2-5.7, panel B: 0.2-6.4). (C) Bar plot showing *E. coli* colony forming unit (CFU) counts per g of feces on day 7, shown is mean ± s.e.m., *P*=0.02857 (Wilcoxon rank sum test). (D) Timeline of longitudinal in vivo recording experiment for illuminating the interaction of *E. coli* with *B. theta* in the mouse gut. Germ-free mice were supplied with aTc in the drinking water and orally gavaged with *E. coli* sentinel cells alone or together with *B. theta*. Fecal Record-seq sampling is indicated. (E) UMAP embedding of Record-seq data from *E. coli* in the presence (yellow) or absence (blue) of *B. theta* on days 4 to 9. Dot sizes denote successive time points. (F) Heatmap showing hierarchical clustering of Record-seq data from *E. coli* in the presence (yellow) or absence (blue) of *B. theta* on indicated days using the identified differentially expressed genes (DEGs). Z-score standardized gene-aligning spacer counts are shown. (G) Scatter plot showing the correlation in log_2FC of DEGs and percentage of these genes regulated in the same direction for the two experiments outlined in **Fig. 5A** and **fig. S7D.** Genes detected as differentially expressed for Record-seq from *E. coli* in the presence (yellow) or absence (blue) of *B. theta* in the experiment outlined in **Fig. 5A** were used to perform this analysis. Panels A to C correspond to the experiment outlined in **Fig. 5A** with n=4 independent biological replicates. Panels E and F correspond to the experiment outlined in **Fig. S7D** with n=4-5 independent biological replicates. Count threshold was $5 \cdot 10^3$. Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

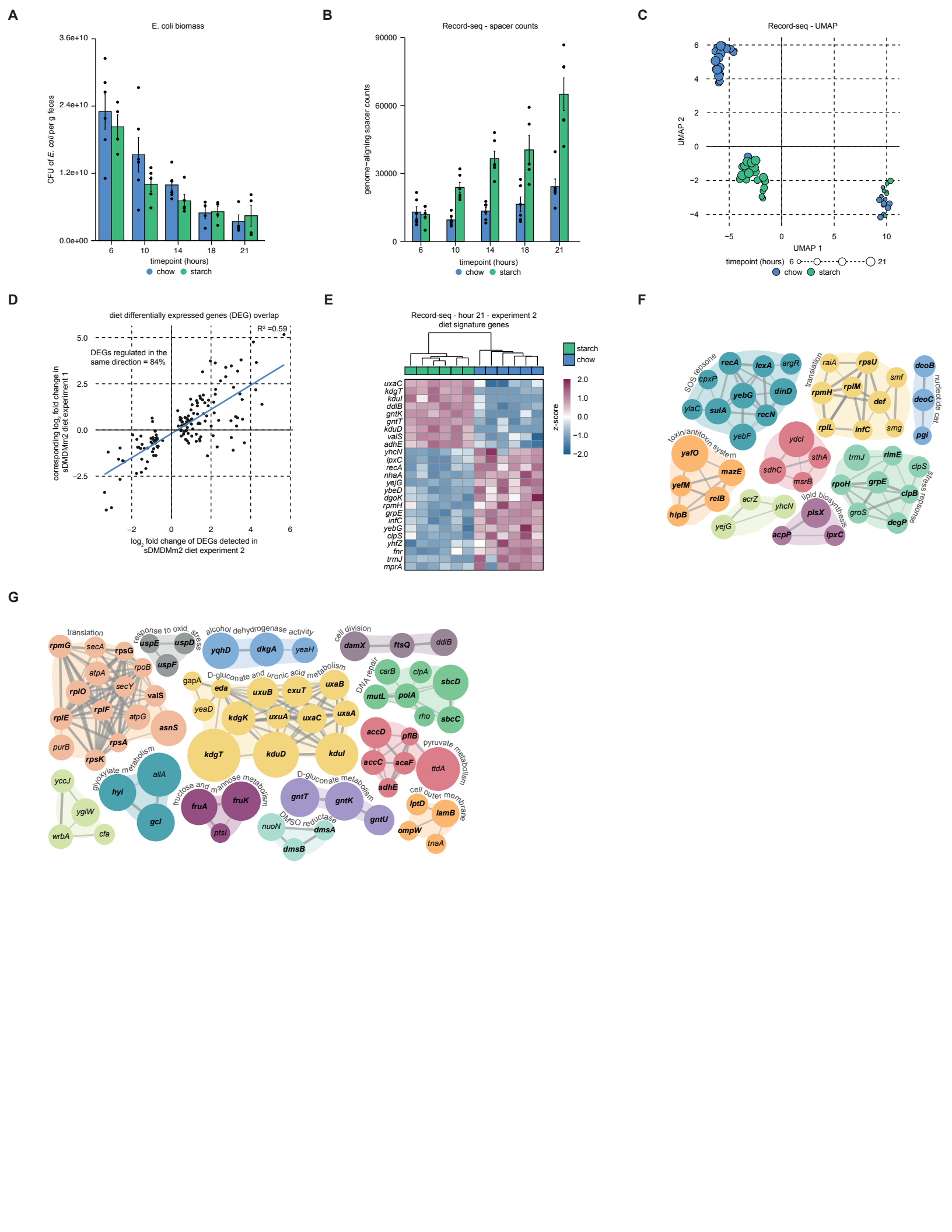

A
E. coli biomass
CFU of E. coli per g feces
3.6e+10
2.4e+10
1.2e+10
0.0e+00
6
10
14
18
21
timepoint (hours)
chow
starch
B
Record-seq - spacer counts
genome−aligning spacer counts
90000
60000
30000
0
6
10
14
18
21
timepoint (hours)
chow
starch
C
Record-seq - UMAP
UMAP 2
6
4
2
0
−2
−4
−5
0
5
10
UMAP 1
timepoint (hours) 6
21
chow
starch
D
diet differentially expressed genes (DEG) overlap
R² =0.59
DEGs regulated in the same direction = 84%
corresponding log2 fold change in sDMDMm2 diet experiment 1
5.0
2.5
0.0
−2.5
−2
0
2
4
6
log2 fold change of DEGs detected in sDMDMm2 diet experiment 2
E
Record-seq - hour 21 - experiment 2
diet signature genes
starch
chow
z-score
2.0
1.0
0.0
−1.0
−2.0
uxaC
kdgT
kduI
ddlB
gntK
gntT
kduD
valS
adhE
yhcN
lpxC
recA
nhaA
yejG
ybeD
dgoK
rpmH
grpE
infC
yebG
clpS
yhfZ
fnr
trmJ
mprA
F
SOS repsone
recA
lexA
argR
cpxP
yebG
dinD
ylaC
sulA
recN
yebF
translation
raiA
rpsU
smf
rpmH
rplM
def
rplL
infC
smg
nucleotide cat.
deoB
deoC
pgi
toxin/antitoxin system
yafO
ydcI
sthA
mazE
sdhC
msrB
yefM
relB
acrZ
yhcN
hipB
yejG
trmJ
rlmE
clpS
rpoH
grpE
clpB
groS
degP
stress response
lipid biosynthesis
plsX
acpP
lpxC
G
translation
rpmG
secA
rpsG
rpoB
atpA
secY
rplO
valS
rplF
rplE
atpG
asnS
purB
rpsA
rpsK
response to oxid. stress
uspE
uspD
uspF
alcohol dehydrogenase activity
yqhD
dkgA
yeaH
cell division
damX
ftsQ
ddlB
D-gluconate and uronic acid metabolism
gapA
eda
yeaD
uxuB
exuT
uxaB
kdgK
uxuA
uxaC
uxaA
kdgT
kduD
kduI
DNA repair
carB
clpA
sbcD
mutL
polA
rho
sbcC
accD
pflB
accC
aceF
ttdA
adhE
pyruvate metabolism
glyoxylate metabolism
allA
hyi
yccJ
ygiW
wrbA
cfa
gcl
fructose and mannose metabolism
fruA
fruK
ptsI
D-gluconate metabolism
gntT
gntK
gntU
cell outer membrane
lptD
lamB
ompW
tnaA
DMSO reductase
nuoN
dmsA
dmsB

Fig. S8. Sentinel cells are deployable within a complex microbiota.

(A) Bar plot showing *E. coli* colony forming unit (CFU) counts per g of feces at the indicated timepoints. (B) Bar plot showing the number of *E. coli* genome-aligning spacers obtained per Record-seq sample from the feces of sDMDMm2 mice at the indicated timepoints after gavage of $7 \cdot 10^{10}$ *E. coli* sentinel cells corresponding to **Fig. 6A**. Shown is the mean ± s.e.m. (C) UMAP embedding of Record-seq data from mice fed a chow (blue), or starch (green) diet at 6, 10, 14, 18, and 21 hours. Dot sizes denote successive points. (D) Scatter plot showing the correlation in log_2FC of DEGs and percentage of these genes regulated in the same direction for the two experiments outlined in **Fig. 6A** and an independent replicate experiment using a gavage dose of $6 \cdot 10^{10}$ *E. coli* sentinel cells. Genes detected as differentially expressed for Record-seq from *E. coli* in the in sDMDMm2 mice on the chow or starch diet in the experiment outlined in **Fig. 6A** were used to perform this analysis. (E) Heatmap showing hierarchical clustering of Record-seq data at 21 hours using identified differentially expressed genes (DEGs). Z-score standardized gene-aligning spacer counts are shown. (F and G) STRING analysis of genes significantly upregulated in *E. coli* under a (F) chow or (G) starch diet in sDMDMm2 mice. Node size indicates log_2FC (F: 0.4-3.2, G: 0.3-5.7). Panels A to G correspond to the experiment outlined in **Fig. 6A**, n=6 independent biological replicates. Panel D additionally uses data from an independent experiment with n=5 biological replicates. Count threshold was $5 \cdot 10^3$. Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

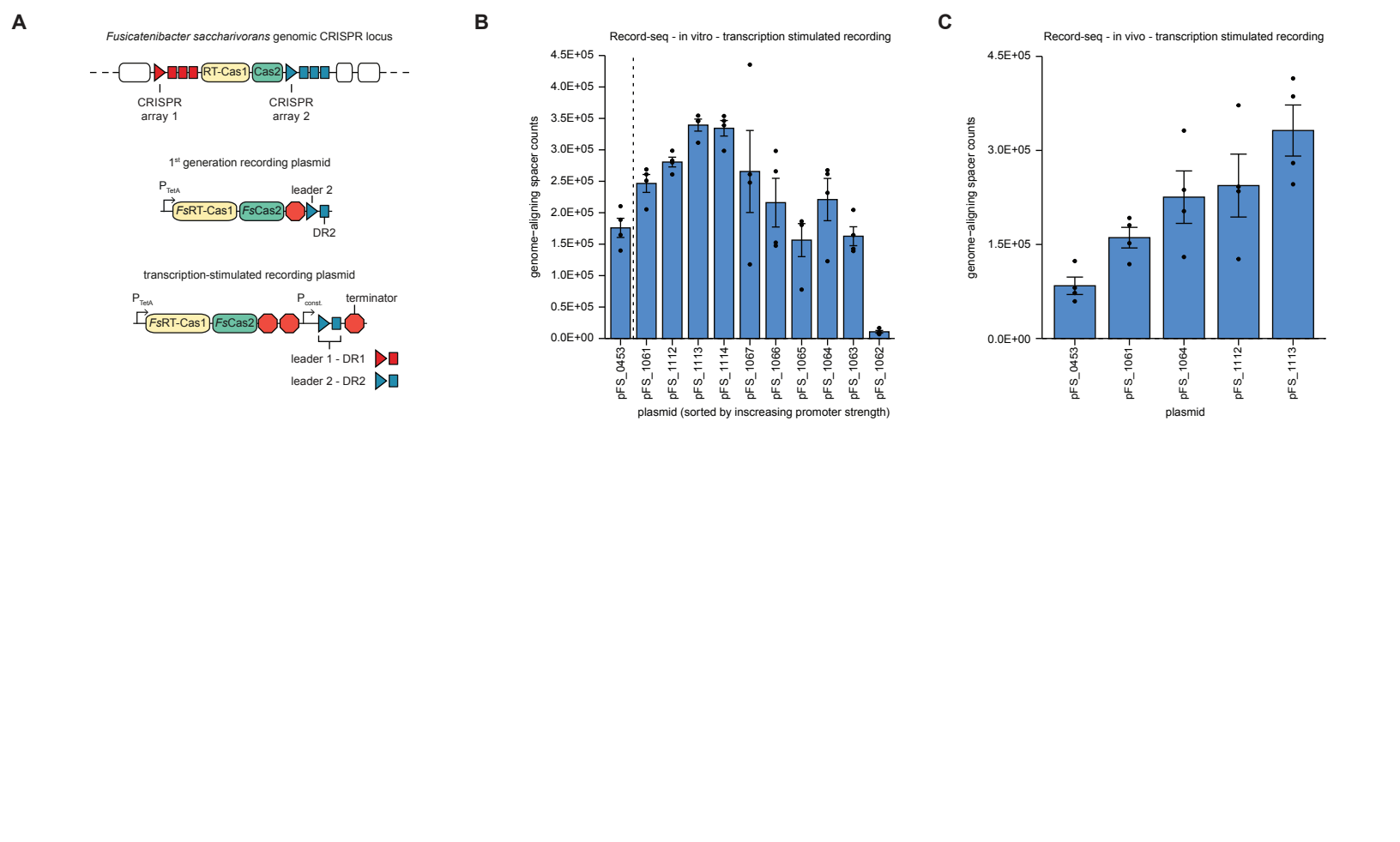
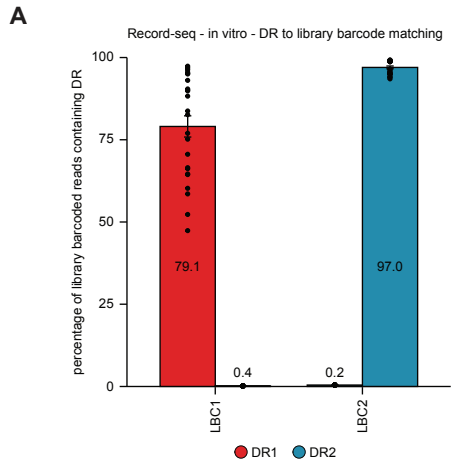
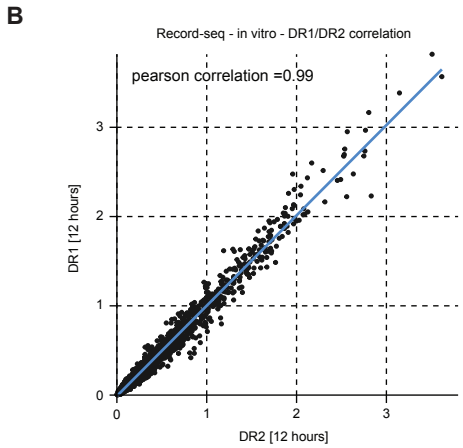
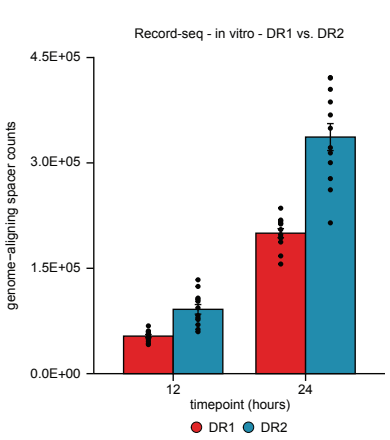

A
Fusicatenibacter saccharivorans genomic CRISPR locus
RT-Cas1
Cas2
CRISPR array 1
CRISPR array 2
1st generation recording plasmid
P TetA
leader 2
FsRT-Cas1
FsCas2
DR2
transcription-stimulated recording plasmid
P TetA
P const.
terminator
FsRT-Cas1
FsCas2
leader 1 - DR1
leader 2 - DR2
B
Record-seq - in vitro - transcription stimulated recording
genome−aligning spacer counts
4.5E+05
4.0E+05
3.5E+05
3.0E+05
2.5E+05
2.0E+05
1.5E+05
1.0E+05
0.5E+05
0.0E+00
pFS_0453
pFS_1061
pFS_1112
pFS_1113
pFS_1114
pFS_1067
pFS_1066
pFS_1065
pFS_1064
pFS_1063
pFS_1062
plasmid (sorted by inscreasing promoter strength)
C
Record-seq - in vivo - transcription stimulated recording
genome−aligning spacer counts
4.5E+05
3.0E+05
1.5E+05
0.0E+00
pFS_0453
pFS_1061
pFS_1064
pFS_1112
pFS_1113
plasmid

Fig. S9. Active transcription of the CRISPR array improves spacer acquisition.

(A) Schematic illustrating: (top) the genomic CRISPR locus of *Fusicatenibacter saccharivorans* (*Fs*), which encodes two CRISPR arrays (CRISPR array 1 and CRISPR array 2) with different leader and direct repeat (DR) sequences; (middle) the first generation recording plasmid encoding *Fs*RT-Cas1–Cas2 under transcriptional control of an anhydrotetracycline (aTc)-inducible promoter and a single terminator upstream of leader-DR2; and (bottom) the transcription-stimulated recording plasmid construct design. A double terminator downstream of *Fs*RT-Cas1–Cas2 minimizes transcriptional readthrough from the P_{TetA} promoter whereas a constitutive promoter upstream of the leader-DR1 or leader-DR2 results in active transcription of the CRISPR array at a strength depending on the chosen promoter. (B) Bar plot showing the number of *E. coli* genome-aligning spacers obtained per Record-seq in vitro sample from *E. coli* cells transformed with the indicated transcriptional recording plasmids employing constitutive *E. coli* promoters from the Anderson promoter library upstream of the CRISPR array (**fig. S9A and Methods**) for DR1 and DR2. (C) Bar plot showing the number of *E. coli* genome-aligning spacers obtained per Record-seq in vivo sample from *E. coli* cells transformed with the indicated transcriptional recording plasmids. Panel B corresponds to an in vitro experiment with n=4 independent biological replicates. Panel C corresponds to an in vivo experiment with n=4 independent biological replicates.


A
Record-seq - in vitro - DR to library barcode matching
percentage of library barcoded reads containing DR
100
75
50
25
0
79.1
0.4
0.2
97.0
LBC1
LBC2
DR1
DR2


B
Record-seq - in vitro - DR1/DR2 correlation
pearson correlation =0.99
DR1 [12 hours]
DR2 [12 hours]
0
1
2
3


C
Record-seq - in vitro - DR1 vs. DR2
genome-aligning spacer counts
4.5E+05
3.0E+05
1.5E+05
0.0E+00
12
24
timepoint (hours)
DR1
DR2

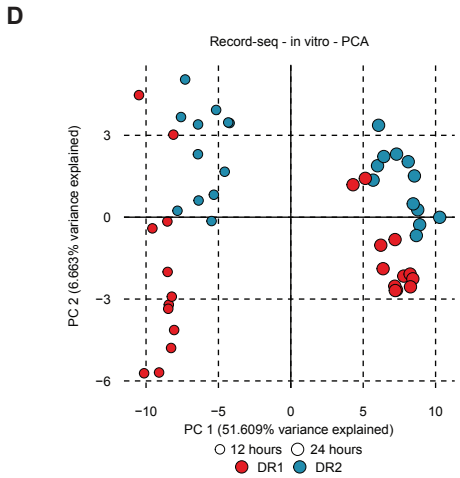

D
Record-seq - in vitro - PCA
PC 2 (6.663% variance explained)
3
0
−3
−6
−10
−5
0
5
10
PC 1 (51.609% variance explained)
12 hours
24 hours
DR1
DR2

Fig. S10. Barcoded CRISPR arrays enable multiplexed Record-seq in vitro.

(A) Bar plot showing the number of reads correctly or erroneously assigned to the DR based on the library barcode (LBC) attached during the adapter ligation procedure in SENECA. Shown is the mean ± s.e.m., n=12 independent biological replicates. (B) Scatter plot showing the correlation between mean normalized gene-aligning spacer-counts for Record-seq in vitro samples from *E. coli* cells transformed with transcriptional recording plasmid encoding *Fs*Leader1-DR1 (pFS_1142) or *Fs*Leader2-DR2 (pFS_1113) from hour 12. Shown is the mean of n=12 independent biological replicates. (C) Bar plot showing the number of *E. coli* genome-aligning spacers obtained per Record-seq in vitro sample from *E. coli* cells transformed with a transcriptional recording plasmid encoding *Fs*Leader1-DR1 (pFS_1142) or *Fs*Leader2-DR2 (pFS_1113). Shown is the mean ± s.e.m. of 12 independent biological replicates. Samples from the 12 and 24-hour timepoints are matched (two timepoints obtained from the same culture). (D) PCA-projected Record-seq in vitro data from *E. coli* cells transformed with a transcriptional recording plasmid encoding *Fs*Leader1-DR1 (pFS_1142) or *Fs*Leader2-DR2 (pFS_1113) from hour 12 and 24. Dot sizes denote successive time points. Panels A to D correspond to an in vitro experiment with n=12 independent biological replicates. Count threshold was $3 \cdot 10^4$. Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

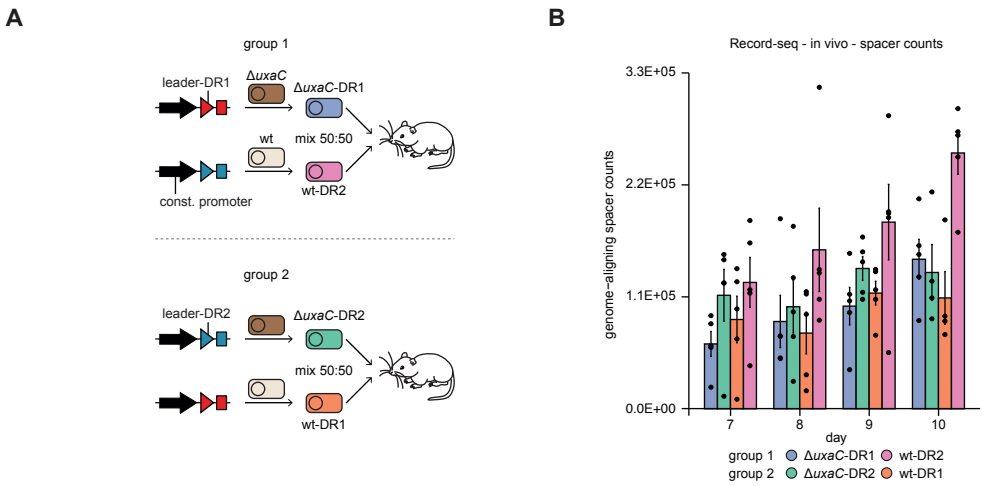

A
group 1
leader-DR1
ΔuxaC
ΔuxaC-DR1
wt
mix 50:50
wt-DR2
const. promoter
group 2
leader-DR2
ΔuxaC-DR2
mix 50:50
wt-DR1
B
Record-seq - in vivo - spacer counts
genome-aligning spacer counts
3.3E+05
2.2E+05
1.1E+05
0.0E+00
7
8
9
10
day
group 1 ΔuxaC-DR1 wt-DR2
group 2 ΔuxaC-DR2 wt-DR1

Fig. S11. Barcoded CRISPR arrays enable multiplexed Record-seq in vivo.

(A) Schematic illustrating full factorial design for multiplexed recording experiment with two experimental groups. Either wild-type (wt) *E. coli* transformed with leader-DR2 recording plasmid (blue) and Δ*uxaC E. coli* transformed with leader-DR1 recording plasmid (pink) are mixed (group 1) or wt *E. coli* transformed with leader-DR1 recording plasmid (green) and ΔuxaC *E. coli* transformed with leader-DR2 recording plasmid (green) are mixed and gavaged into germ-free mice. Since a stretch of sequence that is distinct between the DRs of these two CRISPR arrays is maintained throughout the library preparation procedure, this sequence could serve as a barcode and enable us to computationally discriminate spacers acquired into the two CRISPR arrays. We had previously demonstrated that both *Fs*CRISPR array-1 and array-2 were capable of spacer acquisition in an *E. coli* host (*17*). (B) Bar plot showing the number of *E. coli* genome-aligning spacers obtained per fecal Record-seq sample on the indicated days after gavage from experimental group 1 – gavaged with ΔuxaC *E. coli* harboring DR1 recording plasmid (blue) and wt *E. coli* cells harboring DR2 recording plasmid (pink) in the same mouse, or experimental group 2 – gavaged Δ*uxaC E. coli* harboring DR2 recording plasmid (green) and wt *E. coli* cells harboring DR1 recording plasmid (orange) in the same mouse. Panels A and B correspond to the experiment outlined in **Fig. 7A** with n=5 independent biological replicates.

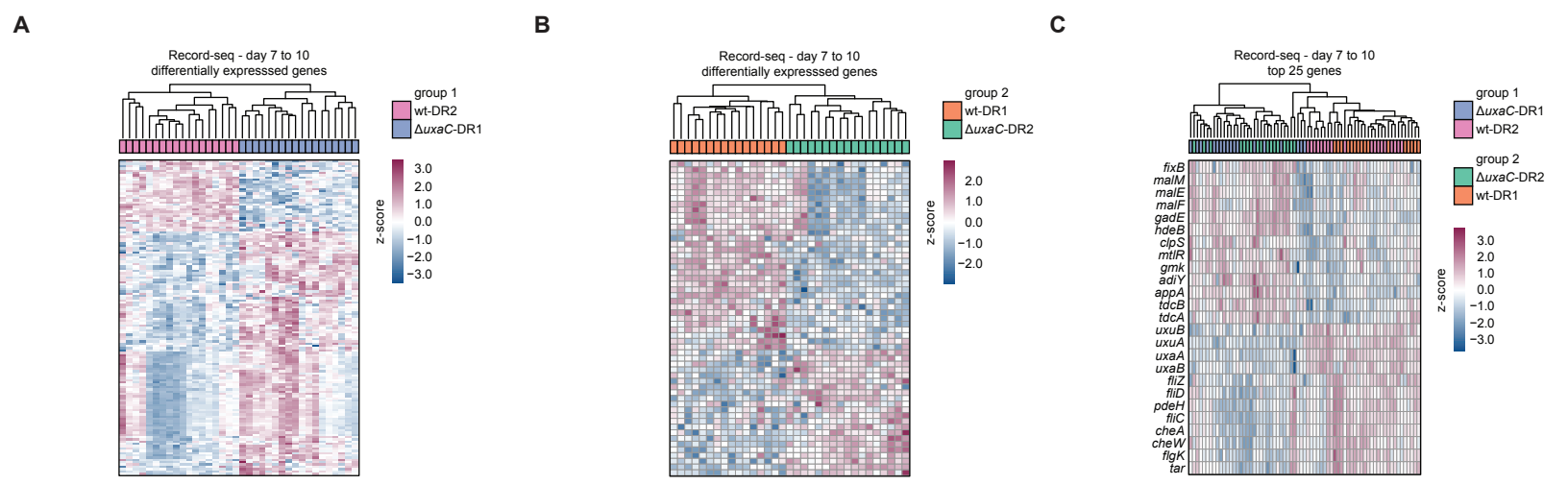

A
Record-seq - day 7 to 10
differentially expresssed genes
group 1
wt-DR2
ΔuxaC-DR1
z-score
3.0
2.0
1.0
0.0
−1.0
−2.0
−3.0
B
Record-seq - day 7 to 10
differentially expresssed genes
group 2
wt-DR1
ΔuxaC-DR2
z-score
2.0
1.0
0.0
−1.0
−2.0
C
Record-seq - day 7 to 10
top 25 genes
group 1
ΔuxaC-DR1
wt-DR2
group 2
ΔuxaC-DR2
wt-DR1
z-score
3.0
2.0
1.0
0.0
−1.0
−2.0
−3.0
fixB
malM
malE
malF
gadE
hdeB
clpS
mtlR
gmk
adiY
appA
tdcB
tdcA
uxuB
uxuA
uxaA
uxaB
fliZ
fliD
pdeH
fliC
cheA
cheW
flgK
tar

Fig. S12. Record-seq enables multiplexed transcriptional profiling of isogenic bacterial strains coinhabiting the mouse intestine.

(A) Heatmap showing hierarchical clustering of Record-seq data from experimental group 1 consisting of Δ*uxaC E. coli* harboring DR1 recording plasmid (blue) in the presence of wt *E. coli* harboring DR2 recording plasmid (pink) in the same mouse, using differentially expressed genes (DEGs) identified from days 7 to 10. Z-score standardized gene-aligning spacer counts are shown. (B) Heatmap showing hierarchical clustering of Record-seq data from experimental group 2 consisting of Δ*uxaC E. coli* harboring DR2 recording plasmid (green) in the presence of wt *E. coli* harboring DR1 recording plasmid (orange) in the same mouse, using identified DEGs identified from days 7 to 10. Z-score standardized gene-aligning spacer counts are shown. (C) Heatmap showing hierarchical clustering of Record-seq data from experimental group 1 consisting of Δ*uxaC E. coli* harboring DR1 recording plasmid (blue) in the presence of wt *E. coli* harboring DR2 recording plasmid (pink) in the same mouse and experimental group 2 consisting of Δ*uxaC E. coli* harboring DR2 recording plasmid (green) in the presence of wt *E. coli* harboring DR1 recording plasmid (orange) in the same mouse. The top 25 DEGs are shown. Z-score standardized gene-aligning spacer counts are shown. Panels A to C correspond to the experiment outlined in **Fig. 7A** with n=5 independent biological replicates. Count threshold was 10^4. Outliers were excluded based on modified Z-score and relative deviation from the mean (see **Methods**).

Supplementary Tables

Table S1. Composition of the standard rodent chow diet, purified starch- and fat-based diets.

Table S2. Genes differentially expressed in *E. coli* as detected by Record-seq on day 7, pairwise comparison of monocolonized with *E. coli* MG1655 (wt) mice on chow, star or fat diets, experiment 1, corresponding to **Fig. 1D**.

Table S3. Genes differentially expressed in *E. coli* as detected by Record-seq on day 7, pairwise comparison of monocolonized with *E. coli* MG1655 (wt) mice on chow, starch or fat diets, experiment 2, corresponding to **fig. S3A**.

Table S4. Genes differentially expressed in *E. coli* MG1655 as detected by Record-seq or RNA-seq and ordered as displayed in the heatmaps of **Fig. 2A; Fig. 3B; Fig. 4C; Fig. 5C; Fig. 6C; Fig. 7B; fig. S2 D and E; fig. S3, G and H; fig. S5E; fig. S6H; fig. S7F and fig. S12, A to C**.

Table S5. Full output of EcoCyc pathway analysis based on Record-seq on day 7, pairwise comparison of *E. coli* in mice on different diets, experiment 1, corresponding to **Fig. 1D**.

Table S6. KEGG- and GO-based OA analysis based on Record-seq on day 7, pairwise comparison of *E. coli* in mice on different diets, experiment 1, corresponding to **Fig. 1D**.

Table S7. Genes differentially expressed in *E. coli* MG1655 as detected by Record-seq in the pairwise comparison of chow-fed mice to starch-fed mice along with corresponding $\boldsymbol{log_2 FC}$ values from RNA-seq of the feces, cecum, proximal colon and distal colon corresponding to **Fig. 3A**.

Table S8. Genes differentially expressed in *E. coli* MG1655 (wt) as detected by Record-seq, pairwise comparison of control mice to mice treated 2% (w/v) DSS in the drinking water, corresponding to **Fig. 4D**.

Table S9. Full output of EcoCyc pathway analysis based on Record-seq, pairwise comparison of *E. coli* in control mice to mice treated with 2% DSS in the drinking water, corresponding to **Fig. 4D**.

Table S10. KEGG- and GO-based OA pathway analysis based on Record-seq, pairwise comparison of *E. coli* in control mice to mice treated with 2% DSS in the drinking water, corresponding to **Fig. 4D**.

Table S11. Genes differentially expressed in *E. coli* as detected by Record-seq, pairwise comparison of *E. coli* in the presence or absence of *B. theta* in the same mouse, experiment 1, corresponding to **Fig. 5A**.

Table S12. Genes differentially expressed in *E. coli* as detected by Record-seq, pairwise comparison of *E. coli* in the presence or absence of *B. theta* in the same mouse, experiment 2, corresponding to **fig. S7D**.

Table S13. Full output of EcoCyc pathway analysis based on Record-seq, pairwise comparison of *E. coli* in the presence or absence of *B. theta*, corresponding to **Fig. 5A**.

Table S14. KEGG- and GO-based OA pathway analysis based on Record-seq, pairwise comparison of *E. coli* in the presence or absence of *B. theta*, corresponding to **Fig. 5A**.

Table S15. Genes differentially expressed in *E. coli* as detected by Record-seq at the 21-hour timepoint, pairwise comparison of sDMDMm2 mice on chow diet and starch-based diet, corresponding to **Fig. 6A**.

Table S16. Genes differentially expressed in *E. coli* as detected by Record-seq at the 21-hour timepoint, pairwise comparison of sDMDMm2 mice on chow diet and starch-based diet, corresponding to **fig. S8D** and an independent experimental replicate.

Table S17. Full output of EcoCyc pathway analysis based on Record-seq, pairwise comparison of sDMDMm2 mice on chow diet and starch-based diet, corresponding to **Fig. 6A**.

Table S18. KEGG- and GO-based OA analysis based on Record-seq, pairwise comparison of sDMDMm2 mice on chow diet and starch-based diet, corresponding to **Fig. 6A**.

Table S19. Genes differentially expressed in wt *E. coli* and *ΔuxaC E. coli* as detected by Record-seq from aggregate counts, pairwise comparison on the starch-based diet, corresponding to **fig. S11A**.

Table S20. Full output of EcoCyc pathway analysis based on Record-seq, pairwise comparison on the starch-based diet, corresponding to **fig. S11A**.

Table S21. KEGG- and GO-based OA analysis based on Record-seq, pairwise comparison on the starch-based diet, corresponding to **fig. S11A**.

Tables S1 to S21 are supplied as excel sheets due to size constrains.

E. coli strain	supplier	order #	genotype
BL21-Gold(DE3)	Agilent Technologies	230132	*E. coli* B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^R gal λ(DE3) *endA* Hte
MG1655 (Bern)	Andrew Macpherson	NA	F^- lambda$^-$ rph-1
MG1655 Str^R Δ*gntK*/Δ*idnK*	Tyrrell Conway	NA	F^- lambda$^-$ rph-1 Δ*gntK* Δ*idnK* Str^R
MG1655 Str^R Δ*uxaC*	Tyrrell Conway	NA	F^- lambda$^-$ rph-1 Δ*uxaC* Str^R Kan^R
MG1655 Str^R Δ*uxaC* ΔKan^R	Tyrrell Conway	NA	F^- lambda$^-$ rph-1 Δ*uxaC* Str^R
MG1655 Str^R	Tyrrell Conway	NA	F^- lambda$^-$ rph-1 Str^R
MG1655 Str^R Nal^R	Tyrrell Conway	NA	F^-, lambda$^-$ rph-1 Str^R Nal^R

Table S22. *E. coli* strains used in this study.

Bacterial species	DSMZ
Lachnoclostridium sp. YL32	DSM 26114
Ruminiclostridium sp. KB18	DSM 26090
Bacteroides sp. I48	DSM 26085
Parabacteroides sp. YL27	DSM 28989
Burkholderiales bacterium YL45	DSM 26109
Erysipelotrichaceae bacterium I46	DSM 26113
Blautia sp. YL58	DSM 26115
Flavonifractor plautii YL31	DSM 26117
Bifidobacterium animalis subsp. animalis YL2	DSM 26074
Lactobacillus reuteri I49	DSM 32035
Akkermansia muciniphila YL44	DSM 26127
Enterococcus faecalis KB1	DSM 32036

Table S23. Taxa of the stable defined moderately diverse mouse microbiota 2 (sDMDMm2).

Primer	Sequence (5′ → 3′)
FS_2814	GTACTGGCGTATGAATCACG
FS_2815	CGAATCAGGATAATACCCGG
FS_2816	HEX-AGCGATCTGAAGAACCAGGAAT-BHQ-1

Table S24. ddPCR primers and probe.

Primer	Sequence (5′ → 3′)
FS_0963	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
FS_0964	AAAGGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
FS_2759	AAAGATTTGTACCAAGGTTCCTAGNNNNNNNNNNGATCGGAAGAGCACACGTCT GAACTCCAGTCAC
FS_2769	CTAGGAACCTTGGTACAAAT
FS_3046	GAGTTGATAGACAATGTAACCCACTCGTGCACCTCGAGCAACTGATCTTATAGA TACAGCATCTTTTACTTTCCTCGAGTAGCCTAGCATAACCCCGCGGGGCCTCTT CGGGGGTCTCGCGGGGTTTTTTGCTATAAAACGAAAGGCTCAGTCGAAAGACTG GGCCTTTCGTTTTATCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA GGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATGTCTTCATGGTAGTACCA AGATACGAAGACATAGTGGCGGGGAAGCTTATGTTCCATAGCAAAAAGTCGGTC AGTCTCGTGGCTGAAATCATGAGTTCCACAAAATGGCTGAAATTCAAGGAAAAT CAGGAATCTCAGAAAAACGATCGACCGACTTTTTCGATAAAATGGTTGCAAAAA TGAGAAAAATCTGATTTAATAGAATCTGAAAACAGCGGAAATGCTGTTGTCGTA CTTTACCTAAAAGGAATTGAAACGTCCCCGCCAGGTTGAATCCGATATTTGGAG GTACGATGGAACAGTCTGGGTGGGATTGAGAAGAGAAAAGAAAACCGCCGATCC TGTCCACCGCATTACTGCAAGGTAGTGGACAAGACCGGCGGTCTTAAGTTTTTT GGCTGAAGCGGCCGCCTCATGGTTATGGCAGCACTGCATAATTTTCTTA
FS_3047	CCGGAACTTGACAATTAATCATCCGGCTCGTATAATGTGTGGAGG
FS_3048	CACTCCTCCACACATTATACGAGCCGGATGATTAATTGTCAAGTT
FS_3049	CCGGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTCTAGT
FS_3050	CACTACTAGAGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAAT
FS_3051	CCGGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTCTAGT
FS_3052	CACTACTAGAGCTAGCACTATACCTAGGACTGAGCTAGCCGTAAAT
FS_3053	CCGGATTTACGGCTAGCTCAGTCCTAGGTACAATGCTAGCTCTAGT
FS_3054	CACTACTAGAGCTAGCATTGTACCTAGGACTGAGCTAGCCGTAAAT
FS_3055	CCGGATTTATAGCTAGCTCAGCCCTTGGTACAATGCTAGCTCTAGT
FS_3056	CACTACTAGAGCTAGCATTGTACCAAGGGCTGAGCTAGCTATAAAT
FS_3057	CCGGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCTCTAGT

FS_3058 CACTACTAGAGCTAGCACAATCCCTAGGACTGAGCTAGCTGTCAAT

FS_3210 CCGGATTTACAGCTAGCTCAGTCCTAGGGACTGTGCTAGCTCTAGT

FS_3211 CACTACTAGAGCTAGCACAGTCCCTAGGACTGAGCTAGCTGTAAAT

FS_3212 CCGGACTGATAGCTAGCTCAGTCCTAGGGATTATGCTAGCTCTAGT

FS_3213 CACTACTAGAGCTAGCATAATCCCTAGGACTGAGCTAGCTATCAGT

FS_3214 CCGGACTGATAGCTAGCTCAGTCCTAGGGATTATGCTAGCTCTAGT

FS_3215 CACTACTAGAGCTAGCATAATCCCTAGGACTGAGCTAGCTATCAGT

FS_3344 GTGATCTAACTCGAGTAGCCTAGCATAACCCCGCGGGGCCTCTTCGGGGGTCTC
GCGGGGTTTTTTGCTATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT
TTTATCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGA
GCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT
GCTGAAAGGAGGAACTATATCCGGACTGATAGCTAGCTCAGTCCTAGGGATTAT
GCTAGCTCTAGTAGTGGAGAATTAAATTGGAAAAAGTCGGTCGATCTCATGCCT
GAAATCATGAATTCCGCAAAATGGCGGAAATTTAAGGAAAATCAGGAATCTCAG
AAAAACGATCGACCGACTTTTGTGATAAAATGGTTGCAAAAAAGAGAAAAATTT
GATTTAATAGAATGTGAAAATAGCGGAAATGCTGATGTTGTACCTTACCTATGA
GGAATTGAAACGTCCCCGCCAGGTTGAATCCGATATTTGGAGGTACGATGGAAC
AGTCTGGGTGGGATTGAGAAGAGAAAAGAAAACCGCCGATCCTGTCCACCGCAT
TACTGCAAGGTAGTGGACAAGACCGGCGGTCTTAAGTTTTTTGGCTGAAGCGGC
CGCTATTCT

FS_3194 AAAGCTAATATACCACCAGCAGTANNNNNNNNNNGATCGGAAGAGCACACGTCT
GAACTCCAGTCAC

FS_3204 TACTGCTGGTGGTATATTAG

FS_3316 TGAGATTACGATCGCCAGGTCATGNNNNNNNNNNGATCGGAAGAGCACACGTCT
GAACTCCAGTCAC

FS_3321 CATGACCTGGCGATCGTAAT

FS_0968 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNCCTAAAAGGAATTGAAAC

FS_0969 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCCTAAAAGGAATTGAAAC

FS_0970 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTAAAAGGAATTGAAA
C

FS_0971 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCCTAAAAGGAATTGAA
AC

FS_0972 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNCCTAAAAGGAATTGA
AAC

FS_0973 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNCCTAAAAGGAATTG
AAAC

FS_0974 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNCCTAAAAGGAATT
GAAAC

FS_3325 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNACCTATGAGGAATTGAAAC

FS_3326 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNACCTATGAGGAATTGAAA
C

FS_3327 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCTATGAGGAATTGAA
AC

FS_3328 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACCTATGAGGAATTGA
AAC

FS_3329 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNACCTATGAGGAATTG
AAAC

FS_3330 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNACCTATGAGGAATT
GAAAC

FS_3331 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNACCTATGAGGAAT
TGAAAC

FS_2238 AAAGCACTTTGGTTATAGAAGAGGGATCGGAAGA
GCACACGTCTGAACTCCAGTCAC

FS_2240 AAAGTCCCATGAATGTTCCACATGATCGGAAGAG
CACACGTCTGAACTCCAGTCAC

FS_2246 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCC
CTCTTCTATAACCAAAGTG

FS_2248 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCA
TGTGGAACATTCATGGGA

FS_2758 AAAGACTTTCCGCACAAACCGTGANNNNNNNNNN
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC

FS_2762	AAAGACAATCCGTCAAGTCACTAGNNNNNNNNNN
	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
FS_2760	AAAGTAAACGACTACACCCGCTCGNNNNNNNNNN
	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
FS_2761	AAAGCGATATCATCGTCCCTTTGTNNNNNNNNNN
	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
FS_2768	TCACGGTTTGTGCGGAAAGT
FS_2770	CGAGCGGGTGTAGTCGTTTA
FS_2771	ACAAAGGGACGATGATATCG
FS_2772	CTAGTGACTTGACGGATTGT
FS_2806	AAAGACGCAGGAAACAGGCTTGAT
FS_2807	ATCAAGCCTGTTTCCTGCGT

Table S25. Oligonucleotides for cloning and SENECA adapter ligation oligonucleotides.