A discovery platform to identify inducible synthetic circuitry from varied microbial sources.

Clare M. Robinson[1,2], David Carreño[1,2], Tim Weber[1,†], Yangyumeng Chen[1], David T. Riglar[1,2,*]

[1] Department of Infectious Disease, Imperial College London, London, SW72AZ, UK.
[2] The Francis Crick Institute, London, NW11AT, UK.
[†] Current address: Department of Molecular Life Sciences, University of Zurich, Zurich, CH-8057, Switzerland.

*To whom correspondence should be addressed: d.riglar@imperial.ac.uk.

Abstract:

Gut microbes encode a variety of systems for molecular sensing and controlling conditional gene expression within the mammalian gut. Synthetic biology approaches such as whole-cell biosensing and 'sense-and-respond' therapeutics aim to tap into this vast sensing repertoire to drive clinical and pre-clinical applications. An ongoing constraint is the limited number of well-characterized inducible circuit components to specifically sense *in vivo* conditions of interest, such as disease. Here, we extend the flexibility and power of a biosensor screening platform using bacterial memory circuits encoded in a gut commensal *E. coli*. We construct libraries driven by potential sensory components derived from a combination of *E. coli* promoters or bacterial two-component systems (TCSs) sourced from diverse gut bacteria. Each is tagged with unique DNA barcodes using a pooled construction method. Using our pipeline, we evaluate sensor activity and performance heterogeneity across *in vitro* and *in vivo* conditions including using a mouse inflammation model. We demonstrate the method's ability to identify biosensors of interest, including the identification of unannotated TCSs. Following the optimisation of library construction, analysis, and delivery to account for the challenges of working with engineered bacteria within a conventional mammalian gut microbiome, we identify and validate several further biosensors of interest responding to the murine gut environment, and specifically during inflammatory conditions. This approach can be applied to transcriptionally activated sensing elements of any type and will allow for rapid development of new biosensors that can advance synthetic biology approaches for complex environments.

Introduction:

A longstanding vision for synthetic biology is the development of bacterial whole-cell biosensors and 'sense-and-respond' live biotherapeutic products. These systems promise to autonomously sense and induce reporters or therapeutics under defined spatial, dietary and disease conditions [1,2]. Potential applications include diagnostics, disease monitoring and biotherapeutics [2]. However, well-characterised biosensing circuits for the mammalian gut remain limited, hindering the complete fulfilment of this vision[3,4].

Synthetic genetic memory circuits, which convert transient signals into sustained responses, for example through transcriptional switches [5,6] or DNA editing [7-10], are powerful tools for non-invasive biosensor reporting and actuation. In rodent models, such circuits have been used to successfully track inflammatory signals [6,8,10,11], record microbial response to dietary perturbations [7,8], detect tumour DNA [12], and actuate sustained biotherapeutic secretion [10]. However, due to the extended, often permanent, nature of the response, input signals typically need to show low background activation combined with strong induction dynamics for best compatibility with these approaches.

We previously developed a high-throughput memory system (HTMS) in mouse commensal *Escherichia coli* NGF-1 [13] to screen *E. coli* promoters as biosensor candidates within the mouse gut [11]. We now expand the power and flexibility of this screening platform to accommodate a diversity of sensory components from varied microbial sources (Figure 1A). These include multi-gene heterologous two-component systems (TCSs), for which we develop a custom computational pipeline for rapid identification of cloning-compatible systems, and shorter native promoter-only-based sensors, identified from previous studies to sense conditions of interest [8,11]. Sensors are constructed by multiplexed golden gate assembly and integrated into the HTMS chassis (Figure 1B). Incorporating unique DNA barcodes into each sensor allows simultaneous testing of variable-sized sensors and flexibility for library combination and re-combination.

Memory recording is based on λ-phage's lysis-lysogeny switch [5]. In the chassis' memory 'OFF' state, *cI* represses β-galactosidase (*lacZ*) and streptomycin/spectinomycin adenyltransferase (*aadA*) reporter cassettes (Figure 1C, left). Upon exposure to an inducing signal, the sensing component triggers expression of cI dominant negative (cI^{DN}), switching cells into memory 'ON' states, in which the repression of *cro*, *lacZ* and *aadA* is relieved (Figure 1C, right). Repression of *cI* by cro maintains switching status irrespective of the presence of the induction signal, resulting in cellular 'memory'. Activated sensors can be identified from a large library through selection by growth in spectinomycin followed by massive parallel sequencing (Figure 1D).

Bacterial TCSs are modular signalling pathways capable of detecting various molecular signals [14]. They are typically composed of a membrane-bound histidine kinase (HK) and partner response regulator (RR) that control the transcription of target genes [15]. While it is not expected that all TCSs will function in a heterologous host exactly as they do in their native host, we and others have successfully transferred sensing capacities between bacterial species. We previously developed a memory biosensor to the inflammatory biomarker tetrathionate, driven by a heterologously expressed TCS from *Salmonella enterica* subsp. *enterica* serovar Typhimurium (herein *S.* Typhimurium), which controlled λ-memory induction through the diverging promoter for the neighbouring ttr_{BCA} operon [6]. Given the ability of TCSs to facilitate non-native sensing capacities in heterologous hosts, they are ideal targets for sensors to diversify engineered chassis biosensing capacity using a screening approach.

Here, we build two new libraries of sensor components: the first focused on TCSs from diverse bacterial origins (Library 1) and a second predominantly containing *E. coli* derived sensors targeted to be responsive to the mammalian gut (Library 2). We test these libraries *in vitro* and *in vivo* in the murine gut, identifying and validating several new sensors of interest. Included are those responsive to the gut environment both generally and specifically during inflammation. Together, this work advances methodology for the application of library-based screens of synthetic biology components in *in vivo* gut contexts and validates the power of this approach for identifying context-dependent sensors that can actuate synthetic circuits within the mammalian gut.

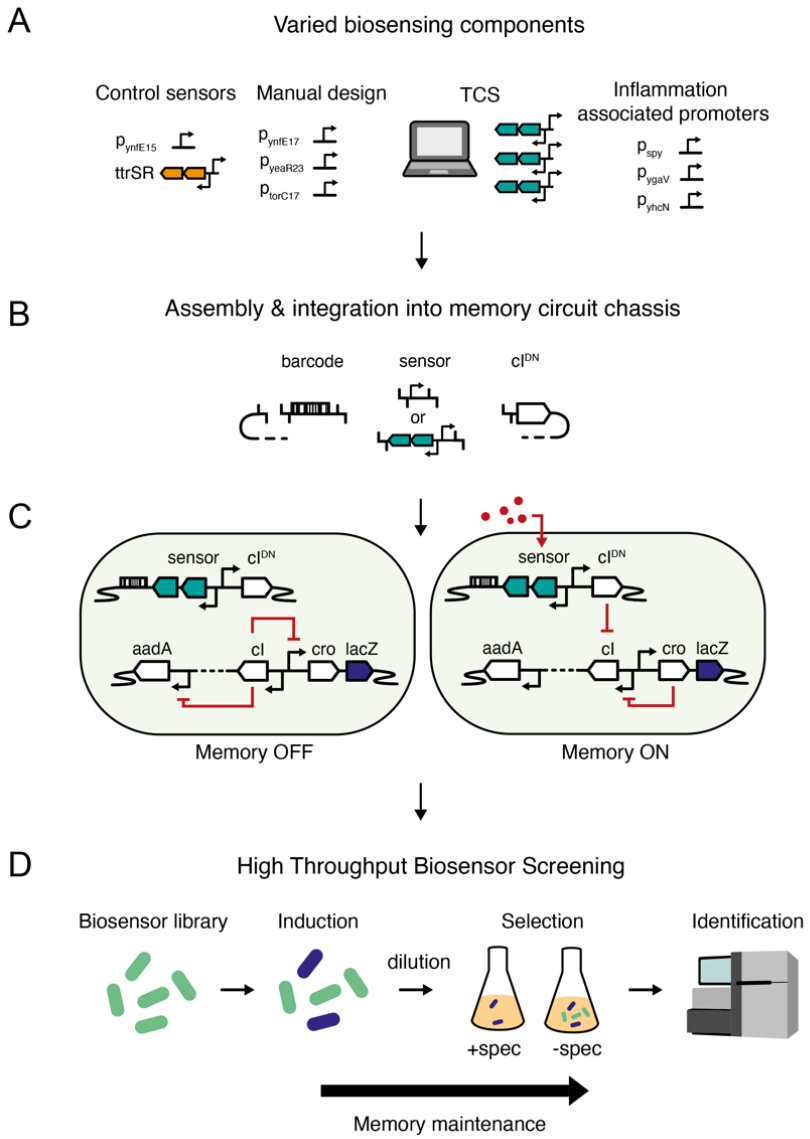

Figure 1. A high-throughput transcriptional screening pipeline based on an engineered memory circuit to identify novel inducible circuits A) Various sensing components can be combined and B) assembled, along with unique DNA barcodes, by multiplexed golden gate assembly to drive switching of C) a λ-phage-based memory circuit. In the presence of an activating stimulus, cI^{DN} is expressed and inhibits *cI*, de-repressing *cro* and allowing expression of reporter genes *lacZ* and *aadA*. Cro expression maintains this memory ON state irrespective of ongoing exposure to the activating stimulus. D) A library of memory-linked bacterial biosensors can be rapidly screened for activated sensors by growth in spectinomycin selective media followed by sequencing of recovered gDNA for identification.

91 **Results:**

92 *Construction of bacterial biosensor libraries encoding diverse native and heterologous sensing machinery:*

93 Most HKs are encoded within 200bp of their partner RR [16]. Where a HK-RR pair sit next to a
94 diverging gene, we term this a 'grouped' TCS (Figure 2A). Many grouped TCSs regulate their adjacent and
95 divergent genes. Of a total of 30 known TCS pairs in *E. coli* MG1655, 13 lie in this 'grouped' formation, and
96 of these, 10/13 regulate the divergent, adjacent gene based on Ecocyc annotations [17]. We, therefore,
97 focused on these 'grouped' TCSs for biosensor library expansion due to the ability to replicate complete
98 sensing machinery (HK, RR and putatively regulated promoter sequence) in a single PCR for transfer to the
99 heterologous *E. coli*.

100 To enable rapid library design, we developed a computational pipeline to identify grouped TCSs
101 from available genomic sequencing data (see methods). Using this pipeline directed against 17 bacterial
102 species for which we had access to both full genome sequences and gDNA of exact or closely related
103 strains, we identified 153 'grouped TCSs', 146 of which were compatible with our downstream cloning
104 strategy (Table 1). Identified TCSs were cloned so that the divergent promoter would drive cI^{DN} expression
105 (Figure 1B). Pooled golden gate assembly tagged each sensor with a 106bp randomised nucleotide
106 hypervariable barcode flanked by sequencing adaptor sites for unique identification. The plasmid library
107 was then transformed, and genome integrated into *E. coli* NGF-1 encoding the high-throughput memory
108 system (PAS811) [11] (Supplementary Figure 1A).

109 **Table 1: Compatible TCSs computationally identified and cloned from diverse gut bacteria.**

Strain	Total TCSs	'Grouped'	SapI compatible	BsaI compatible (not SapI)	Successfully amplified	Assigned Library 1	Assigned Library 2
Bacteroides caccae	9	5	5	0	4	1	1
Bacteroides dorei	14	7	3	4	2	0	2
Bacteroides fragilis	13	8	6	0	5	2	0
Bacteroides thetaiotaomicron	12	8	3	3	1	1	0
Blautia producta	90	26	19	6	14	5	5
Citrobacter rodentium	22	14	11	2	13	11	2
Clostridium difficile	43	8	8	1	3	0	1
Edwardsiella tarda	20	11	8	1	5	5	0
Enterococcus faecalis	13	4	3	0	3	2	0
Klebsiella oxytoca	25	12	8	5	2	1	0
Klebsiella pneumoniae	25	12	10	0	8	7	0
Lactobacillus gasseri	6	1	1	0	1	1	0
Lactobacillus plantarum	13	2	2	1	2	1	0
Ruminococccus gnavus	36	13	8	4	6	2	3
Salmonella Typhimurium	22	11	6	2	5	3	0
Vibrio cholerae	24	13	8	3	5	3	2
Yersinia enterocolitica	20	8	7	1	5	4	0
Total	407	163	116	33	84	49	16
						Total TCS	65

110 Barcodes were identified and associated with their corresponding sensing components by a
111 combination of massive parallel long- and short-read sequencing (Supplementary Figure 2A). Sequencing
112 confirmed successful barcode assignment for 49 unique SapI compatible TCSs in the first instance. We
113 supplemented these strains with 9 additional individually cloned 'curated' sensors identified in our previous
114 studies [6,11] to afford 'Library 1' (Table 1, Data File 1). Curated sensors included two controls: i) a *S.*
115 Typhimurium *ttrSR*-P_{ttrBCA} sensor, edited with a synthetic ribosome binding site, (ST TCS1*) that previously
116 showed response to the inflammatory biomarker tetrathionate [6], and ii) the *E. coli* Nissle 1917 $P_{ynfEFGH}$
117 promoter edited with a synthetic 5' untranslated region and ribosome binding site, MCD15 (Ec ynfE15)[18]
118 that previously showed switching under *in vitro* anaerobic growth conditions and in the murine gut [11].
119 Library 1 contained 504 assigned unique barcodes linked to these 58 sensors.

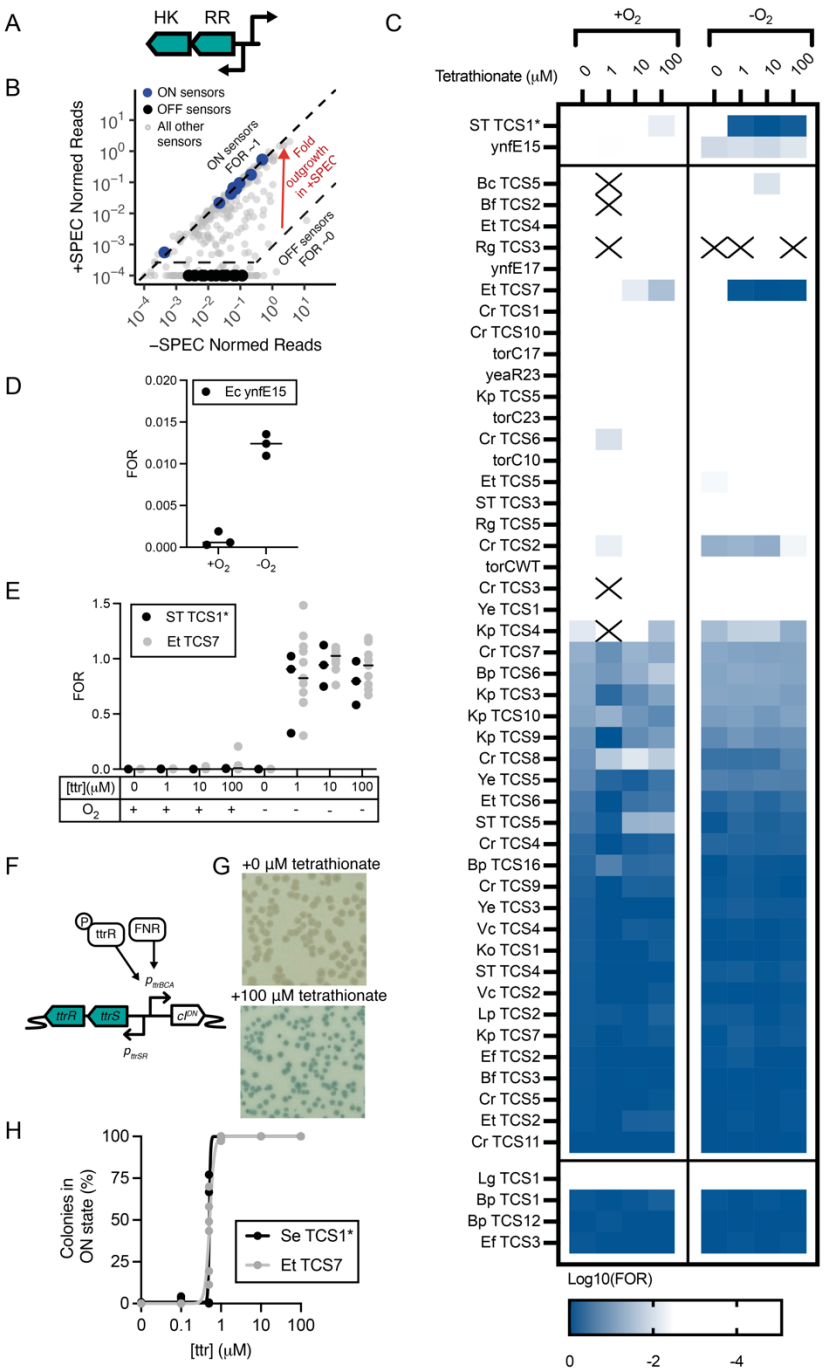

Figure 2: Barcoded biosensor screening of Library 1 identifies specific biosensors *in vitro*. A) 'Grouped' TCS arrangements are ideal for straightforward cloning of libraries encoding heterologous sensing components together with a probable regulated promoter to trigger memory formation. B) Normalised reads of a typical sample, with ON and OFF control sensors highlighted. ON sensors will have equal growth in the presence (+SPEC) or absence (-SPEC) of spectinomycin selective media, resulting in a fractional odds ratio (FOR) of ~1. OFF sensors will not grow in spectinomycin selective media (+SPEC), resulting in a FOR of ~0. C) Average log10 transformed fractional odds ratio of all sensors within the library in the presence and absence of oxygen and varying levels of tetrathionate. D) Ec ynfE15 sensors showed increased activation under anaerobic conditions. E) Sensor activation of barcodes from Et TCS7, ST TCS1* under anaerobic conditions across all oxygen and tetrathionate combinations tested. For D) and E) each point represents an individual barcode's FOR, with medians marked. F) Putative sensor regulation of both Et TCS7 and ST TCS1* G) Et TCS7 was individually cloned (DTR302) and tested by exposure to anaerobic conditions in the presence and absence of 100μM tetrathionate and indicator plating, confirming the sensor is responsive to tetrathionate under anaerobic conditions. H) Response dynamics across various tetrathionate concentrations were similar to memory driven by ST TCS1*(DTR277).

135 Building on the flexibility of the method to easily combine sensors, barcodes and libraries, we
136 subsequently cloned a set of additional triggers, targeting those responsive to gut and gut inflammation
137 related signals. These included *E. coli* promoter regions previously induced in mouse models, including
138 during inflammation [8]; 2) computationally identified TCSs that were not assigned in Library 1 construction,
139 including those with BsaI but not SapI compatibility; and 3) several additional curated sensors of interest
140 (Table 1, Data File 1). Library 2 contained 10,063 assigned unique barcodes linked to 17 additional TCSs and
141 72 *E. coli* promoter regions of interest.

142 The number of assigned barcodes per sensor ranged between 1 and 44 (median 5) for Library 1 and
143 1 and 1001 (median 52) for Library 2 (Supplementary Figure 2B). Each library demonstrated similar
144 abundance profiles (Supplementary Figure 2C). Between the two libraries we confirmed at least 1 barcode
145 for ~80% of successfully PCR-amplified TCS sensor regions with this method (Table 1).

146 *In vitro validation confirms barcoded library biosensor screening capabilities:*

147 To quantify the state of each sensor we compared barcode sequencing read counts for each
148 barcode from samples split and grown in the presence or absence of spectinomycin (Figure 2B). Only ON
149 sensors grow with spectinomycin due to *aadA* expression (Figure 1C-D). Degree of switching is represented
150 as a fractional odds ratio (FOR) by normalisation to reads from known positive and negative strains (see
151 methods) (Figure 2B). FOR ~1 corresponds to an ON sensor, and ~ 0 to an OFF sensor.

152 To validate barcoded library screening, we cultured pooled sensors from Library 1 *in vitro* with
153 varying tetrathionate concentrations and with or without oxygen (Figure 2C). Various sensors showed
154 differential activation between the conditions. As expected, around half of the sensors were consistently ON
155 in all conditions, suggesting a level of expression from these promoters during *in vitro* growth, resulting in
156 activation of the memory circuit under all conditions. Mirroring our previous results, Ec ynfE15 sensors had
157 elevated odds ratios in anaerobic conditions (Figure 2C-D) [11] and ST TCS1* sensors activated with
158 tetrathionate in the absence of oxygen (Figure 2C) [6]. Together, these results demonstrate the functionality
159 of the barcoding approach to combine sensors of diverse size and origins.

160 *Barcode screening successfully identifies unannotated biosensor candidates*

161 The 22 barcodes linked to *Edwardsiella tarda* TCS7 (Et TCS7) were also consistently induced at levels
162 comparable to ST TCS1* across the conditions tested (Figure 2C, E). While the available gene annotations
163 for the strain used for TCS extraction (*E. tarda* KC-Pc-HB1) did not predict this behaviour, BLAST searches
164 identified homology to a putative tetrathionate reductase promoter region with its associated TCS in *E.*
165 *tarda* strains Et54[19] and FL6-60 [20] (Figure 2F). Et TCS7 was therefore cloned as an individual strain (DTR302)
166 and its response to tetrathionate was confirmed by plating of induced cultures on X-gal indicator plates for
167 blue-white colony screening. When grown anaerobically, both DTR302 and an individual strain triggered by
168 ST TCS1* (DTR277) demonstrated strong induction in the presence of >1µM tetrathionate (Figure 2G-H).
169 This demonstrates the ability of our screening pipeline to identify unannotated biosensors with specific
170 sensing functions using our screening approach.

171 *Barcode and strain diversity is maintained over short periods within the murine gut*

172 We have previously demonstrated robust *E. coli* NGF-1 colonisation of the conventional mouse gut,
173 typically following a dose of streptomycin to reduce colonisation resistance, but in select cases also without
174 pre-treatment [5,6]. To test colonisation characteristics in our current facility, we administered sensor Library 1
175 by oral gavage (~10^{11} bacteria/mouse) to groups of 2 healthy, conventional mice 24 hours following
176 administration of streptomycin antibiotics (1mg or 5mg/mouse) or mock treatment (Supplementary Figure
177 3A). Faecal bacterial CFU counts reflected streptomycin provision, with engineered bacteria in untreated
178 animals dropping below the detection limit at day 3, and in mice administered 1mg streptomycin on day 4-5
179 (Supplementary Figure 3B). Engineered *E. coli* successfully colonised mice provided with 5mg streptomycin
180 throughout the full 53-day experiment (Supplementary Figure 3B).

181 Strain diversity, as estimated by percentage of barcodes identified via sequencing, was high in all
182 animals at day 1 post-administration but not measurable by day 4 in untreated and 1mg streptomycin-
183 treated mice as no engineered bacteria were recovered from these animals on these days (Supplementary
184 Figure 3C). Even in 5mg streptomycin-treated mice that were effectively colonised, barcode and sensor

diversity dropped quickly following administration (Supplementary Figure 3C and D). By day 4 only 3 – 4 % of total barcoded strains remained, with a handful of strains dominating colonisation within each mouse (Supplementary Figure 3E). Such colonisation bottlenecks are consistent with other work examining barcoded isogenic strains in the mouse gut [21,22]. Consequently, all further library analyses focussed on short (2-day) experimental timeframes to maintain barcode diversity.

Barcode screening successfully identifies biosensors in vivo:

To test sensor activation in the gut, we administered Library 1 sensors by oral gavage (~10^{12} bacteria/mouse) to mice (n = 4) 24 hours following administration of 5mg streptomycin (Figure 3A). Nearly all (>98%) barcodes and sensors (100%) were recovered on both days (Figure 3B-C). Various sensors showed elevated switching in animals compared to gavage cultures (Figure 3D; full data set: Supplementary Figure 4A). Consistent with previous testing [11], Ec ynfE15 strains showed odds ratio increases most prominently by day 2 (Figure 3E). The top-ranked activated sensor was *Citrobacter rodentium* TCS7 (Figure 3F), identified as the *ygiW* gene promoter paired with the adjacent and opposing TCS, QseCB. Expression of *ygiW* is regulated by the QseB homologue, PreA, in *S.* Typhimurium [23], and deletion of the HK QseC in K. pneumonia resulted increase of *ygiW* expression, suggesting the gene is directly or indirectly regulated by the QseBC system in this species also [24]. However, in *E. coli*, expression of *ygiW* has not been linked to QseB, but rather hydrogen peroxide and cadmium stress [25], the gluconate utilisation response regulator, GlaR, and the leucine responsive global transcriptional regulator, Lrp (Figure 3G) [17].

To validate the observed sensor response, we cloned Cr TCS7 as an individual strain (DTR307) and delivered it to mice via oral gavage 24 hours following streptomycin treatment (5mg) (Figure 3H). DTR307 response to the gut was tested over the 48 hours following administration via plating of faecal samples on indicator plates (Figure 3I). Similarly to our library-based screen, the sensor was off in the gavage sample, and showed consistent switching in ~20% of bacteria collected from faecal pellets from all mice on both days post gavage (Figure 3J). This validates the signal seen within the library-based screen and demonstrates the ability for the screening pipeline to identify sensors of interest within the mouse gut.

Another biosensor of interest ranked as showing higher response within the murine gut than our control Ec ynfE15 is Cr TCS2, identified as the *C. rodentium* intergenic region leading to *citC* with adjacent TCS DpiAB (Figure 3D, G, Supplementary Figure 5A-C). We further observed Ye TCS1, the *Y. enterocolitica* intergenic region leading to *macA* with adjacent TCS CpxAR (Supplementary Figure 4D-F), showing response to the gut, albeit below the threshold of the positive control Ec ynfE15. Homologous TCSs to these systems in *E. coli* are well characterised, with known inducers. When cloned and tested *in vitro* as individual sensors we confirmed the response of these heterologous sensors to known regulating signals for their cognate *E. coli* homologues (Supplementary Figure 5A-B, D-F). For Cr TCS2 we additionally tested truncated variants of the sensor, lacking the heterologous *C. rodentium dpiA* and *dpiB* genes, or DpiA binding sites (Supplementary Figure 5A-C). We observed a response to citrate in the truncated mutant lacking *dpiA* and *dpiB*, suggesting native *E. coli* host machinery is capable of regulation of the non-native intergenic promoter region, but no response to citrate upon the removal of DpiA binding sites, confirming regulation by DpiA is necessary for citrate response (Supplementary Figure 5C).

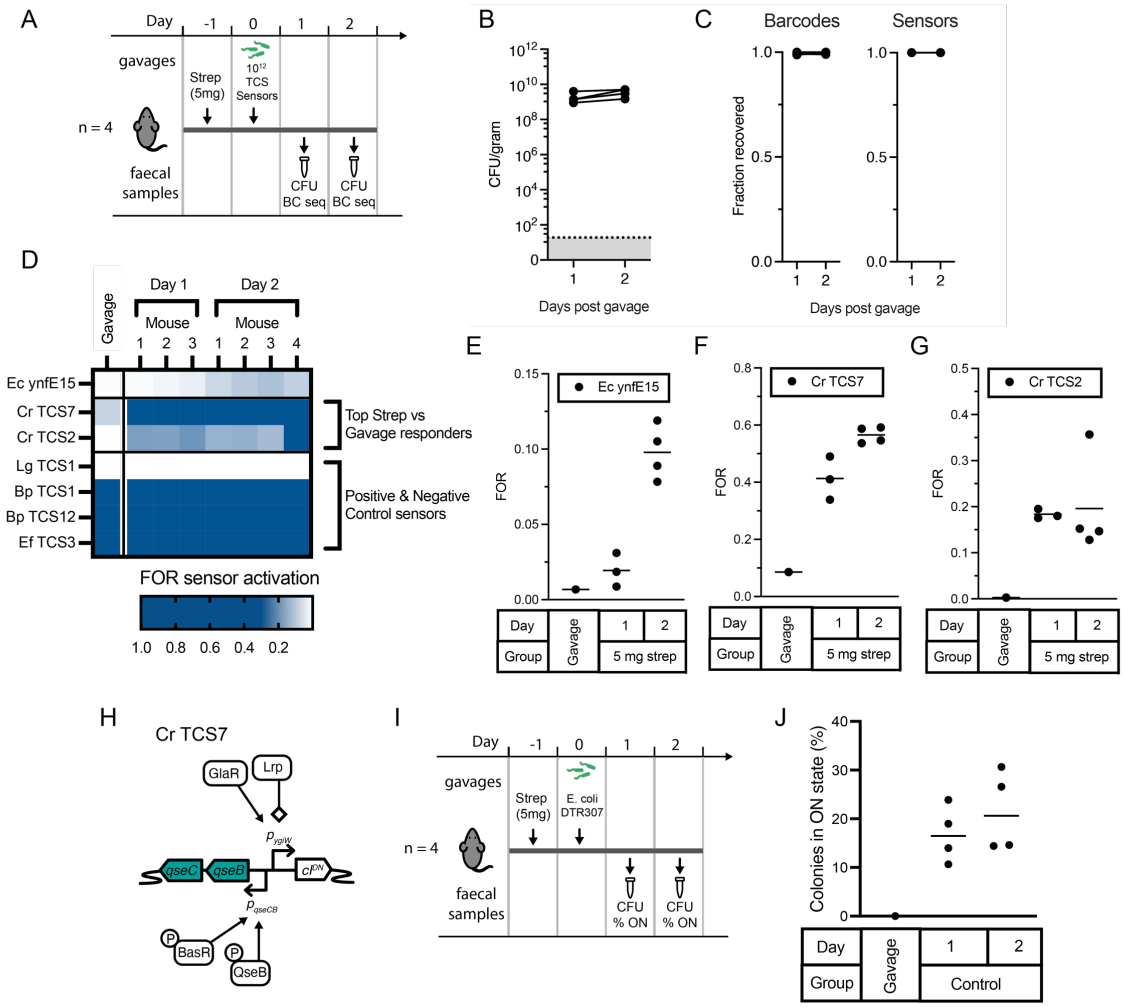

Figure 3: Barcoded memory screening of Library 1 successfully identifies biosensors for in vivo growth. A) Library 1 was administered to mice 1 day post-administration of 5 mg streptomycin (n=4). Faecal samples were collected for colonisation and memory switching analyses the two days following. B) Colonisation levels and C) fractions of recovered barcodes and sensors showed high retention in all mice. D) Sensor activation of top-ranked in vivo responsive sensors, and control sensors in mice on days 1 and 2 post library administration. Full dataset is shown in Supplementary Figure 4. E) The response of the control sensor, Ec ynfE15 and top-ranked in vivo responsive sensors F) Cr TCS7 and G) Cr TCS2 to the gut environment. Panels show FOR of faecal pellets from each mouse & gavage samples with mean response marked. H) Potential regulatory control of Cr TCS7 (QseBC-P_{ygiW}) based on regulation of C. rodentium and E. coli homologues. I) Experiment timeline for testing the response of Cr TCS7 (DTR307) within the mouse gut. J) Percentage of activated sensor colonies from plating of faecal contents on days 1 and 2 after sensor administration, panel shows counts from plated individual mouse faecal pellets and the gavage sample.

Identification and validation of inflammation responsive biosensors:

To specifically target inflammation-responsive circuit development, we combined Library 1 with Library 2 – the latter being predominantly built using *E. coli* promoters previously associated with response in the murine gut including during inflammation [8]. We administered the combined libraries via oral gavage to mice (n=4 per group; ~10^{11} bacteria/ mouse) pre-treated with streptomycin 24 hours prior (5mg/mouse) and with or without 3% DSS in their drinking water for the preceding 4 days to induce intestinal inflammation (Figure 4A). Colonisation levels were high in all animals and similar across both groups (Figure 4B). Reduction of colon length at endpoint and loss of body weight throughout the experiment confirmed the expected impact of DSS in treated animals (Figure 4C-D).

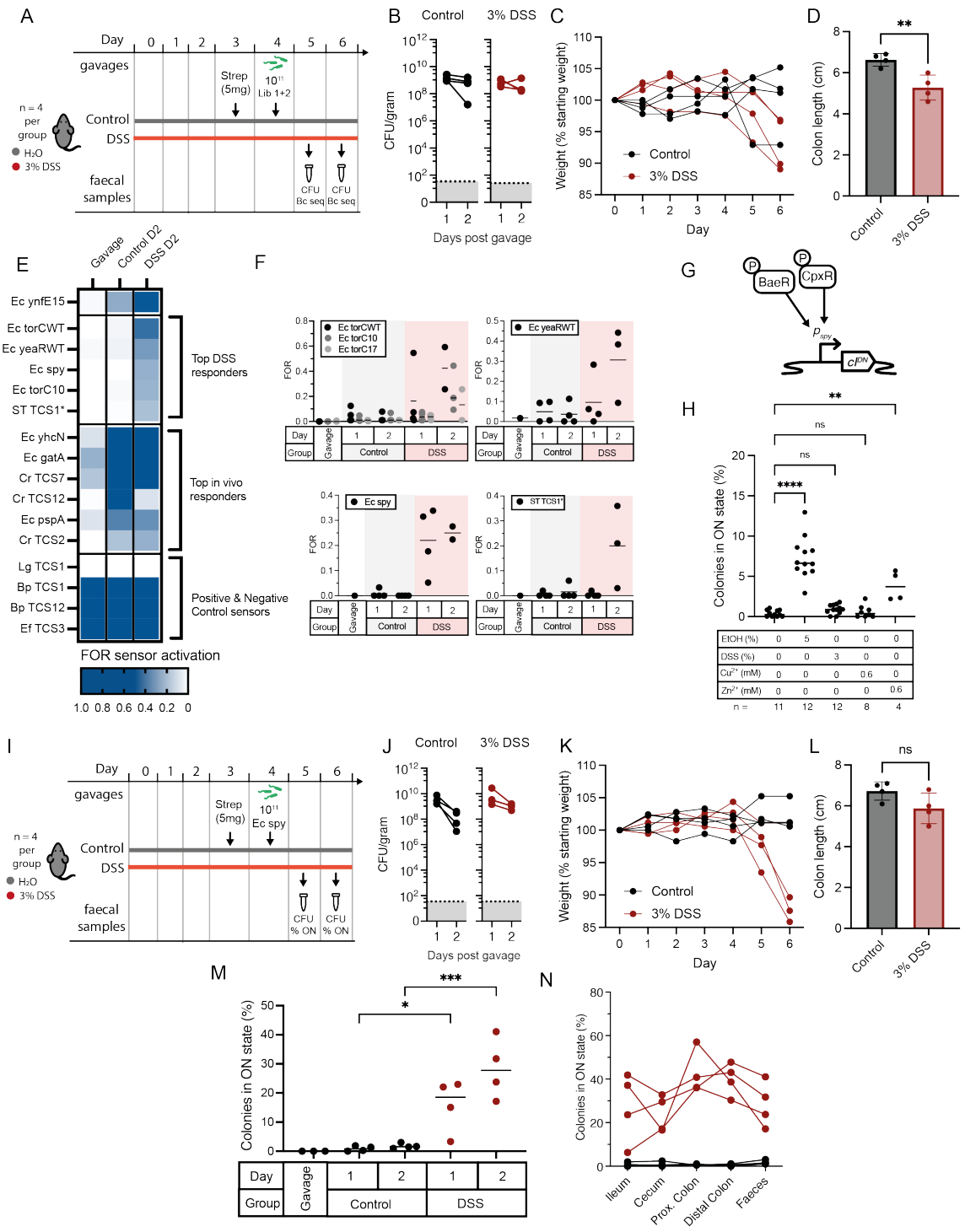

Figure 4: Identification and validation of inflammation-specific biosensors through screening of Library 1 + 2. A) Experimental timeline for biosensor library screening of combined Library 1 + 2 sensors in animals with or without intestinal inflammation induced by DSS administration. B) Colonisation levels in all mice were comparable. C) Fractional weight change of mice and D) colon lengths measured at the experimental endpoint on day 6 post DSS administration/ day 2 post bacterial administration were typical of DSS induced inflammation. Graph shows mean ± standard deviation and results from an unpaired T-test, p-value: 0.0075. E) Sensor activation of the top-ranked DSS responsive, in vivo responsive, and control sensors as averaged FOR across all mice on day 2 post library administration. The full dataset is shown in Supplementary Figure 6. F) individual sensor activation (FOR) from each animal for top-ranked inflammation responsive sensors. Graphs also show mean. G) Predicted regulation of E. coli P_{spy}. H) Response of DTR306 Ec spy sensor when exposed to different inducers in vitro. Graphs show mean. Statistics calculated by one-way ANOVA between groups as shown, with Dunnet's multiple comparisons adjustment. Adjusted p-values: ** = 0.001, **** = <0.0001, NS = not significant I) Experimental timeline to test DTR306 Ec spy sensor in mice with or without intestinal inflammation induced by 3% DSS administration. J) Colonisation levels between groups and animals remained high across the days following sensor bacteria administration. K) Fractional weight and L) colon

lengths on day 6 post DSS administration/ day 2 post bacterial administration were indicative of DSS-induced inflammation (P value: 0.0988 by unpaired t-test). M) Response of DTR306 Ec spy sensor by plating of faecal contents on days 1 and 2 post bacterial administration showed elevated response in DSS-treated animals. Graphs show mean. Statistics calculated by one-way ANOVA between groups as shown, with Šídák's multiple comparison adjustment. Adjusted p-values: * = 0.0195, *** = 0.0003. N) Plating of contents from different intestinal regions at experimental endpoint on day 2 post bacterial administration confirms elevated sensor memory response in all tested regions within the DSS-treated gut.

To account for the ~10-fold increase in barcode abundance in the combined library compared with Library 1 alone (Supplementary Figure 2B-C, Data File 1), increased volumes of faecal pellet supernatants (corresponding to 30 – 40 mg worth of faecal pellet per mouse, and approximately 10^7 CFU of engineered bacteria as estimated from colonisation levels) were inoculated into outgrowths before splitting into +/- spectinomycin cultures. Following barcode sequencing, fractional odds ratios were then calculated at the sensor level to account for the expected stochasticity of the lower abundance barcodes within the recovered fractions.

Several sensors showed elevated FOR in response to either the mouse gut or the DSS-treated gut specifically (Figure 4E, full dataset: Supplementary Figure 6A). The top 5 DSS responsive sensors included two *E. coli torC* promoter variants (with wt or synthetic 5'UTR MCD10), likely activated by trimethylamine-N-oxide [26]; *E. coli yeaR* promoter, potentially activated by nitrate, nitrite or nitric oxide [27]; *E. coli spy* promoter, reported to be regulated by BaeR and CpxR [28,29] and previously identified as being upregulated in response to DSS treatment [8]; and, ST TCS1* (TtrRS-P_{ttrBCA}) (see Figure 2), previously implicated in tetrathionate sensing during inflammation [6] (Figure 4E-F). As expected, the control strain, Ec *ynfE15*, showed activation in all animals, with stronger sensor activation in faecal samples from day 2 (Figure 4E, Supplementary Figure 6A-B). Confirming our previous results (Figure 3), top-ranked sensors from Library 1 that activated in all animals included Cr TCS7 (QseCB-P_{ygiW}) and Cr TCS2 (DpiAB-P_{citC}). Additional gut-responsive sensors of interest identified in this screen include the promoter region of *E. coli yhcN*, reported to be induced by cytoplasmic acid response [30], Cr TCS12 (*C. rodentium* NarXL TCS and diverging promoter region to the NarK homologue), and the promoter regions of *E. coli gatA* and *pspA* genes (Figure 4E, Supplementary Figure 6A-B).

Of the top-ranked inflammation responsive sensors, Ec *spy*, showed particularly low response in control animals and consistent response across DSS-treated animals (Figure 4F). To explore its induction behaviour in the context of our memory circuit more extensively, we therefore cloned and tested the sensor individually (DTR306). DTR306 cultures were grown for 4 hours in aerobic SOC broth in the presence of previously reported inducers of the *spy* promoter: EtOH (5%), zinc (0.6mM) and copper (0.6mM) (Figure 4H) [28,31]. Colony screening on indicator plates demonstrated clear induction of DTR306 by ethanol and zinc, but not copper. Neither did we see *in vitro* induction of DTR306 in the presence of 3% DSS, indicating the substance itself is not inducing sensor activation (Figure 4H).

To confirm *in vivo* sensor behaviour, DTR306 was administered by oral gavage as an individual strain to mice. Animals (n = 4 per group) were treated with or without 3% DSS in drinking water for four days (~10^{11} bacteria/ mouse) (Figure 4I), and streptomycin was provided (5mg/ mouse by oral gavage) 24 hours prior to DTR306 administration. Colonisation levels were high across the 2 days post administration in both groups (Figure 4J). Weight loss throughout the experiment was indicative of the impact of DSS treatment (Figure 4K). Average colon length was also reduced in DSS-treated mice, albeit without statistical significance (Figure 4L). Analysis of DTR306 response via growth on indicator plates demonstrated consistent increase in switching specifically in DSS-treated mice on both days (Figure 4M), consistent with the library -based screening results. Growth of DTR306 from endpoint dissection of the gut also demonstrated consistent switching in all regions of the gut (Figure 4N). Together, these results demonstrate the potential for the library-based construction and screening approach to identify compatible inducers for the development of memory biosensors responsive to a multitude of conditions.

314 **Discussion:**

315 Our work expands the use of a robust bacterial memory circuit to simultaneously evaluate
316 hundreds of uniquely barcoded biosensor strains in the murine gut or *in vitro* culture. By incorporating a
317 DNA barcoding strategy, using a single, hyper-variable oligonucleotide for barcode delivery, we generate
318 large libraries of diversly sized sensors in a low-cost, single-pot reaction (Figure 1). Barcoding also facilitates
319 iterative addition, combination and revision of libraries over time, as demonstrated throughout this current
320 project. The successful identification and validation of several inducible circuits, both novel and previously
321 identified, highlights the effectiveness of our pipeline in facilitating biotechnology tool development (Figure
322 2-4).

323 Novel sensor development is an area of need for synthetic biology in the quest to expand the
324 toolbox of health- and disease- relevant inducible circuit components. Barcoded HTMS analysis is
325 compatible with any transcriptional activation-based sensing circuits that function in the commensal *E. coli*
326 host. This is a particular strength of our method, sourcing and testing diverse heterologous sensing
327 components to drive memory circuit activity. By comparison, another powerful parallelised bacterial
328 memory approach, RecordSeq [8,32], provides a transcriptome-wide view of chassis cell gene expression, akin
329 to a recording of an RNAseq dataset over time. The flexibility of GGA for allowing simultaneous assembly of
330 multiple fragments, facilitates future expansion of the approach to more complicated sensor designs, for
331 example non-grouped TCSs, or those that regulate non-proximal or multiple promoters throughout the
332 genome. Furthermore, use of alternative type IIS restriction enzymes can be used to expand cloning
333 compatibility of sensors, as demonstrated here during Library 2 construction which used both SapI and BsaI
334 enzymes.

335 The cloning strategy afforded libraries of thousands of uniquely barcoded strains, encoding
336 hundreds of variant sensors. Not all targeted sensors were successfully cloned, with lack of PCR
337 amplification the greatest source of loss within the pipeline. This is indicative of the challenges of
338 undertaking PCR primer design and reactions in bulk using diverse bacterial genomes. Of the amplified
339 sensors, >80% were successfully assigned to at least 1 barcode in the library. A fraction of the sensors were
340 constitutively ON in all tests, which is expected in a memory system for any gene promoters that are
341 routinely activated during *in vitro* culture conditions. In some cases, this may also result from a lack of
342 native regulation and thus high basal expression of cI^{DN} in the heterologous host. TCSs derived from species
343 more closely related to the *E. coli* NGF-1 host were, unsurprisingly, more likely to show signs of
344 functionality, consistent with systematic assessment of porting non-native genetic systems into new hosts
345 [33]. Many promoters are controlled by combinations of TCSs, and most TCSs regulate gene expression from a
346 variety of targets [34], suggesting that cross-regulation between host and heterologous TCSs and gene
347 expression is both likely and necessary. Tetrathionate sensing, driven by TtrRS regulation of the adjacent
348 *ttrBCA* promoter from *S.* Typhimurium [6] or *E. tarda* (Figure 2F) requires cross-regulation by host *E. coli* FNR.
349 Our experiments with truncated *Cr* TCS2 sensor variants also demonstrated cross-regulation by host *E. coli*
350 DpiAB, NarL and cAMP:CRP (Supplementary Figure 5C). Future libraries may take this systems biology
351 perspective into account by focussing on candidate TCSs from closely related species. Together, these
352 challenges point to the power of library screening approaches for rapid sensor identification and validation
353 within specific engineering biology contexts.

354 During *in vivo* library screening experiments, population bottlenecking caused barcode and strain
355 diversity loss following administration to the gut, ultimately limiting experiment length (Supplementary
356 Figure 3). Barcoded *C. rodentium* libraries have showed similar dynamics [21,22], indicating that bottlenecking
357 is likely an inevitable physiological constraint. Nevertheless, our colonisation data suggests that library
358 screening approaches are feasible and valuable over short experimental periods (eg. up to 2 days) to
359 simultaneously test thousands of uniquely barcoded strains within the gut. Testing directly in the λ-memory
360 system encoded in mouse commensal *E. coli* NGF-1, which has previously demonstrated low cellular
361 burden, long-term stability, and function over many months in the murine gut [5,6,11] allows for rapid
362 progression from screening to follow up tests as a single sensor in animals where these bottlenecking issues
363 are no longer a concern (eg. Figure 2-4).

364 This pipeline not only enables the construction of biosensors tailored to the gut environment but
365 also offers the potential for screening biosensors in other microbial growth permissible and potentially
366 inaccessible environments. The ability to build and test new whole-cell biosensors at high throughput is a
367 powerful method for the identification of novel inducible synthetic circuit components and prototype
368 disease diagnostics.
369
370 **Methods:**

371 *Media and culture conditions:*

372 Bacterial cultures of individual strains were grown at 37˚C in LB broth or on LB agar plates. Library
373 cultures were grown in LB peptone special (LBPS) containing 10 g/L peptone (Sigma product no. 68971), 10
374 g/L NaCl, and 5 g/L yeast extract in deionized water. LBPS was used because it does not contain lactose so
375 avoids the potential for biased outgrowth of ON sensors expressing *lacZ*. For anaerobic growth, cultures
376 were grown statically in pre-reduced media in a BD GasPak EZ container system with container system
377 sachets to generate the anaerobic environment (BD, product no. 260001) or using an Invivo$_2$ workstation
378 with I-CO_2N_2IC gas controller fitted for anoxic conditions (Baker Ruskinn). For memory response
379 quantification and blue-white screening, cultures were plated on LB agar plates containing 60 µg/mL 5-
380 bromo-4-chloro-3-indolyl-β-d-galactopyranoside (X-gal; Thermo Fisher Scientific catalogue no. R0402).
381 Antibiotics in media for bacterial growth were used at the following working concentrations; streptomycin
382 (Sigma-Aldrich, S650) 200 µg/mL, chloramphenicol (Sigma-Aldrich, C0378) 25 µg/mL, carbenicillin
383 (ThermoFisher Scientific, BP2648) 100 µg/mL, and spectinomycin (MP Biomedical, 158993) 50 µg/mL.

384 Individual biosensor inductions *in vitro* were conducted in SOC media (2% Tryptone, 0.5% yeast
385 extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) with the specified
386 concentration of inducer: potassium tetrathionate (Sigma-Aldrich, P2926), sodium citrate tribasic dihydrate
387 (Sigma-Aldrich, 71406), sodium nitrate (VWR, 27955), copper chloride dihydrate (Sigma-Aldrich, C6641), or
388 zinc sulphate (Sigma-Aldrich, Z0251). SOC media was used for individual sensor testing for comparability
389 with previous individual memory sensor testing (Naydich 2019, Riglar 2017). M9 (1x Difco M9 Minimal Salts
390 (Scientific Laboratory Supplies,: 248510), 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.4% filter sterilised glucose, 0.2%
391 casamino acids (Fisher Scientific Fisher,12747109) and 1µg/mL niacinamide) media was used for *in vitro*
392 library screening.

393 *Computational pipeline for TCS sequence extraction:*

394 A custom Python script was used to extract TCS DNA sequences from the Microbial Signal
395 Transduction Database (MiST 3.0) [35] and NCBI RefSeq databases. All HK and RR TCS signalling genes were
396 extracted from MistDB for a given bacterial strain and pairs of partner HK and RR were identified by
397 proximity (<200bp apart) and orientation (both genes must be on the same strand). Genes neighbouring
398 the paired TCS were considered likely regulated targets if they were encoded adjacent to, and on the
399 opposing strand of, the HK and RR pair. The strain's DNA sequence and annotations were automatically
400 downloaded from Refseq's FTP site. The TCS's gene IDs from MistDB were then matched to the RefSeq
401 annotations and the combined DNA sequence for the TCS and intergenic region between the TCS and its
402 putative target were extracted. Next, the combined TCS & Target sequences were screened for SapI or BsaI
403 sites, and, if absent, were deemed 'SapI or BsaI compatible'. Primers for compatible TCSs were designed
404 using Primer3 (in Geneious Prime software version 2020.1.1-2023.1.1; Biomatters Ltd) with SapI or BsaI
405 overhangs for Golden Gate Assembly (GGA). A full list of predicted grouped TCS amplicons (including HK, RR
406 and intergenic region) and primers used for TCS amplification is provided (Data File 1).

407 *TCS Library construction:*

408 Additional oligonucleotide and primer sequences used for library construction and sequencing
409 (Table 2) and bacterial strains used throughout the study (Supplementary Table 3) are provided.

410

411

412 **Table 2:** Primers and DNA fragments

Name	Sequence
Ultramer DNA Barcode cassette	ATGCTCTTCTAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNN

415 Purified bacterial gDNA was sourced from Public Health England (NCTC 10396 *Edwardsiella tarda*
416 and NCTC 11829 *Blautia producta*), by donation from other laboratories, or, for *Citrobacter rodentium*,
417 purified using the Gentra Puregene Yeast/Bacteria gDNA extraction kit (Qiagen catalogue no. 158567). TCSs
418 were amplified by PCR using Q5 High-Fidelity DNA polymerase (New England Biolabs), and amplified PCR
419 fragments were purified by column purification using the Monarch PCR Cleanup Kit (New England Biolabs)
420 following the manufacturer's protocols.

421 DNA barcodes were synthesised as an Ultramer DNA Oligo (IDT) consisting of a 106bp variable
422 nucleotide region flanked by Illumina sequencing primer binding sites, allowing direct sequencing of the
423 barcode region, and SapI restriction sites, for GGA assembly (Table 2). The DNA barcode oligo was made
424 double stranded by Q5 PCR with 5 cycles with 8 pmol of starting oligo template and 32 pmol of each
425 duplexing primer (primers p070 and p075 for SapI based GGA, and p079 and p100 for BsaI based GGA, in
426 Table 2).

427 The destination vector was pCR06 (for SapI assembly), a modified Tn7 transposon backbone derived
428 from pGRG36 [36] via pDR07 [11], or pDR07 (for BsaI assembly). Amplified sensors were pooled and assembled
429 with the double stranded barcode oligo into the Tn7 transposon machinery containing backbone (pCR06 or
430 pDR07) by multiplexed GGA using SapI or BsaI. SapI GGA reactions were set up using: 3nM of each DNA
431 fragment, 3μL(15 units) of SapI, 0.25μL(500 units) of T4 DNA Ligase, 2μL of T4 DNA Ligase buffer and H20 to
432 a final volume of 20μL. BsaI GGA reactions were set up using: 3nM of each DNA fragment, 1.5μL (30 units)
433 of BsaI-HFv2, 0.5μL (1000 units) of T4 DNA Ligase, 2.5μL of T4 DNA Ligase buffer and H_20 to a final volume
434 of 25μL. Cycling conditions for all GGA were: (37°C, 5 min and 16°C, 5 min) x 30, followed by 60°C, 5 min,
435 then storage at 4˚C. For each library, 4 μL of the GGA mix was transformed into 50 μL of ElectroMAX DH5α-E
436 competent cells (Thermo Fisher Scientific) by electroporation according to the supplier's protocol and
437 plated onto 5 carbenicillin selective LB-agar plates. An additional 1:100 dilution of the transformation
438 reaction was plated to estimate library size. All transformants were scraped from LB-agar transformant
439 plates using 1 mL of pre-warmed LB, vortexed to mix thoroughly, and frozen as a glycerol stock. A large
440 inoculum of the glycerol stock was grown in 5 mL of LB overnight, miniprepped and transformed into
441 electrocompetent E. coli NGF-1 encoding HTMS memory (PAS811). Transformant plates were scraped with
442 pre-warmed LB, pooled, diluted 1:1000 and grown for 6 hours at 30˚C in LBPS-streptomycin-carbenicillin-
443 chloramphenicol with 0.1% L-arabinose to induce Tn7-based genome integration. Cells were then diluted
444 1:1000 and grown overnight at 42˚C three times for plasmid curing. Libraries were frozen as 25% glycerol
445 stocks until use.

446 Individually cloned sensors (Ec yeaRWT, Ec yeaR23, Ec ynfE17, Ec torC10, Ec torC17, Ec torC23, Ec
447 torCWT, Ec hycAWT), control sensors (Ec ynfE15 and ST TCS1*), and fabR positive normalisation strains were
448 cloned using identical protocols to the library construction, with the exception of using individual sensor
449 PCRs for GGA reactions (Data File 1) and picking multiple separate (so separately barcoded) individual
450 colonies for plasmid integration and curing steps. Sensors and barcodes of individually cloned constructs
451 were verified and identified by sanger sequencing (Eurofins) (Data File 1). Correctly assembled individually
452 cloned sensors were combined with the pooled library at ~1:1000.

453 *Truncation mutant cloning:*

454 DpiA binding sites within the Cr TCS2 intergenic promoter region were identified by sequence
455 alignment to the corresponding genomic region in *E. coli* MG1655 within the Ecocyc database [17]. Plasmids
456 for truncated sensor mutants DTR304 and DTR305 were cloned by PCR amplifying appropriate regions
457 (using DTR304_Fwd and DTR304_5_Rev or DTR305_Fwd and DTR304_5_Rev respectively), assembled by
458 GGA using SapI along with unique barcodes, and transformed and genome integrated into PAS811 as above.

459 *Library sequencing and barcode assignment:*

460 For nanopore long read sequencing, the barcode and sensor regions of the biosensor libraries were
461 amplified using Q5 PCR (New England Biolabs) directly off 5 μL of glycerol stock using primers p117 and
462 p118 (Table 2). PCR products were purified using a Monarch PCR Cleanup Kit (New England Biolabs) and
463 prepared for nanopore long read sequencing using the Ligation Sequencing Kit (Oxford Nanopore
464 Technologies, SQK-LSK109). Library 1 was run on a Flongle flow cell (R9.4.1) and Library 2 was run on a

465 Minion flow cell (R10.4.1). Basecalling from Fast5 files was completed using MinKNOW software (version
466 22.12.5). Nanopore adapters were trimmed in Geneious (adapter sequence:
467 CCTGTACTTCGTTCAGTTACGTATTGCT) with 5bp allowed mismatches, minimum 10bp match to adapter and
468 0.1 error probability limit (regions with more than a 10% chance of error per base were trimmed).

469 Trimmed nanopore data was filtered for reads >400bp (a read length sufficient to at minimum
470 partially cover both the barcode and sensor regions), aligned to TCS reference sequences (excluding
471 barcode regions) of all successfully PCR amplified TCS using Minimap2 (version 2.17). Long read basepairs
472 that extend past the TCS reference sequence were trimmed using a custom python script and aligned to the
473 inferred barcodes from short read sequencing data (see below) using Bowtie2 [37]. Barcodes were assigned to
474 a specific sensor (Data File 1) if that sensor had both the highest number of trimmed long reads aligning to
475 the barcode, and highest fraction of aligning trimmed long reads normalised by total long reads for the
476 specific sensor, with a minimum of 5 long reads for assignment. Any barcodes with different assignments
477 between those values, or to which no trimmed long read data aligned were discarded for odds ratio
478 calculations.

479 For short read sequencing, genomic DNA (gDNA) was extracted from ± spectinomycin overnight
480 cultures manually using the Gentra Puregene Yeast/Bacteria gDNA extraction kit (Qiagen catalogue no.
481 158567) or on a FeliX liquid handler using the innuPREP DNA/RNA kit with an initial manual lysozyme step
482 for bacterial cell lysis. For the lysozyme bacterial cell lysis steps, 500µL of bacterial overnight culture was
483 pelleted, resuspended in 200µL of enzymatic buffer consisting of 20mM TrisCl pH 8, 2mM sodium EDTA,
484 1.2% Triton X-100, and 20mg/mL lysozyme and incubated at 37˚C for 30 minutes. Barcodes were first Q5
485 PCR amplified from extracted gDNA using primers p145 and p146 (Table 2), then purified by AMPure XP
486 beads at a ratio of 1.8x beads to sample following the manufacturer's protocol (Beckman Coulter). Purified
487 barcodes were then Q5 PCR amplified (using a modified protocol with 12 cycles) using IDT UDIs and purified
488 again by AMPure XP beads, as above. Final fragments for short read sequencing were quantified using Qubit
489 dsDNA BR assay kit (Invitrogen) following the manufacturer's protocol and pooled at equimolar
490 concentrations. Sequencing was undertaken on an Illumina NovaSeq sequencer (150bp Paired-end)
491 (Novogene), with an average of 1.4 million reads per sample.

492 *Short read sequence barcode analysis:*

493 Short read sequencing of library screening experiments was analysed using Dada2 [38] with a custom
494 R script to generate inferred barcodes and barcode counts for each sample (settings: truncLen=106, max
495 expected errors maxEE = 1, truncQ=2). Inferred barcodes were filtered to the expected barcode length of
496 106bp and clustered on 95% sequence similarity using the seq_cluster function from bioseq [39]. The exact
497 barcode sequence with the highest read number in each cluster was kept as the cluster's reference barcode
498 and read counts from other barcode members of the cluster added to the reference barcode.

499 *Fractional Odds ratio calculation and sample QC:*

500 Sensor activation is calculated using a fractional odds ratios defined for each sensor. First, the odds
501 ratio for each barcode in a sample is calculated as:

502 $$OR_{sensor} = \frac{BC_{+SPECT} / mean(PNB_{+SPECT})}{BC_{-SPECT} / mean(PNB_{-SPECT})}$$

503 Where BC_{+SPECT} and BC_{-SPECT} are reads from an individual barcode from samples grown with and without
504 spectinomycin respectively, and PNB_{+SPECT} and PNB_{-SPECT} are combined read numbers from several positive
505 normalisation barcodes in samples grown with and without spectinomycin respectively. Odds ratios were
506 initially calculated using the averaged results from four barcoded strains triggered by our previously used
507 100% positive normalisation promoter from *E. coli* MG1655, pFabR [11]. However, in certain samples we
508 observed variability in the outgrowth of spiked-in strains compared to sensors contained within the library,
509 which appeared to grow more slowly (Supplementary Figure 7A). We hypothesised this may result from the
510 differential growth conditions of normalisation strains and library prior to mixing, leading to differential lag
511 phases prior to growth. Thus, we developed a more robust normalisation strategy using 14 barcodes linked
512 to 3 sensors contained within the library (Bp TCS1, Bp TCS12 and Ef TCS3) that consistently showed odds

513 ratios ~1 when normalised using fabR strains, and which were subsequently validated as constitutively
514 switching ON by independent cloning and indicator plating (Supplementary Figure 7C). When samples were
515 normalised using these barcodes instead of the fabR barcoded strains it corrected skewness that could lead
516 to shifted odds ratio (Supplementary Figure 7B).

517 Because spectinomycin is bacteriostatic, reads from OFF sensors may remain in the culture at an
518 abundance determined by the fold change of the population's outgrowth. This will vary based on the
519 dilution factor used (typically 1:1000 for *in vitro* data, 1:100 for *in vivo* data) and *E. coli* cell density in the
520 original sample (which is variable). Therefore, we further calculated a fractional OR (FOR) that takes into
521 account observed differences in outgrowth, defined as:

522 $$FOR = \frac{OR_{sensor} - OR_{neg}}{OR_{pos} - OR_{neg}}$$

523 Where OR_{pos} is the average odds ratios of all positive control strains (Ef TCS3, Bp TCS12 and Bp
524 TCS1), and OR_{neg} is the average odds ratio of an internally selected negative control (OFF) strain. Lg TCS1
525 was selected as an internal negative control, as it was the sensor with >5 assigned barcodes that
526 consistently showed the lowest odds ratios across all analysed samples.

527 Finally, a set of QC parameters was developed to ensure any given sample would provide reliable
528 sensor activation quantification. Replicate barcodes of individual sensors within the library allowed
529 assessment of odds ratio variability caused by differences in bacterial starting concentration. Unsurprisingly,
530 low concentration samples display higher variability in their outgrowth, resulting in higher deviation from
531 the expected odds ratio of 1 for the 14 positive barcodes (Supplementary Figure 8G, H, I). Slope and R^2 of a
532 linear fit of OR of positive control strains are calculated to exclude samples with high levels of observed
533 outgrowth variation (thresholds set at an R^2 >0.6 and slope>0.5 and <2). We additionally set a conservative
534 rationally calculated sensitivity threshold. The libraries contain sensors with abundances across ranges of
535 orders of magnitude. We filtered barcodes to be above a set threshold of an abundance in the final
536 sequenced -spectinomycin sample of >1/5000 of the fraction of reads. This is because any fractions below
537 this threshold would indicate tens to a couple of hundred bacteria' cells being split between
538 +spectinomycin and -spectinomycin cultures (from a stationary phase *E. coli* culture containing 10^8 – 10^9
539 CFU/mL, a sensor at 1/5000 of the library is at a concentration of around $2*10^4$ – $2*10^5$ CFU/mL, after a
540 1:1000 dilution into ± spectinomycin cultures would be at 20 – 200 CFU/mL). To avoid any impacts from low
541 abundance positive normalisation barcodes, barcodes with <100 reads in ± spectinomycin cultures from a
542 given sample were excluded.

543 Due to the roughly 10-fold higher number of barcodes in Library 2, reads from all barcodes for a
544 specific sensor in these samples were pooled before OR, FOR and QC calculations.

545 *High throughput in vitro library screening with Tetrathionate:*

546 Approximately 15µL of a glycerol stock of Library 1 was inoculated into a 4mL LBPS overnight
547 culture with streptomycin, then back diluted 1:1000 into the relevant test condition in 0.5mL of LBPS in a
548 96-well deep-well plate and grown for 4 hours at 37˚C aerobically or anaerobically. For aerobic growth
549 plates were shaken at 1200 RPM. Liquid cultures of barcoded fabR strains for positive normalisation were
550 spiked in at 1:100 and cultures from deep well plates were then split and back diluted 1:100 into 0.5mL
551 cultures ± spectinomycin and grown overnight aerobically. Cells were pelleted and gDNA extracted using the
552 FeliX liquid handling robot (as above).

553 *In vivo library and individual strain screening:*

554 All animal experiments were carried out under approval of the local Ethical Review Committee at
555 Imperial College London according to UK Home Office guidelines. Experiments were conducted using 7-9
556 week old C57BL/6J female mice (Charles River UK). Upon arrival at the facility mice were given one week to
557 acclimatise to housing conditions and diet. All animals were sacrificed humanely by cervical dislocation at
558 the endpoint of the experiments. Mice were provided with Teklad 2019 Global Rodent Diet (Envigo). For all
559 experiments, 24 hours prior to engineered bacterial administration, mice were administered 5mg, 1mg or
560 0mg of streptomycin sulphate as specified (VWR, 0382-EU-50G) by oral gavage in 100µL PBS and confirmed

free of streptomycin resistant bacteria by faecal plating on streptomycin selective LB agar plates 24 hours post streptomycin oral gavage but prior to engineered bacterial delivery. Where specified, colitis was induced via administration of 3% dextran sodium sulphate in drinking water for the specified number of days before bacterial gavage.

For library screening experiments, libraries were grown overnight, washed in PBS and concentrated 10x (Library 1) or 5x (Library 1 + 2). 200μL of the washed library were then administered by oral gavage (~10^{10-12} cells per mouse). For sensor activation and bacterial load quantification, fresh faecal pellets were collected and stored immediately in 100μL of PBS, then resuspended to 100 mg/mL in PBS, vortexed for 2 mins, pulse centrifuged to remove large debris, before being diluted and plated to track bacterial load. The remainder of the faecal supernatant (up to a limit of 400 μL for Library 1) was inoculated into 4 mL streptomycin chloramphenicol LBPS and grown for 4 hours. For Library 1, positive normalisation fabR strains were also diluted 1:1000 from overnight cultures and grown for 4 hours, before being diluted 1:100 into 5mL of the faecal pellet outgrowths. After thorough mixing, cultures were diluted 1:100 into 10mL fresh LBPS, split into 2x 5mL culture, to which spectinomycin was added to one, and grown overnight. Overnight cultures were pelleted, and gDNA extracted and amplified for short read sequencing.

For testing Ec spy, mice were administered 3% DSS for the specified number of days before bacterial gavage. To test sensor activation, faecal pellets were collected in 100 μL of sterile PBS, then resuspended to 100 mg/mL, vortexed, pulse centrifuged to remove large debris and diluted and plated on indicator X-gal LB-agar plates for blue-white screening. To test sensor activation within different regions of the gut, at the endpoint of the experiment mice were dissected and the entire contents including mucus of the small intestine, cecum, proximal colon or distal colon were scraped, collected, resuspended to 100 mg/mL, vortexed, centrifuged, diluted and plated as for faecal pellets. The proximal colon was selected as the top half of the colon (from the bottom of the cecum to halfway to the anus), and distal colon is defined as the bottom half of this region.

Statistics and Sensor Rankings:

Statistical tests were performed in Graphpad Prism v9 or v10 software. For ranking responsive sensors in the Library 1 *in vivo* screen (Figure 3D) and Library 2 *in vivo* screen (Figure 4E), sensors were first filtered for those OFF in the gavage sample (FOR <0.25). To rank *in vivo* responsive For Library 1 screening (Figure 3), sensors were ranked by greatest difference between average faecal FOR across all mouse samples and gavage FOR. Sensors were then classified as 'top *in vivo* responders' if they had a response higher than the control Ec ynfE15. To rank inflammation responsive sensors in the Library 1 + 2 screen (Figure 4), sensors were ranked by difference in average faecal FOR in -treated animals (D2) and average faecal FOR in control animals (D2). Sensors were classified as 'top DSS responders'.

Data availability:

TCS DNA sequences, including intergenic regions up to neighbouring genes and primer sequences used to amplify the TCS, and library barcode sequences are available in Supplementary Data File 1.

Code availability:

All custom code used to analyse sequencing data or search relevant databases and extract TCS are publicly available at https://github.com/ClareRobin/Robinson2023_BiosensorsLibrary.

References:

1 Robinson, C. M., Short, N. E. & Riglar, D. T. Achieving spatially precise diagnosis and therapy in the mammalian gut using synthetic microbial gene circuits. *Frontiers in Bioengineering and Biotechnology* **10** (2022).

2 Riglar, D. T. & Silver, P. A. Engineering bacteria for diagnostic and therapeutic applications. *Nat Rev Microbiol* **16**, 214-225, doi:10.1038/nrmicro.2017.172 (2018).

3 Tanna, T., Ramachanderan, R. & Platt, R. J. Engineered bacteria to report gut function: technologies and implementation. *Curr Opin Microbiol* **59**, 24-33, doi:10.1016/j.mib.2020.07.014 (2021).

4 Barra, M., Danino, T. & Garrido, D. Engineered Probiotics for Detection and Treatment of Inflammatory Intestinal Diseases. *Frontiers in Bioengineering and Biotechnology* **8**, 265, doi:10.3389/fbioe.2020.00265 (2020).

5 Kotula, J. W. *et al.* Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci U S A* **111**, 4838-4843, doi:10.1073/pnas.1321321111 (2014).

6 Riglar, D. T. *et al.* Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat Biotechnol* **35**, 653-658, doi:10.1038/nbt.3879 (2017).

7 Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst* **1**, 62-71, doi:10.1016/j.cels.2015.06.001 (2015).

8 Schmidt, F. *et al.* Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **376**, eabm6038, doi:10.1126/science.abm6038 (2022).

9 Inda-Webb, M. E. *et al.* Sub-1.4 cm(3) capsule for detecting labile inflammatory biomarkers in situ. *Nature* **620**, 386-392, doi:10.1038/s41586-023-06369-x (2023).

10 Zou, Z. P., Du, Y., Fang, T. T., Zhou, Y. & Ye, B. C. Biomarker-responsive engineered probiotic diagnoses, records, and ameliorates inflammatory bowel disease in mice. *Cell Host Microbe* **31**, 199-212 e195, doi:10.1016/j.chom.2022.12.004 (2023).

11 Naydich, A. D. *et al.* Synthetic Gene Circuits Enable Systems-Level Biosensor Trigger Discovery at the Host-Microbe Interface. *mSystems* **4**, e00125-00119, doi:10.1128/mSystems.00125-19 (2019).

12 Cooper, R. M. *et al.* Engineered bacteria detect tumor DNA. *Science* **381**, 682-686, doi:10.1126/science.adf3974 (2023).

13 Ziesack, M. *et al.* Escherichia coli NGF-1, a Genetically Tractable, Efficiently Colonizing Murine Gut Isolate. *Microbiol Resour Announc* **7**, doi:10.1128/MRA.01416-18 (2018).

14 Shaw, C., Hess, M. & Weimer, B. C. Two-component systems regulate bacterial virulence in response to the host gastrointestinal environment and metabolic cues. *Virulence* **13**, 1666-1680, doi:10.1080/21505594.2022.2127196 (2022).

15 Jacob-Dubuisson, F., Mechaly, A., Betton, J. M. & Antoine, R. Structural insights into the signalling mechanisms of two-component systems. *Nat Rev Microbiol* **16**, 585-593, doi:10.1038/s41579-018-0055-7 (2018).

16 Williams, R. H. N. & Whitworth, D. E. The genetic organisation of prokaryotic two-component system signalling pathways. *BMC Genomics* **11**, 720, doi:10.1186/1471-2164-11-720 (2010).

17 Keseler, I. M. *et al.* The EcoCyc Database in 2021. *Front Microbiol* **12**, 711077, doi:10.3389/fmicb.2021.711077 (2021).

18 Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* **10**, 354-360, doi:10.1038/nmeth.2404 (2013).

19 Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Centre for Biotechnology Information; [1988] - Accession No. BAE19899.1, Edwardsiella tarda tetrathionate reductase complex response regulator [cited 2023 08 25] Available from: https://www.ncbi.nlm.nih.gov/protein/BAE19899.1.

20 Daeffler, K. N. *et al.* Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. *Mol Syst Biol* **13**, 923, doi:10.15252/msb.20167416 (2017).

21 Campbell, I. W., Hullahalli, K., Turner, J. R. & Waldor, M. K. Quantitative dose-response analysis untangles host bottlenecks to enteric infection. *Nat Commun* **14**, 456, doi:10.1038/s41467-023-36162-3 (2023).

22 Woodward, S. E. *et al.* Gastric acid and escape to systemic circulation represent major bottlenecks to host infection by Citrobacter rodentium. *ISME J* **17**, 36-46, doi:10.1038/s41396-022-01321-9 (2023).

23 Merighi, M. *et al.* Genome-wide analysis of the PreA/PreB (QseB/QseC) regulon of Salmonella enterica serovar Typhimurium. *BMC Microbiol* **9**, 42, doi:10.1186/1471-2180-9-42 (2009).

24 Lv, J. *et al.* The Role of the Two-Component QseBC Signaling System in Biofilm Formation and Virulence of Hypervirulent Klebsiella pneumoniae ATCC43816. *Front Microbiol* **13**, 817494, doi:10.3389/fmicb.2022.817494 (2022).

25 Lee, J., Hiibel, S. R., Reardon, K. F. & Wood, T. K. Identification of stress-related proteins in Escherichia coli using the pollutant cis-dichloroethylene. *J Appl Microbiol* **108**, 2088-2102, doi:10.1111/j.1365-2672.2009.04611.x (2010).
26 Baraquet, C. *et al.* TorT, a member of a new periplasmic binding protein family, triggers induction of the Tor respiratory system upon trimethylamine N-oxide electron-acceptor binding in Escherichia coli. *J Biol Chem* **281**, 38189-38199, doi:10.1074/jbc.M604321200 (2006).
27 Constantinidou, C. *et al.* A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as Escherichia coli K12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281**, 4802-4815, doi:10.1074/jbc.M512312200 (2006).
28 Srivastava, S. K., Lambadi, P. R., Ghosh, T., Pathania, R. & Navani, N. K. Genetic regulation of spy gene expression in Escherichia coli in the presence of protein unfolding agent ethanol. *Gene* **548**, 142-148, doi:10.1016/j.gene.2014.07.003 (2014).
29 Kwon, E., Kim, D. Y., Gross, C. A., Gross, J. D. & Kim, K. K. The crystal structure Escherichia coli Spy. *Protein Sci* **19**, 2252-2259, doi:10.1002/pro.489 (2010).
30 Kannan, G. *et al.* Rapid acid treatment of Escherichia coli: transcriptomic response and recovery. *BMC Microbiol* **8**, 37, doi:10.1186/1471-2180-8-37 (2008).
31 Bury-Mone, S. *et al.* Global analysis of extracytoplasmic stress signaling in Escherichia coli. *PLoS Genet* **5**, e1000651, doi:10.1371/journal.pgen.1000651 (2009).
32 Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380-385, doi:10.1038/s41586-018-0569-1 (2018).
33 Wang, G. *et al.* CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat Microbiol* **4**, 2498-2510, doi:10.1038/s41564-019-0573-8 (2019).
34 Rajeev, L., Garber, M. E. & Mukhopadhyay, A. Tools to map target genes of bacterial two-component system response regulators. *Environmental Microbiology Reports* **12**, 267-276, doi:10.1111/1758-2229.12838 (2020).
35 Gumerov, Vadim M., Ortega, Davi R., Adebali, O., Ulrich, Luke E. & Zhulin, Igor B. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic Acids Research* **48**, D459-D464, doi:10.1093/nar/gkz988 (2020).
36 McKenzie, G. J. & Craig, N. L. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol* **6**, 39, doi:10.1186/1471-2180-6-39 (2006).
37 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
38 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).
39 Keck, F. Handling biological sequences in R with the bioseq package. *Methods in Ecology and Evolution* **11**, 1728-1732, doi:https://doi.org/10.1111/2041-210X.13490 (2020).

Acknowledgements:

We would like to thank R Jackson, M Crone, N Matthews, and K Rekopoulou for assistance with sequencing and sequencing analysis; M Tran and K Jensen for assistance with automation; G Frankel, J Marchesi, C Mullineaux Sanders, L Roberts and D Chrystomou, for gifts of, and assistance with, genomic DNA. PAS811 was a gift from P Silver. This research was funded by a Sir Henry Dale Fellowship to D.T.R (211230/Z/18/Z). C.M.R. was funded by a President's PhD Scholarship through Imperial College London.

Author contributions:

D.R., C.M.R and D. C. conceived of the study and experimental designs. D.R. and C.M.R. wrote the manuscript. All authors contributed to the editing of the manuscript. T.W. and D.R. aided library construction design, and C.M.R finalised and cloned all sensors. C.M.R. performed sequencing and computational analyses. C.M.R and D.C. performed the animal experiments. Y.C. built and characterised the C. rodentium DpiAB-p_{citc}C variants.

Competing interests:
The authors declare no competing interests.