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-resident microbes encode a variety of systems capable of sophisticated sensing and conditional gene expression control within the mammalian gut during health and disease. Bacterial synthetic biology approaches, such as whole-cell biosensing and 'sense-and-respond' therapeutics, aim to re-purpose such systems for research and clinical applications; however, the toolbox of well characterised *in vivo* and disease-related inducible systems remains limited. Here, we dramatically expand the flexibility and power of a high-throughput approach to biosensor evaluation using bacterial memory circuits encoded in a gut commensal *E. coli* species. We construct a novel library of bacterial two-component system (TCS) based biosensors sourced from gut pathogens, tagged with unique DNA barcodes using a single-pot construction method. Using our pipeline, we evaluate sensor activity and performance heterogeneity across *in vitro* and *in vivo* conditions leading to novel sensor discovery. In the process, we optimise library construction, sample quality thresholds, and parameters for engineered bacterial biosensor screening in the murine gut. This approach can be applied to transcriptionally activated sensing elements of any type and will allow rapid development of new biosensors that can advance synthetic biology approaches for complex environments, such as the gut.

The bacteria that make up the gut microbiome are increasingly recognised as key responders to, and influencers of, local and systemic elements of health and disease. A longstanding vision for synthetic biology is the development of bacterial 'sense-and-respond' live biotherapeutic products, capable of autonomously inducing synthetic circuits under certain spatial or disease conditions (Riglar and Silver, 2018; Robinson et al., 2022). However, the catalogue of biosensing circuits with well demonstrated and understood functionality in the mammalian gut environment remains limited (Barra et al., 2020; Tanna et al., 2021), failing to take advantage of the microbiome's extensive range of sensing capabilities (Gumerov et al., 2020; Guo et al., 2020).

In recent years, synthetic memory circuits, capable of converting transient transcriptional activity into sustained responses, for example through induction of transcriptional switches (Kotula et al., 2014; Riglar et al., 2017) or DNA editing (Mimee et al., 2015; Schmidt et al., 2022 Inda-Webb, 2023, Nature; Zou et al., 2023), have proven particularly powerful for the assessment of bacterial function within the gut. We previously demonstrated the potential for a synthetic memory circuit based on the λ-phage lysis-lysogeny switch (Kotula et al., 2014) to simultaneously screen activity of libraries of candidate *E. coli* promoters within the mouse gut (Naydich et al., 2019). In this high-throughput memory system (HTMS), the 'memory OFF' state is characterised by expression of *cI*, encoding the λ-repressor that represses expression of other components of the circuit (Figure 1A, left). Production of a dominant negative cI variant (cI^{DN}), driven by a modular trigger element, switches cells into the 'memory ON' state, activating expression of *cro, lacZ and aadA* genes which encode i) the cro repressor to repress *cI* and maintain cells in the 'memory ON' state, ii) a β-galactosidase enzymatic reporter, and iii) a streptomycin/spectinomycin adenyltransferase which

confers resistance to these aminoglycosides, respectively (Figure 1A, right). Differential selection of activated from non-activated sensors within a pooled library can therefore be achieved by growth in the presence of spectinomycin, with sensors identified by massive parallel sequencing (Figure 1B) (Naydich et al., 2019). Owing to memory maintenance and the use of a mouse commensal *E. coli* NGF-1 as a host (Kotula et al., 2014), sensors can be tested in inaccessible locations such as during long-term colonisation of the murine gut via non-invasive faecal sampling (Riglar et al., 2017).

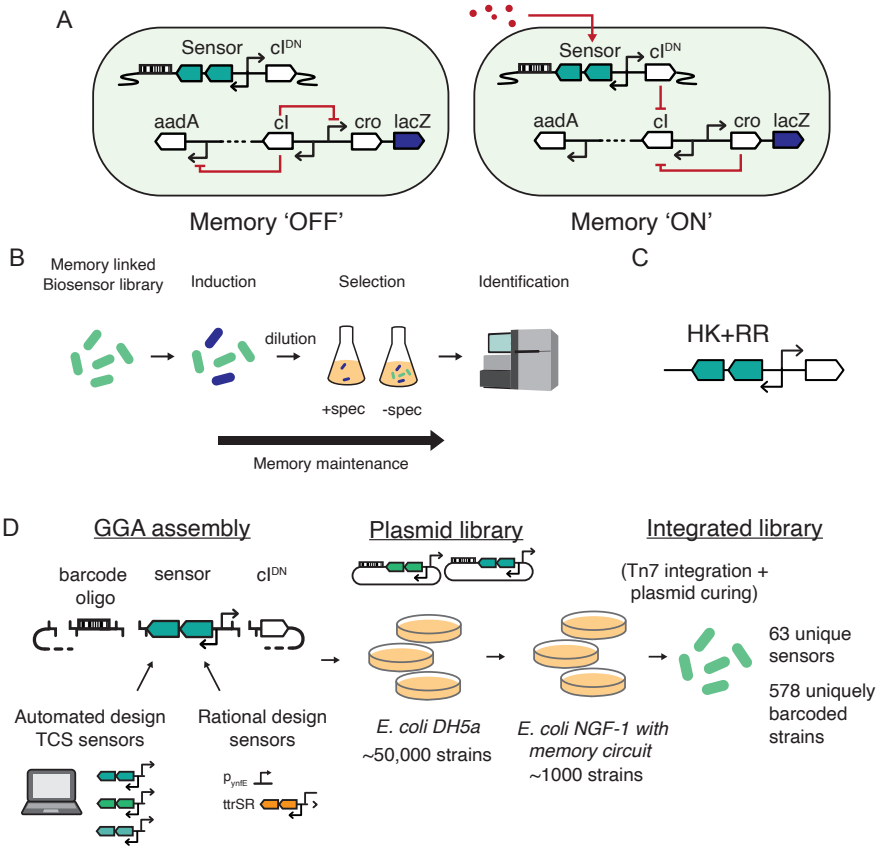

Figure 1: A high-throughput transcriptional screening pipeline based on an engineered memory circuit. A) A λ-phage based memory circuit including a bistable switch activated in the presence of cI^{DN} protein expression to de-repress cro and reporter gene expression (lacZ and aadA for screening and selecting respectively). Driving cI^{DN} expression from an inducible sensor affords biosensor capabilities. B) 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. C) 'Grouped' TCS arrangements are ideal for straightforward cloning of libraries encoding heterologous sensing components together with a probable regulated promoter to trigger memory formation. D) Library construction strategy for bulk assembled sensors included i) pooled GGA assembly of backbone, double-stranded barcode, and PCR-amplified sensor, followed by ii) plasmid library amplification and retrieval, iii) transformation into commensal E. coli PAS811, iv) genome integration using Tn7 transposon insertion, and finally v) plasmid curing.

One key limitation of this method was its inability to accommodate varied sizes of trigger/sensor insert within a single library, thus limiting the opportunity to screen a broad diversity of biosensing components. A major fundamental mechanism for bacterial biosensing is through two-component systems (TCSs): abundant, diverse, and modular signalling pathways capable of detecting a variety of molecular signals (Shaw et al., 2022). TCSs are typically composed of a membrane-bound histidine kinase (HK) and a partner response regulator (RR) that, once activated, alters transcription from target promoters through a DNA-binding output domain (Jacob-Dubuisson et al., 2018). We previously demonstrated the ability for a heterologously expressed TCS from *Salmonella enterica* subsp. *enterica* serovar Typhimurium (herein *S.* Typhimurium) to control the λ-memory circuit

through induction of the diverging promoter for a neighbouring operon (Riglar et al., 2017). In this way, bacteria were engineered to detect and record the presence of tetrathionate as an inflammatory biomarker in the gut.

To accommodate a wider variety of sensory components we now dramatically expand the power and flexibility of our λ memory-based transcriptional screening platform through inclusion of barcoded library construction and readout methods. This allows simultaneous screening of sensors with varying sizes, for example combined libraries of heterologous sensors encoding accompanying genes such as TCSs with shorter promoter-only sensors that rely on the host's endogenous regulation machinery. Pairing this advance with development of a computational pipeline for rapid identification of compatible TCSs from diverse microbiota, we demonstrate the identification of valuable sensing components using a combination of *in vitro* screening against specific chemical and growth conditions and *in vivo* testing in mouse models. Our work highlights the power of screening numerous uniquely barcoded sensor replicates, which allow simultaneous assessment of biosensor strain responses and response heterogeneity.

Materials and Methods

Media and culture conditions:

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

Individual biosensor inductions *in vitro* used the following liquid media: LB or SOC media (2% Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) with the specified concentration of inducer: potassium tetrathionate (Sigma-Aldrich, P2926), sodium citrate tribasic dihydrate (Sigma-Aldrich, 71406) or sodium nitrate (VWR, 27955).

For the *in vitro* library screen the following media and induction conditions were used: M9 media (1x Difco M9 Minimal Salts (Scientific Laboratory Supplies,: 248510), 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.4% filter sterilised glucose, 0.2% casamino acids (Fisher Scientific Fisher,12747109) and 1µg/mL niacinamide), Low potassium M9 (identical to M9 media but with modified salts: 240 mM Na_2HPO_4, 11 mM KH_2PO_4, 42.8 mM NaCl 93 mM NH_4Cl), Low phosphate M9 media (identical to M9 media but with modified salts: 24 mM Na_2HPO_4, 11 mM KH_2PO_4, 42.8 mM NaCl 93 mM NH_4Cl), SOC (as above), M9 with potassium tetrathionate at 1, 10 or 100 uM, M9 with sodium nitrate at 2mM or 20mM, M9 with 50 mM epinephrine bitartrate salt (ThermoFisher Scientific, 430140010), M9 with sodium citrate tribasic dihydrate at 10 or 50 mM (Sigma-Aldrich, 71406), M9 with 20 mM sodium fumarate dibasic (Sigma-Aldrich, F1506), M9 with 0.1M NaCl, M9 with 8% sucrose, M9 adjusted to pH5.

Computational pipeline for TCS sequence extraction:

A custom Python script was used to extract TCS DNA sequences from the Microbial Signal Transduction Database (MiST 3.0) (Gumerov et al., 2020) and NCBI RefSeq databases. All HK and RR TCS signalling genes were extracted from MistDB for a given bacterial strain and pairs of partner HK and RR were identified by proximity (<200bp apart) and orientation (both genes must be on the same strand). Genes neighbouring the paired TCS were considered likely regulated targets if they

were encoded adjacent to, and on the opposing strand of, the HK and RR pair (Figure 1C). The strain's DNA sequence and annotations were automatically downloaded from Refseq's FTP site. The TCS's gene IDs from MistDB were then matched to the RefSeq annotations and the combined DNA sequence for the TCS and intergenic region between the TCS and its putative target were extracted. Next, the combined TCS & Target sequences were screened for SapI sites, and, if absent, were deemed 'SapI compatible'. Primers for SapI-compatible TCSs were designed using Primer3 (in Geneious Prime software version 2020.1.1-2023.1.1; Biomatters Ltd) with SapI overhangs for Golden Gate Assembly (GGA). A full list of predicted grouped TCS amplicons (including HK, RR and intergenic region) and primers used for TCS amplification is provided (Data File 1).

TCS Library construction:

Oligonucleotide and primer sequences used for library construction (Supplementary Table 1) and bacterial strains used throughout the study (Supplementary Table 2) are provided.

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

DNA barcodes were synthesised as an Ultramer DNA Oligo (IDT) consisting of a 106bp variable nucleotide region flanked by Illumina sequencing primer binding sites, allowing direct sequencing of the barcode region, and SapI restriction sites, for GGA assembly (Supplementary Table 1). The DNA barcode oligo was made double stranded by Q5 PCR with 5 cycles with 8 pmol of starting oligo template and 32 pmol of each duplexing primer (primers p070 and p075 in Supplementary Table 1).

The destination vector was pCR06, a modified Tn7 transposon backbone derived from pGRG36 (McKenzie and Craig, 2006) via pDR07 (Naydich et al., 2019). Amplified TCSs were pooled and assembled with the double stranded barcode oligo into the Tn7 transposon machinery containing backbone (pCR06) by multiplexed GGA using SapI. GGA reactions were set up using: 3nM of each DNA fragment, 3µL(15 units) of SapI, 0.25µL(500 units) of T4 DNA Ligase, 2µL of T4 DNA Ligase buffer and H20 to a final volume of 20µL. Cycling conditions were: (37°C, 5 min and 16°C, 5 min) x 30, followed by 60°C, 5 min, then stored at 4C̊. 4µL of the assembled plasmid GGA mix were then transformed into 50µL of ElectroMAX DH5α-E competent cells (Thermo Fisher Scientific) by electroporation following the supplier's protocol, recovered in 500µL of SOC media and plated onto 5 ampicillin LB-agar plates (100 µL of transformed cells per plate). An additional 1:100 dilution of the transformation was plated for estimation of library size. In this case, the total plasmid library size after transformation was ~50,000 strains. All colonies were scraped from LB-agar transformant plates using 1 mL of pre-warmed LB, and thoroughly mixed and frozen as a glycerol stock. A large inoculum of the glycerol stock was grown in 5 mL LB overnight culture, miniprepped (Monarch Plasmid Miniprep Kit, New England Biolabs), and transformed into *E. coli* NGF-1 encoding HTMS memory (PAS811) by electroporation. 75ng of DNA was transformed into 200µL of competent PAS811 cells, approximately 1000 transformant CFUs were counted, indicating a library size of roughly 1000 strains before curing steps. Transformant plates were scraped, pooled, and grown for 6h at 30C̊ in LBPS-streptomycin-carbenicillin-chloramphenicol with 0.1% L-arabinose to induce Tn7-based genome integration. Cells were then diluted 1:1000 into 4mL LBPS-streptomycin-chloramphenicol and grown overnight at 42˚C three times with 1:1000 backdilutions for plasmid curing before the integrated library was frozen as a glycerol stock.

Individually cloned sensors (yeaR23, ynfE17, torC10, torC17, torC23, torCWT, hycAWT), control sensors (ynfE15 and ST1*), and fabR positive normalisation strains were cloned using identical protocols to the library construction, with the exception of using individual sensor PCRs for GGA reactions (Data File 1) and picking individual colonies for plasmid integration and curing steps (Supplementary Table 2). Barcodes of individually cloned sensors were identified by sanger

sequencing (Eurofins) (Data File 1). Individually cloned sensors were combined with the pooled library at ~1:1000.

Truncation mutant cloning:

DpiA binding sites within the Cr TCS2 intergenic promoter region were identified by sequence alignment to the corresponding genomic region in *E. coli* MG1655 within the Ecocyc database (Keseler et al., 2021). Plasmids for truncated sensor mutants DTR304 and DTR305 were cloned by PCR amplifying appropriate regions (using DTR304_Fwd and DTR304_5_Rev or DTR305_Fwd and DTR304_5_Rev respectively), assembled by GGA using SapI along with unique barcodes, and transformed and genome integrated into PAS811 as above.

Library sequencing and barcode assignment:

For nanopore long read sequencing, 5µL of glycerol stocked integrated library in *E. coli* NGF-1 was amplified directly using Q5 PCR (New England Biolabs), using primers p117 and p118 (Supplementary Table 1) and following manufacturer's protocols. PCR products were purified using a Monarch PCR Cleanup Kit (New England Biolabs) and prepared for nanopore long read sequencing using the Ligation Sequencing Kit (Oxford Nanopore Technologies, SQK-LSK109). The final prepared DNA was sequenced on a Flongle flow cell (FLO-MIN106D). Bases were called from Fast5 files using MinKNOW software (version 22.12.5). Nanopore adapters were trimmed in Geneious (adapter sequence: CCTGTACTTCGTTCAGTTACGTATTGCT), using 5bp allowed mismatches, minimum 10bp match to adapter, and 0.1 error probability limit (i.e. regions with more than a 10% chance of error per base were trimmed). Trimmed nanopore data was aligned to TCS reference sequences (excluding barcode regions) of all successfully PCR amplified TCS using Minimap2 (version 2.17). Reads aligned to 62 out of 82 successfully amplified TCS. All long read basepairs that extend past the TCS reference sequence were trimmed using a custom python script and aligned to the inferred barcodes from short read sequencing data (see below) using Bowtie2 (Langmead and Salzberg, 2012). Barcodes were assigned to a specific sensor (Data File 1) if that sensor had both the highest number of trimmed long reads aligning to the barcode, and highest fraction of aligning trimmed long reads normalised by total long reads for the specific sensor. Any barcodes with different assignments between those values, or to which no trimmed long read data aligned were discarded for odds ratio calculations.

For short read sequencing, gDNA was extracted from +spectinomycin and -spectinomycin overnight cultures manually using the Gentra Puregene Yeast/Bacteria gDNA extraction kit (Qiagen catalogue no. 158567) or on a FeliX liquid handler using the innuPREP DNA/RNA kit with an initial manual lysozyme step for bacterial cell lysis. For the lysozyme bacterial cell lysis steps, 500µL of bacterial overnight culture was pelleted, resuspended in 200µL of enzymatic buffer consisting of 20mM TrisCl pH 8, 2mM sodium EDTA, 1.2% Triton X-100, and 20mg/mL lysozyme and incubated at 37˚C for 30 minutes. Barcodes were first Q5 PCR amplified from extracted gDNA using primers p145 and p146 (Supplementary Table 1), then purified by AMPure XP beads at a ratio of 1.8x beads to sample following the manufacturer's protocol (Beckman Coulter). Purified barcodes were then Q5 PCR amplified (using a modified protocol with 12 cycles) using IDT UDIs and purified again by AMPure XP beads, as above. Final fragments for short read sequencing were quantified using Qubit dsDNA BR assay kit (Invitrogen) following the manufacturer's protocol and pooled at equimolar concentrations. Sequencing was undertaken on an Illumina NovaSeq sequencer (150bp Paired-end) (Novogene), with an average of 1.4 million reads per sample.

Short read sequence barcode analysis:

Short read sequencing of library screening experiments was analysed using Dada2 (Callahan et al., 2016) with a custom R script to generate inferred barcodes and barcode counts for each sample (settings: truncLen=106, max expected errors maxEE = 1, truncQ=2). Inferred barcodes were filtered to the expected barcode length of 106bp and clustered on 95% sequence similarity using the seq_cluster function from bioseq (Keck, 2020). The exact barcode sequence with the highest read number in each cluster was kept as the cluster's reference barcode and read counts from other barcode members of the cluster added to the reference barcode.

Odds ratio calculation:

Odds ratios were calculated using the following formula:

$$\frac{BC_{+SPECT}/mean(PNB_{+SPECT})}{BC_{-SPECT}/mean(PNB_{-SPECT})}$$

Where BC_{+SPECT} and BC_{-SPECT} are reads from an individual barcode from samples grown with and without spectinomycin respectively, and PNS_{+SPECT} and PNS_{-SPECT} are combined read numbers from several positive normalisation barcodes in samples grown with and without spectinomycin respectively. Barcodes with <10 reads in -SPECT were excluded from analyses, as were samples with <50,000 total reads.

Normalisation was first calculated using the averaged results from four barcoded strains triggered by our previously used 100% positive normalisation promoter from *E. coli* MG1655, pFabR (Naydich et al., 2019), spiked into cultures prior to outgrowth ± spectinomycin. However, in certain samples we observed variability in the outgrowth of spiked-in strains compared to sensors contained within the library, which appeared to grow more slowly (Supplementary Figure 1A-B). We hypothesised this may result from the differential growth conditions of normalisation strains and library prior to mixing, leading to differential lag phases prior to growth. Thus, we developed a more robust normalisation strategy using 14 barcodes linked to 3 sensors contained within the library (Bp TCS1, Bp TCS12 and Ef TCS3) that consistently showed odds ratios ~1 when normalised using fabR strains, and which were subsequently validated as constitutively switching ON by independent cloning and indicator plating. To avoid any impacts from low abundance positive normalisation barcodes, barcodes with <50 -SPECT or <25 +SPECT reads from a given sample were excluded. When samples were normalised using these barcodes instead of the fabR barcoded strains it corrected skewness that could lead to shifted odds ratio (Supplementary Figure 1B-C).

Replicates of individual sensors within the library allowed assessment of odds ratio variability caused by differences in bacterial starting concentration. Unsurprisingly, low concentration samples display higher variability in their outgrowth, resulting in higher deviation from the expected odds ratio of 1 for the 14 positive barcodes (Supplementary Figure 1D-E). We therefore set a quality spread threshold of an R^2 >0.7 on a linear fit of the positive normalisation barcode strains to eliminate samples with higher outgrowth variability and thus unreliable results.

High throughput in vitro library screening:

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

In vivo library and individual strain screening:

All animal experiments were carried out under approval of the local Ethical Review Committee according to UK Home Office guidelines. Experiments were conducted using 7-9 week old C57BL/6J female mice (Charles River UK). Mice were provided with Teklad 2019 Global Rodent Diet (Envigo). Mice were administered 5mg of streptomycin sulphate (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. The sensor library was grown overnight, washed in PBS and concentrated 1:10, with 200µL administered by oral gavage (~10^{10} cells per mouse). 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 was inoculated into 4 mL streptomycin+chloramphenicol LBPS and grown for 4 hours. 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 for citrate response *in vivo*, individual strains were prepared and gavaged as above. One day before administration of bacteria, mice were administered 5mg of streptomycin (as above) by oral gavage. 24h post bacterial gavage, 50 mM sodium Citrate (E331, Classikool Ltd) with 5% sucrose and ± 0.5 g/L streptomycin was introduced in their drinking water and continued throughout the experiment. Faecal pellets, collected and prepared as detailed above, were plated on X-gal indicator LB plates for blue-white screening.

Statistics

Statistical tests were performed in Graphpad Prism v9 software. For count data using blue-white screening, datapoints derived from less than 140 counted colonies due to limited colonies on plates are noted (corresponding to 95% confidence of MOE ~8% at 50% population proportion) and excluded if less than 50 (corresponding to 95% confidence of MOE >13% at 50% population proportion). For comparison of switching within library tests, a Mann-Whitney test was used to account for the non-normal distribution of datasets, with Bonferroni adjustment for multiple testing correction. For visualisation on log-scale graphs, we consider all sensors with OR <0.001 (*in vitro* data) or <0.01 (*in vivo* data) as OFF, and assign their value at this limit of detection.

Results:

Construction of bacterial biosensor libraries encoding diverse heterologous sensing machinery:

To build new libraries of putative biosensors with maximum potential to sense diverse molecules within the gut we took advantage of the diversity and genomic organisation of TCSs, which are catalogued within databases such as the microbial signal transduction database (MiST3) (Gumerov et al., 2020). Most HKs are encoded genomically within 200bp of their partner RR (Williams and Whitworth, 2010), and a proportion of these are known to regulate adjacent genes within the genome. Where a HK-RR pair are situated next to a diverging gene, we have termed this a 'grouped' TCS arrangement (Figure 1C). Manual inspection of the 14 grouped TCSs contained in the *E. coli* MG1655 genome using the Ecocyc database (Keseler et al., 2021) suggests that in 10 of 14 cases the divergent, adjacent gene is regulated in some way by the neighbouring TCS. This is therefore a convenient conformation for the construction of biosensor libraries that can expand on the natural sensing capacity of the chassis strain. Complete sensing machinery (HK, RR and diverging promoter sequence) can be extracted in a single PCR for cloning, and there is a strong likelihood that the expressed TCS will regulate the included promoter region, at least in its natural host.

To enable rapid library design, we developed a computational pipeline to identify grouped TCSs from available genomic sequencing data and extract DNA sequences for downstream primer design, PCR amplification and cloning (see methods). Using our pipeline directed against the genome sequences of 16 bacterial species for which we had access to gDNA of exact or closely related strains (Table 1), we identified 401 total TCS sequences (excluding hybrid and orphan TCSs), 153 of which were classified as 'grouped TCSs' (Table 1). Of these 113 were free of SapI restriction enzyme recognition sites making them compatible with our downstream cloning strategy (Table 1).

Identified TCSs were cloned so that the divergent promoter would drive cI^{DN} expression within a plasmid backbone containing Tn7 transposon machinery for genomic integration (Figure 1D) (McKenzie and Craig, 2006). Golden gate assembly also included a 106bp randomised nucleotide barcode flanked by sequencing adaptor sites, to uniquely identify all strains constructed in a single pot reaction. The hypervariability of these barcodes allows cloning of large libraries with essentially no chance of any single barcode labelling more than one construct. The library was integrated into a commensal mouse *E. coli*, *E. coli* NGF-1, encoding the λ-phage high-throughput memory system (PAS811) (Naydich et al., 2019). Sequencing of the library identified 54 unique TCSs (Table 1, Data File 1). Taking advantage of the flexibility afforded by the barcoding strategy, we supplemented these

strains with 9 additional 'curated' sensors that were identified in our previous studies (Riglar et al., 2017; Naydich et al., 2019), that were cloned individually and added to the library (Data File 1). Curated sensors included two controls: i) a *S.* Typhimurium *ttrSR*-P_{ttrBCA} sensor, edited with a synthetic ribosome binding site, (ST TCS1*) that previously showed response to the inflammatory biomarker tetrathionate (Riglar et al., 2017), and ii) the *E. coli* Nissle1917 $P_{ynfEFGH}$ promoter edited with a synthetic 5' untranslated region and ribosome binding site, MCD15 (Ec ynfE15) (Mutalik et al., 2013) that previously showed some switching under *in vitro* anaerobic growth conditions and complete switching in the murine gut (Naydich et al., 2019).

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

Strain	Total TCSs	'Grouped'	SapI compatible	Successfully amplified	In final library
Bacteroides caccae	9	5	5	4	2
Bacteroides fragilis	13	8	6	5	3
Blautia producta	90	26	19	14	5
Citrobacter rodentium	22	14	11	11	11
Clostridium difficile	43	8	8	3	1
Clostridium perfringens	20	5	3	1	1
Edwardsiella tarda	20	11	8	5	5
Enterococcus faecalis	13	4	3	3	3
Klebsiella oxytoca	25	12	8	2	1
Klebsiella pneumoniae	25	12	10	8	7
Lactobacillus gasseri	6	1	1	1	1
Lactobacillus plantarum	13	2	2	2	1
Ruminococccus gnavus	36	13	8	6	2
Salmonella typhimurium s. enterica	22	11	6	5	4
Vibrio cholera	*24*	*13*	*8*	*4*	*3*
Yersinia enterocolitica	*20*	*8*	*7*	*5*	*4*
Total	401	153	113	79	54

Barcode assignment:

Following cloning, barcodes were identified and associated with their linked TCS/promoter sequence by a combination of massive parallel long and short read sequencing (Supplementary Figure 2A). The final library contained 578 assigned unique barcodes, 550 of which were linked to TCSs generated through bulk construction and 28 of which were cloned individually and added to the library at known concentration. The number of barcodes identified per sensor range between 1 and 44 for the bulk cloned TCS sensors, with a median of 6 (Supplementary Figure 1B). An additional 270 barcodes were detected in the library but unable to be assigned to a sensor, either due to assignment strategy disagreement (28 barcodes) or the lack of available aligned long reads (242 barcodes). Assessment of the relative abundance of barcodes in three representative samples showed that unassigned barcodes were typically 3 to 5 orders of magnitude lower abundance within the library than assigned barcodes, meaning they make up a small fraction of the assembled library and explaining the absence of long-read sequence coverage (Supplementary Figure 1C). Indeed, 44 of the 242 unassigned barcodes were detected in less than 100 reads across all analysed samples. For comparison the average number of reads per assigned barcode across all library testing samples passing QC was 83,222 reads with a median of 13,234 reads.

In vitro validation of library biosensor screening:

To quantify the switching state of each sensor during library screening we compared the number of barcode sequencing reads for each barcode from samples split and grown as two cultures in the presence (+SPEC) or absence (-SPEC) of spectinomycin (Figure 1B). Only ON sensors will grow in the presence of spectinomycin due to *aadA* expression (Figure 1A). Degree of switching is represented as an odds ratio by normalisation to sequencing reads from known positive strains (see methods). An odds ratio close to 1 corresponds to an ON sensor, and close to 0 to an OFF sensor (Figure 2A-B). Sensors will regularly switch at partial levels, and empirically, odds ratio values as low as ~0.01 can be considered likely to be at least partially ON (Naydich et al., 2019). Because spectinomycin is bacteriostatic, reads from OFF sensors may remain in the culture at an abundance determined by the fold change of the population's outgrowth. This will vary based on the dilution factor used (1:1000 for *in vitro* data, 1:100 for *in vivo* data) and *E. coli* cell density in the original sample (which is variable). Therefore, we conservatively consider odds ratios <0.001 (*in vitro* data) or <0.01 (*in vivo* data) as OFF sensors in line with the dilution factor for each experiment type.

To demonstrate the functionality of the library screening approach, we cultured pooled sensors *in vitro* in the presence or absence of oxygen and at different concentrations of tetrathionate (Figure 2C). Various sensors showed differential activation between the conditions tested. As expected, ynfE15 sensors had consistently elevated odds ratios in anaerobic conditions (eg. aerobic M9: median OR = <0.001; anaerobic M9: median OR = 0.014) (Figure 2C-D). This mirrors previous results demonstrating partial switching of a similar sensor during *in vitro* anaerobic growth (Naydich et al., 2019). Also similar to previous results (Riglar et al., 2017), ST TCS1* sensors activated consistently in the presence of tetrathionate and absence of oxygen (eg. anaerobic 0 µM ttr: median OR = <0.001; anaerobic 100 µM ttr: median OR = 0.820) (Figure 2C, E). Interestingly, low level switching was detected in 100 µM tetrathionate conditions in the presence of oxygen (aerobic 100 µM ttr: median OR = 0.00583) (Figure 2C, E), which may indicate either a degree of aerobic response for this promoter or alternatively limitations in the aeration under culture conditions used. Together, these results demonstrate the functionality of the approach and ability to combine sensors of diverse size and origins.

Tetrathionate response variation

The 22 barcodes linked to *Edwardsiella tarda* TCS7 (Et TCS7) were also consistently induced in anaerobic conditions in the presence of tetrathionate, at levels comparable to ST TCS1* (eg. anaerobic 0µM ttr: OR = 0.000 (0.000, 0.000) median(IQR); aerobic 100 µM ttr: 0.000(0.000, 0.0153); anaerobic 100µM ttr: 0.874(0.745, 1.07)) (Figure 2C, E). While the available gene annotations for the strain used for TCS extraction (*E. tarda* KC-Pc-HB1) did not predict this behaviour, BLAST searches identified homology to a putative tetrathionate reductase and associated TCS in *E. tarda* strains Et54 (Protein Accession No. BAE19899.1) and FL6-60 (Daeffler et al., 2017). To validate these findings, EtTCS7 was re-cloned as an individual strain (DTR302) and its response to tetrathionate confirmed by plating of anaerobically induced cultures (100µM tetrathionate) on X-gal indicator plates for blue-white colony screening (Figure 2F). Confirming the library-based testing results, when grown anaerobically with a range of tetrathionate concentrations (0, 0.1, 0.5, 1, 10, 100 µM) both DTR302 and a strain triggered by ST TCS1* (DTR277) demonstrated almost complete induction in the presence of >1µM tetrathionate (Figure 2G). Together, these results demonstrate the identification of novel biosensors with specific sensing functions using our screening approach.

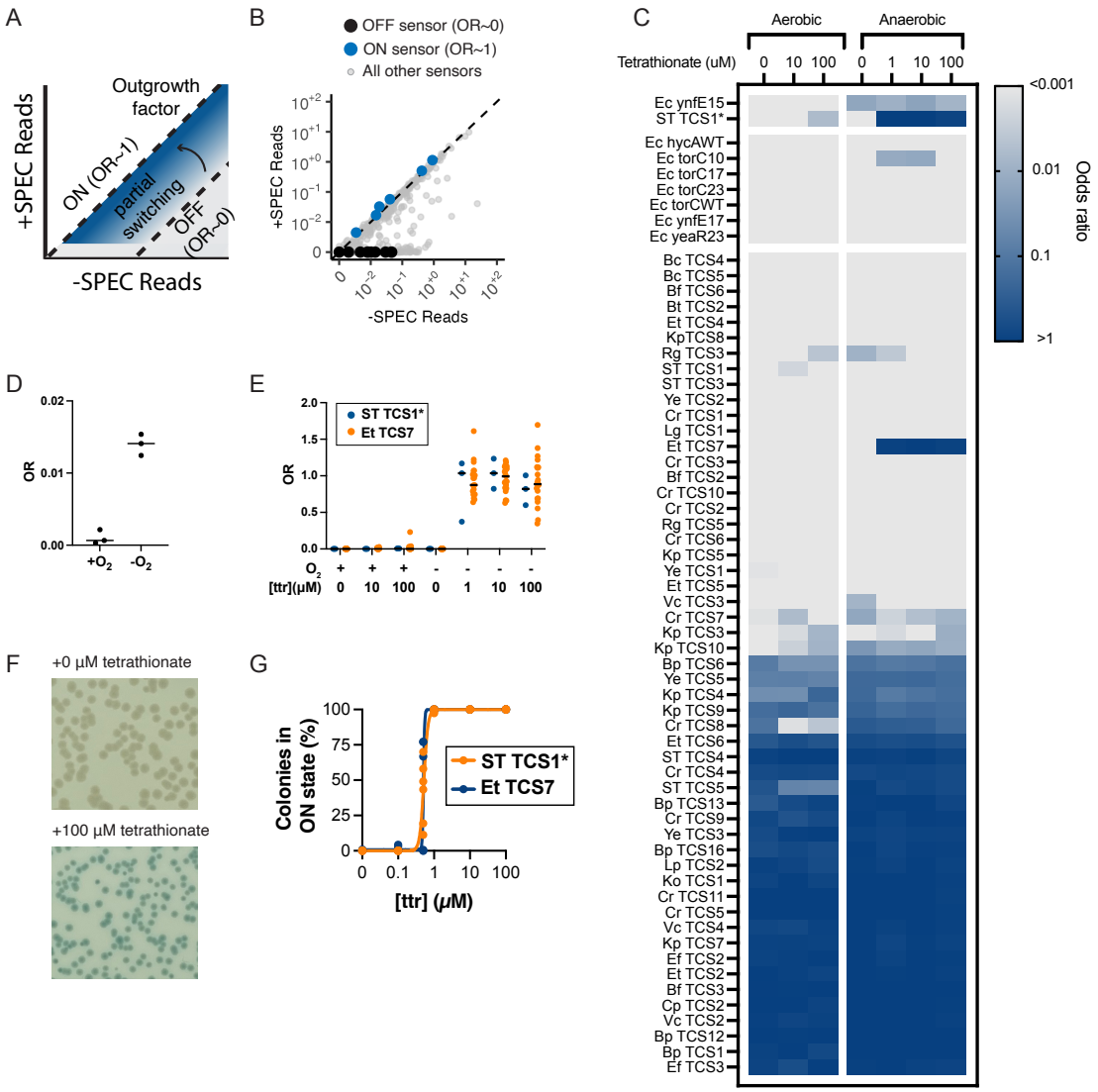

Figure 2: Pooled screening identifies novel inducible sensors. A) Schematic and B) normalised reads of example ON and OFF sensors, ON sensors will have equal growth in the plain (-SPEC) and spectinomycin selective (+SPEC) media, resulting in an odds ratio of ~1, OFF sensors will not grow in the spectinomycin selective media, resulting in an odds ratio of ~0. C) average log10 transformed odds ratio of all sensors within the library in the presence and absence of oxygen and varying levels of tetrathionate. D) Individual Ec ynfE15 sensors showed increased odds ratios under anaerobic conditions. E) Odds ratios of individual barcodes from Et TCS7, ST TCS1* under anaerobic conditions across all oxygen and tetrathionate combinations tested. F) An isolated Et TCS7 sensor showed high levels of switching when grown in anaerobic 100μM tetrathionate. G) Isolated Et TCS7 and ST TCS1* sensors both responded strongly to anaerobic growth across a range of tetrathionate concentrations when plated on indicator plates.

Impact of streptomycin pre-treatment on colonisation and barcode diversity.

The memory function of our sensors is a key advantage to this screening pipeline, allowing recording of transient signals from within the murine gut. We have previously demonstrated robust, long-term colonisation of the conventional mouse gut by *E. coli* NGF-1 strains, including, in select instances, without the need for any antibiotic pre-treatment - a common requirement to overcome colonisation resistance (Riglar et al., 2017). Interestingly, another mouse commensal *E. coli* strain, AZ-1, which is closely related to NGF-1, has also demonstrated favourable murine colonisation characteristics (Russell et al., 2022).

However, colonisation is strongly dependent on the background microbiome of animals which vary considerably between mouse vendors and mouse facilities. To test colonisation characteristics of our library in our current facility, we therefore administered the sensor library by oral gavage (~10^{10} bacteria/mouse) to groups of 2 healthy, conventional mice 24 hours following administration of streptomycin antibiotics (1mg or 5mg/mouse) or mock treated controls (Supplementary Figure 3A). Faecal samples were collected daily for colonisation level analysis from days 1-7 and day 11 for plated colony forming unit counts and on days 1, 4, 6 and 11 following bacterial administration for strain diversity analysis by barcode sequencing (Supplementary Figure 3A). Faecal bacterial counts correlated with the amount of streptomycin provided, with engineered bacteria in untreated animals dropping below detection limit at day 3, and in mice administered 1mg streptomycin on day 4-5 (Supplementary Figure 3B). Mice provided 5mg streptomycin became colonised over longer periods, in line with our previous results (Kotula et al., 2014; Riglar et al., 2017; Naydich et al., 2019). Barcode diversity in all animals was high 24 hours after bacterial administration but could not be measured due to limited colonisation on day 4 in untreated and 1mg streptomycin treated mice (Supplementary Figure 3C). Even in 5mg streptomycin treated mice, barcode diversity dropped quickly over the days following administration (day 1: 556 and 555 barcodes, day 4: 19 and 24 barcodes; Day 6: 10 and 9 barcodes and Day 11: 4 and 3 barcodes) (Supplementary Figure 3C). We therefore reasoned that experiments over short timeframes would be necessary to maintain sufficient barcode diversity for library sensor function analyses.

High throughput identification of novel biosensors in vivo:

We therefore administered our sensor library by oral gavage (~10^{10} bacteria/mouse) to cohorts of mice (n = 4) 24 hours following administration of 5mg streptomycin or to mock treated animals (Figure 3A). As previously, levels of engineered bacteria in faeces were lower in mock treated animals (Supplementary Figure 3E), although in this instance levels at d2 post administration remained several orders of magnitude higher than previously (Supplementary Figure 3B,E). The vast majority of barcodes and thus sensors were recovered from streptomycin treated animals on both days (day 1: 523 ± 19 barcodes and 65 ± 1 sensors; day 2: 523 ± 6 barcodes and 64 ± 3 sensors) (Supplementary Figure 3E-F). In contrast, by day 2 mock treated mice had lost 75% of the library barcode diversity, and nearly 50% of sensor diversity (day 1: 373 ± 186 barcodes and 58 ± 13 sensors; day 2 96 ± 41 barcodes and 38 ± 6 sensors). Samples from mock treated animals also showed considerably more variation, with only 1 of 8 samples passing our combined quality thresholds (see methods). We therefore analysed sensor response solely in samples from streptomycin treated mice.

Various sensors showed indications of elevated switching in animals compared to aerobic culture conditions (Figure 3B-F; Supplementary Figure 4A). Increases in odds ratio from barcoded strains linked to the *E. coli* ynfE15 sensor, which previously demonstrated complete switching within the murine gut (Naydich et al., 2019), were apparent from day 1, but more prominent on day 2 (Figure 3C). This is consistent with previous testing of individual sensors within the gut (Naydich et al., 2019). Thus, we focussed analyses on the response of sensors 48 hours after delivery to streptomycin treated mice (Figure 3B,D-F; Supplementary Figure 4A).

Statistical testing using adjustment for multiple testing (Mann Whitney test with Bonferroni adjustment) identified only two sensors with significantly higher response in mice than in gavage controls (Cr TCS7: corrected p = <0.0001, median OR = 0.0245 control and OR = 0.399 mouse D2; Cr TCS2: corrected p = 0.0189, median OR = <0.0100 control and OR = 0.0358 mouse D2) (Figure 3D-E). However, notably this does not include our previously validated sensor, Ec ynfE15 (Figure 3C), suggesting the stringency of this multiple comparison adjustment may lead to false negatives. Thus, we also considered another sensor with elevated OR at statistical significance equivalent to that of Ec ynfE15 for follow up: Ye TCS1 (OR = 0.009 (0.005, 0.16) control and 0.090 (0.029, 0.180) mouse D2) (Figure 3F).

C. rodentium TCS7 is identified as QseCB. QseC is an adrenergic receptor reported to be capable of sensing the host neurotransmitters epinephrine and norepinephrine to activate its cognate RR QseB (Moreira et al., 2016). *C. rodentium* TCS2 is identified as DpiAB. DpiAB is reported in

E. coli to regulate genes involved in anaerobic citrate fermentation (Kaspar and Bott, 2002; Yamamoto et al., 2008). *Y. enterocolitica* TCS1 is identified as CpxAR. CpXAR is associated with membrane stress response in gram negative bacteria (Cho et al., 2023). Together, these results demonstrate the importance of internal replicates afforded by barcoding to further identify robust sensors.

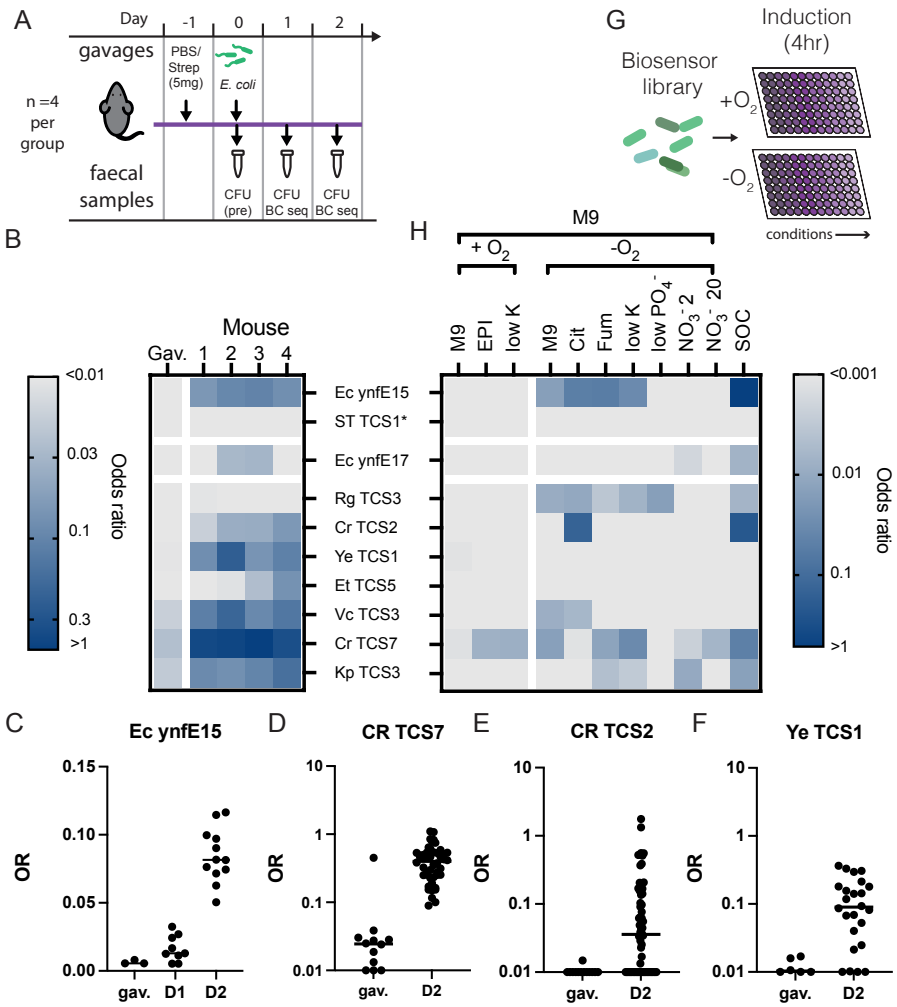

Figure 3: High throughput identification of novel biosensors *in vivo* and *in vitro* A) Set up of the *in vivo* screening of the library. B) Median odds ratio across sensors of the library screened in mice (n=4) pre-treated with 5mg streptomycin prior to administration. Data shown is from D2 post administration and for select sensors. Full dataset shown in Supplementary Figure 4. C) Ec ynfE15 individual barcode response for gavage and mice on day 1 and 2 post administration (response of each bc in 4 mice shown with median marked). D-F) Individual barcode response for gavage and mice day 2 post administration (response of each bc in 4 mice shown) for D) Cr TCS7, E)Cr TCS2, F) Ye TCS1/ Graphs also show median. G) Set up of *in vitro* screening of the library. H) Median odds ratio across sensors of library screened under a range of in vitro growth conditions. Data shown from select sensors and conditions. Full dataset shown in Supplementary Figure 4.

In vitro fingerprinting of biosensor response:

To further decipher the signals to which these, and other sensors within the library may be responding within the mammalian gut, we took advantage of the high-throughput characteristics of this method to perform *in vitro* fingerprinting by exposing the sensor library to a variety of metabolites and conditions potentially experienced in the gut: nitrate, citrate, epinepherine, fumerate, low pH, low potassium, low phosphate, high osmolarity (sucrose), and high salt (NaCl) in many cases both with or without oxygen (Figure 3G).

Several sensors showed differential response across the conditions tested (Figure 3H, Supplementary Figure 4B). Barcodes linked to the Ec ynfE15 sensor had elevated odds ratios during growth in various anaerobic conditions, especially anaerobic SOC, mirroring our previous findings with this sensor (Naydich et al., 2019) (Figure 3H, Supplementary Figure 4B). Also, as expected, the response of this sensor was repressed in the presence of anaerobic nitrate (Figure 3H, Supplementary Figure 4B). Lack of switching under a range of other anaerobic growth (pH, osmolarity, NaCl, limited potassium) may be indicative of these environments slowing growth of the *E. coli* NGF-1 host (Supplementary Figure 4B) given that growth of the bacterial cells is required for effective switching.

Validation of citrate response *in vitro* and *in vivo*:

Of the sensors showing elevated response in the murine gut, Cr TCS2 (*C. rodentium* DpiAB-p_{CitC}) was of note due to the specificity of *in vitro* signals that showed elevated OR, limited to anaerobic growth in the presence of sodium citrate (50mM) and anaerobic rich growth medium (SOC) (Figure 3H). This suggests a potential mechanism for induction of this sensor within the murine gut.

To further validate this response, we cloned and tested the Cr TCS2 sensor individually (DTR303). Growth of the sensor in anaerobic SOC media with or without citrate (50mM) followed by plating on X-gal indicator plates confirmed elevated switching in the presence of citrate (control: 0.1 $\pm$ 0.1% SEM; citrate: 41.4 $\pm$ 3.4% SEM) (Figure 4A). The absence of glucose (SOB + citrate: 6.3 $\pm$ 0.5% SEM) and presence of nitrate (20mM) (citrate + nitrate: 0.0 $\pm$ 0.0% SEM) suppressed response to citrate (Figure 4A), consistent with predicted negative regulation by cyclic-AMP: Catabolite Repressor Protein and activated NarL (Figure 4B). These results validate the findings of our library screen and confirm that the host *E. coli* NarL can cross-regulate the *C. rodentium citC* promoter region.*E. coli* NGF-1 natively encodes a TCS homologous to *C. rodentium's* DpiAB (DpiA 86.7% AA sequence identify; DpiB 87.2% AA sequence identity). To explore whether host *E. coli* DpiA may cross-react with the heterologous *C. rodentium citC* promoter, we therefore cloned truncated variants of Cr TCS2 lacking the HK and RR (DTR304) or lacking the HK and RR along with predicted dpiAB binding operators in the *C. rodentium citC* gene promoter (DTR305) driving bacterial memory (Figure 4B). DTR303, DTR304 and DTR305 were grown in anaerobic SOC in the presence or absence of citrate (10mM). DTR303 and DTR304 both showed low levels of switching in SOC only controls (14.8 $\pm$ 5.9% SEM and 1.3 $\pm$ 0.5% SEM, respectively) with significantly elevated response in the presence of citrate (85.3 $\pm$ 1.4% SEM and 83.5 $\pm$ 1.3% SEM) (Figure 4C). Thus, *E. coli* DpiA is able to regulate *C. rodentium* p_{citC} to equivalent levels as the *C. rodentium* heterologous gene. No switching was detected in sensors lacking dpiAB binding sites in the presence of citrate, confirming that requirement of DpiA for the observed response (SOC only 0.0 $\pm$ 0.0% SEM; citrate: 0.0 $\pm$ 0.0% SEM) (Figure 4C).

We next tested the response of DTR303 to citrate in the murine gut. Citrate can make its way to the gut from many sources including from citrus fruit and as a commonly used preservative in food. It can be metabolized by a variety of gut bacteria and may be implicated in competitiveness of *Vibrio cholerae* during infection (Liu et al., 2019). Streptomycin has been implicated in slight elevations to citrate levels within the gut (Liu et al., 2019). To avoid the potential for antibiotics to interfere with results, we therefore administered DTR303 bacteria to mice (n=4 per group) via oral gavage without prior streptomycin treatment (Figure 4D). One group of mice were administered 50mM sodium citrate in their drinking water throughout the course of the experiment, with faecal pellets collected for plating on indicator plates (Figure 4D). Colonisation levels were comparable between groups throughout the course of the experiment (Supplementary Figure 5A). We observed switching of a small but variable subset of sensor bacteria in citrate supplemented animals only (Figure 4E; Supplementary Figure 5B) consistent with citrate supplementation leading to elevated sensing in the gut.

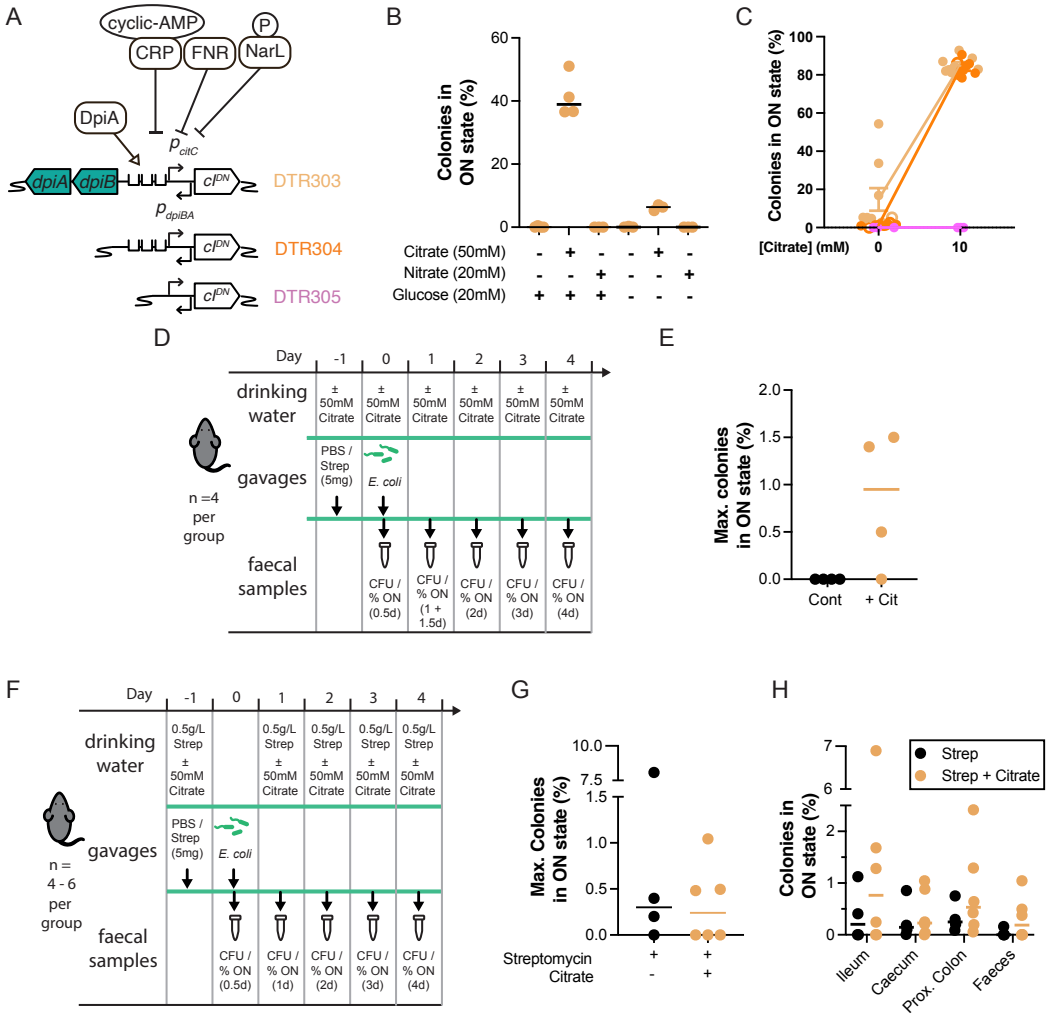

Figure 4: Characterisation of the *C. rodentium* DpiAB-p_{citC} biosensor A) C. rodentium DpiAB-p_{citC} (Cr TCS2) is regulated by various signalling systems. Truncation mutants were cloned to investigate the sensor's regulation. B) Induction dynamics of Cr TCS2 DTR303 (n=3-4) in the presence/absence of 50mM citrate, 20mM nitrate and 20mM glucose. Graph shows median. C) Cr TCS2 DTR303(n=9), ΔdpiAB DTR304 (n=9) and ΔdpiABΔdpiA binding sites DTR305 (n=3) truncation mutants in the presence or absence of citrate in anaerobic SOC media. Graph shows mean ± SEM D) Strain testing of the citrate responsive biosensor DTR303 in control and mice (n=4 per group) ± 50 mM citrate in drinking water. E) Maximum colonies in on state (%) from faecal pellets plated on selective indicator plates over 4 days for control and 50mM citrate supplemented mice. F) Additional DTR303 testing of DTR303 in streptomycin (n=4) and streptomycin + 50mM citrate supplemented mice. G) Maximum colonies in on state (%) from faecal pellets plated on selective indicator plates over 4 days and F) from dissected regions of the gut at endpoint. Graphs show median.

Citrate availability within the distal gut is likely low, with effective absorption of this nutrient in the proximal small intestine and potential utilisation by commensal and pathogenic bacteria within the gut (Liu et al., 2019). We therefore reasoned that citrate may be a spatially distinct cue within proximal gut, with low levels of response reflecting limited citrate availability at sites of *E. coli* growth within the gut. Thus, we administered DTR303 to streptomycin treated mice (n=4 or 6 per group) via oral gavage (Figure 4F), with (n=6) or without (n=4) 50 mM sodium citrate supplemented in their drinking water. Colonisation levels were elevated due to ongoing streptomycin administration, and comparable between groups (Supplementary Figure 5C-D). Faecal sampling and indicator plating demonstrated a low but detectable sensor response in both groups (Figure 4G; Supplementary Figure 5E). Upon dissection and indicator plating from different regions of the gut, all mice showed

sensor activation within at least one region of the gut (Figure 4H) with maximal switching detected in dissection samples rather than faecal pellets in all instances (5/10 max in ileum, 4/10 max in proximal colon). Combined, these results are consistent with sensor activation occurring in more proximal regions of the gut, especially the small intestine. Detection in faecal samples is likely based on memory retention, however, signal may be diluted by the comparatively high colonisation levels of *E. coli* within the distal colon that are more likely to be recovered within faecal pellets.

Discussion:

Our work expands on the use of a robust bacterial memory circuit to simultaneously evaluate hundreds of uniquely barcoded sensor strains in the murine gut or *in vitro* culture. By incorporating a DNA barcoding strategy, using a single, hyper-variable oligonucleotide for barcode delivery, we accommodate the development of large libraries of diversly sized sensors in a low-cost, single pot reaction (Figure 1). The inclusion of uniquely barcoded isogenic replicates within the library allows evaluation of sensor performance heterogeneity, improving robustness against sources of variation that may impact results, for example genetic mutations, microbial growth variability, or spatiotemporal variations that are known to impact bacterial function within the mammalian gut (Figure 2-3). The successful identification and validation of several inducible circuits, both novel and previously identified, highlights the effectiveness of our pipeline in facilitating both biotechnology tool development and biological discovery (Figure 2 - 4).

Barcoded HTMS analysis is ideally suited to the rapid cloning and evaluation of heterologous sensing circuits, such as TCSs, in a commensal *E. coli* host. This is an area of need for synthetic biology in the quest to expand the toolbox of health- and disease- relevant inducible circuit components and a particular strength of this method in comparison to other powerful parallelised bacterial memory approaches, such as RecordSeq (Schmidt et al., 2018; Schmidt et al., 2022). By comparison, RecordSeq provides a transcriptome-wide view of host cell gene expression, more akin to a recording of an RNAseq dataset over time. Testing directly in the λ-memory system, which has previously demonstrated low cellular burden, long-term stability and function in the murine gut (Kotula et al., 2014; Riglar et al., 2017; Naydich et al., 2019), allows for rapid progression from screening to follow up tests in animals (eg. Figure 3-4).

The library cloning strategy afforded a library of hundreds of unique strains and should support development of far greater library sizes where needed. However, *in vitro* experiments can most readily capitalise on this level of variability. By comparison, population bottlenecking caused barcode and strain diversity to drop quickly following administration to the gut, limiting experiment length (Supplementary Figure 3). This is similar to the observed dynamics of barcoded libraries of the pathogen *C. rodentium* (Campbell et al., 2023; Woodward et al., 2023). Nevertheless, higher numbers of barcodes per sensor should increase the robustness of results from short-term animal experiments. The modular nature of the golden gate trigger assembly method is also suited to developing larger combinatorial libraries, for example linking TCSs to a variety of promoters, removing the current dependency on neighbouring gene regulation, or creating different promoter sensitivities in a single pot construction.

Of the TCSs tested, those derived from species more closely related to the *E. coli* NGF-1 host were, unsurprisingly, more likely to show signs of functionality in at least one condition. Many promoters are controlled by combinations of TCSs, and most TCSs regulate gene expression from a variety of targets (Rajeev et al., 2020), suggesting that cross-regulation between host and heterologous TCSs and gene expression is both likely and necessary. Tetrathionate sensing, driven by TtrRS TCS control over the adjacent *ttrBCA* gene promoter from *S.* Typhimurium (Riglar et al., 2017) or *E. tarda* (Figure 2) requires the heterologous TtrRS system's activity with additional regulation provided by host *E. coli* FNR. Our experiments with truncated *Cr* TCS2 sensor variants also demonstrated the ability for host *E. coli* DpiAB to activate the heterologous *C. rodentium*-derived circuit, even in the absence of the heterologous TCS (Figure 4).

During individual testing of the *Cr* TCS2 (DTR303) sensor in mice, we observed response in animals supplemented with streptomycin and/or citrate (Figure 4), validating our screening results from streptomycin pre-treated mice (Figure 3). However, response was typically very low, variable between animals and inconsistent within individuals across time. Such variability was also detected in screening results, where not all barcode/mouse combinations showed increased OR (Figure 3E). This may reflect the expected spatial restriction of citrate availability in the gut or other sources of variation, such as those that led to some switching under anaerobic rich media growth conditions. Nevertheless, our results point to the utility of the memory element for reporting transient signals from the gut in faeces and of our library-based assay as a high sensitivity method to pick up even very low signals *in vivo*. The spatial restriction within the gut and safety profile of citrate in humans could make for interesting future applications for this promoter, however, further engineering, for example signal amplification, and more extensive understanding of variation within and between individuals would be important for this.

This pipeline not only enables the construction of biosensors tailored to the gut environment but also offers the potential for screening biosensors in other microbial growth permissible and potentially inaccessible environments. The ability to combine *in situ* and *in vitro* fingerprinting data at high throughput is a powerful method for the identification of novel inducible synthetic circuit components, biosensors and prototype disease diagnostics.

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.

References:

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

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7)**,** 581-583. doi: 10.1038/nmeth.3869.

Campbell, I.W., Hullahalli, K., Turner, J.R., and Waldor, M.K. (2023). Quantitative dose-response analysis untangles host bottlenecks to enteric infection. *Nat Commun* 14(1)**,** 456. doi: 10.1038/s41467-023-36162-3.

Cho, S.H., Dekoninck, K., and Collet, J.F. (2023). Envelope-Stress Sensing Mechanism of Rcs and Cpx Signaling Pathways in Gram-Negative Bacteria. *J Microbiol* 61(3)**,** 317-329. doi: 10.1007/s12275-023-00030-y.

Daeffler, K.N., Galley, J.D., Sheth, R.U., Ortiz-Velez, L.C., Bibb, C.O., Shroyer, N.F., et al. (2017). Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. *Mol Syst Biol* 13(4)**,** 923. doi: 10.15252/msb.20167416.

Gumerov, Vadim M., Ortega, Davi R., Adebali, O., Ulrich, Luke E., and Zhulin, Igor B. (2020). MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic Acids Research* 48(D1)**,** D459-D464. doi: 10.1093/nar/gkz988.

Guo, Y., Kitamoto, S., and Kamada, N. (2020). Microbial adaptation to the healthy and inflamed gut environments. *Gut Microbes* 12(1)**,** 1857505. doi: 10.1080/19490976.2020.1857505.

Jacob-Dubuisson, F., Mechaly, A., Betton, J.M., and Antoine, R. (2018). Structural insights into the signalling mechanisms of two-component systems. *Nat Rev Microbiol* 16(10)**,** 585-593. doi: 10.1038/s41579-018-0055-7.

Kaspar, S., and Bott, M. (2002). The sensor kinase CitA (DpiB) of Escherichia coli functions as a high-affinity citrate receptor. *Archives of Microbiology* 177(4)**,** 313-321. doi: 10.1007/s00203-001-0393-z.

Keck, F. (2020). Handling biological sequences in R with the bioseq package. *Methods in Ecology and Evolution* 11(12)**,** 1728-1732. doi: https://doi.org/10.1111/2041-210X.13490.

Keseler, I.M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martinez, C., Caspi, R., et al. (2021). The EcoCyc Database in 2021. *Front Microbiol* 12**,** 711077. doi: 10.3389/fmicb.2021.711077.

Kotula, J.W., Kerns, S.J., Shaket, L.A., Siraj, L., Collins, J.J., Way, J.C., et al. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci U S A* 111(13)**,** 4838-4843. doi: 10.1073/pnas.1321321111.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4)**,** 357-359. doi: 10.1038/nmeth.1923.

Liu, M., Hao, G., Li, Z., Zhou, Y., Garcia-Sillas, R., Li, J., et al. (2019). CitAB Two-Component System-Regulated Citrate Utilization Contributes to Vibrio cholerae Competitiveness with the Gut Microbiota. *Infect Immun* 87(3). doi: 10.1128/IAI.00746-18.

McKenzie, G.J., and Craig, N.L. (2006). 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.

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

Moreira, C.G., Russell, R., Mishra, A.A., Narayanan, S., Ritchie, J.M., Waldor, M.K., et al. (2016). Bacterial Adrenergic Sensors Regulate Virulence of Enteric Pathogens in the Gut. *mBio* 7(3)**,** 10.1128/mbio.00826-00816. doi: 10.1128/mbio.00826-16.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., et al. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 10(4)**,** 354-360. doi: 10.1038/nmeth.2404.

Naydich, A.D., Nangle, S.N., Bues, J.J., Trivedi, D., Nissar, N., Inniss, M.C., et al. (2019). Synthetic Gene Circuits Enable Systems-Level Biosensor Trigger Discovery at the Host-Microbe Interface. *mSystems* 4(4)**,** e00125-00119. doi: 10.1128/mSystems.00125-19.

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

Rajeev, L., Garber, M.E., and Mukhopadhyay, A. (2020). Tools to map target genes of bacterial two-component system response regulators. *Environmental Microbiology Reports* 12(3)**,** 267-276. doi: 10.1111/1758-2229.12838.

Riglar, D.T., Giessen, T.W., Baym, M., Kerns, S.J., Niederhuber, M.J., Bronson, R.T., et al. (2017). Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat Biotechnol* 35(7)**,** 653-658. doi: 10.1038/nbt.3879.

Riglar, D.T., and Silver, P.A. (2018). Engineering bacteria for diagnostic and therapeutic applications. *Nat Rev Microbiol* 16(4)**,** 214-225. doi: 10.1038/nrmicro.2017.172.

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

Russell, B.J., Brown, S.D., Siguenza, N., Mai, I., Saran, A.R., Lingaraju, A., et al. (2022). Intestinal transgene delivery with native E. coli chassis allows persistent physiological changes. *Cell* 185(17)**,** 3263-3277.e3215. doi: 10.1016/j.cell.2022.06.050.

Schmidt, F., Cherepkova, M.Y., and Platt, R.J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* 562(7727)**,** 380-385. doi: 10.1038/s41586-018-0569-1.

Schmidt, F., Zimmermann, J., Tanna, T., Farouni, R., Conway, T., Macpherson, A.J., et al. (2022). Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* 376(6594)**,** eabm6038. doi: 10.1126/science.abm6038.

Shaw, C., Hess, M., and Weimer, B.C. (2022). Two-component systems regulate bacterial virulence in response to the host gastrointestinal environment and metabolic cues. *Virulence* 13(1)**,** 1666-1680. doi: 10.1080/21505594.2022.2127196.

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

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

Woodward, S.E., Vogt, S.L., Pena-Diaz, J., Melnyk, R.A., Cirstea, M., Serapio-Palacios, A., et al. (2023). Gastric acid and escape to systemic circulation represent major bottlenecks to host infection by Citrobacter rodentium. *ISME J* 17(1)**,** 36-46. doi: 10.1038/s41396-022-01321-9.

Yamamoto, K., Matsumoto, F., Oshima, T., Fujita, N., Ogasawara, N., and Ishihama, A. (2008). Anaerobic Regulation of Citrate Fermentation by CitAB in Escherichia coli. *Bioscience, Biotechnology, and Biochemistry* 72(11)**,** 3011-3014. doi: 10.1271/bbb.80301.

Zou, Z.P., Du, Y., Fang, T.T., Zhou, Y., and Ye, B.C. (2023). Biomarker-responsive engineered probiotic diagnoses, records, and ameliorates inflammatory bowel disease in mice. *Cell Host Microbe* 31(2)**,** 199-212 e195. doi: 10.1016/j.chom.2022.12.004.