RESEARCH ARTICLE SUMMARY

MICROBIOLOGY

Type III secretion system effectors form robust and flexible intracellular virulence networks

David Ruano-Gallego*, Julia Sanchez-Garrido*, Zuzanna Kozik, Elena Núñez-Berrueco, Massiel Cepeda-Molero, Caroline Mullineaux-Sanders, Jasmine Naemi-Baghshomali Clark, Sabrina L. Slater, Naama Wagner, Izabela Glegola-Madejska, Theodoros I. Roumeliotis, Tal Pupko, Luis Ángel Fernández, Alfonso Rodríguez-Patón, Jyoti S. Choudhary†, Gad Frankel†

INTRODUCTION: Infections with many Gram-negative pathogens, including *Escherichia coli*, *Salmonella*, *Shigella*, and *Yersinia*, rely on the injection of effectors via type III secretion systems (T3SSs). The effectors hijack cellular processes through multiple mechanisms, including molecular mimicry and diverse enzymatic activities. Although in vitro analyses have shown that individual effectors can exhibit complementary, interdependent, or antagonistic relationships, most in vivo studies have focused on the contribution of single effectors to pathogenesis.

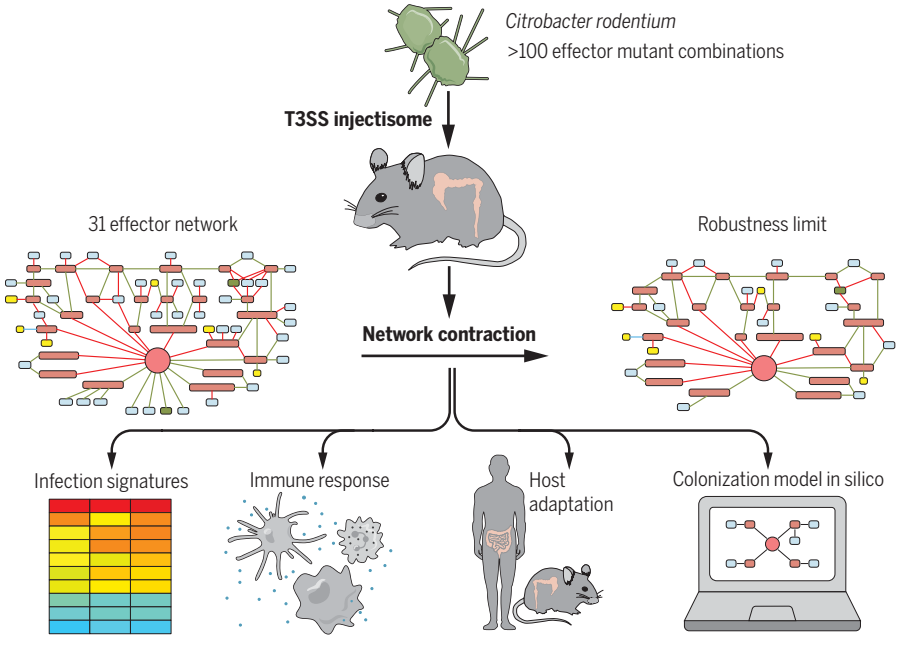
Citrobacter rodentium is a natural mouse pathogen that shares infection strategies and virulence factors with the human pathogens enteropathogenic and enterohemorrhagic *E. coli* (EPEC and EHEC). The ability of these pathogens to colonize the gastrointestinal tract is mediated by the injection of effectors via a T3SS. Although *C. rodentium* infects 31 effectors, the prototype EPEC strain E2348/69 translocates 21 effectors.

RATIONALE: The aim of this study was to test the hypotheses that, rather than operating individually, the T3SS effectors form robust intracellular networks that can sustain large contractions and that expanded effector repertoires play a role in distinct disease phenotypes and host adaption.

RESULTS: We tested the effector-network paradigm by infecting mice with >100 *C. rodentium* effector mutant combinations. First, using machine learning prediction algorithms, we discovered additional effectors, NleN and NleO. We then sequentially deleted effector genes from two distinct starting points to reach sustainable endpoints, which resulted in strains missing 19 unrelated effectors (CR14) or 10 effectors involved in the modulation of innate immune responses in intestinal epithelial cells (IECs) (CRi9). Moreover, we deleted Map and EspF, which target the mitochondria and disrupt tight junctions. Unexpectedly, all strains colonized the colon and activated conserved metabolic and antimicrobial processes in the IECs while eliciting distinct cytokine and immune cell infiltration responses. In particular, although infection with *C. rodentium* Δ*map*/Δ*espF* failed to induce secretion of interleukin-22 (IL-22), CR14 and CRi9 triggered heightened secretion of IL-6 and granulocyte-macrophage colony-stimulating factor (GM-CSF) and of IL-22, interferon-γ (IFN-γ), and IL-17 from colonic explants, respectively. Nonetheless, infection with CR14 or CRi9 induced protective immunity against secondary infections.

Although Tir, EspZ, and NleA are essential, other effectors exhibit context-dependent essentiality in vivo. Moreover, *C. rodentium* expressing the effector repertoire of EPEC E2348/69 failed to efficiently colonize mice. We used curated functional information and our in vivo data to train a machine learning model that predicted values for colonization efficiency of previously uncharacterized mutant combinations. Notably, a mutant with a low predicted value, lacking only *nleF*, *nleG8*, *nleG1*, *nleB*, and *espL*, failed to colonize.

CONCLUSION: Our analysis revealed that T3SS effectors form robust networks, which can sustain substantial contractions while maintaining virulence, and that the composition of the effector network contributes to host adaptation. Alternative effector networks within a single pathogen triggered markedly different immune responses yet induced protective immunity. CR14 did not tolerate any further contraction, which suggests that this network reached its robustness limit with only 12 effectors. As the robustness limits of other effector networks depend on the contraction starting point and the order of the deletions, machine learning models could transform our ability to predict alternative network functions. Together, this study demonstrates the robustness of T3SS effector networks and the ability of IECs to withstand drastic perturbations while maintaining antibacterial functions. ■



T3SS effectors form robust intracellular networks. T3SS effector networks can sustain substantial contractions while maintaining virulence. Using *C. rodentium* as a model showed that although triggering the conserved infection signatures in IECs, distinct networks induce divergent immune responses and affect host adaption. Because the robustness limit depends on the contraction sequence, machine learning models could transform our ability to predict the virulence potential of alternative networks.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: g.frankel@imperial.ac.uk (G.F.); jyoti.choudhary@icr.ac.uk (J.S.C.)
Cite this article as D. Ruano-Gallego *et al.*, *Science* **371**, eabc9531 (2021). DOI: 10.1126/science.abc9531

RESEARCH ARTICLE

MICROBIOLOGY

Type III secretion system effectors form robust and flexible intracellular virulence networks

David Ruano-Gallego[1]*†, Julia Sanchez-Garrido[1]*, Zuzanna Kozik[2], Elena Núñez-Berrueco[3], Massiel Cepeda-Molero[1], Caroline Mullineaux-Sanders[1], Jasmine Naemi-Baghshomali Clark[1], Sabrina L. Slater[1]‡, Naama Wagner[4], Izabela Glegola-Madejska[1], Theodoros I. Roumeliotis[2], Tal Pupko[4], Luis Ángel Fernández[5], Alfonso Rodríguez-Patón[3], Jyoti S. Choudhary[2]§, Gad Frankel[1]§

Infections with many Gram-negative pathogens, including *Escherichia coli*, *Salmonella*, *Shigella*, and *Yersinia*, rely on type III secretion system (T3SS) effectors. We hypothesized that while hijacking processes within mammalian cells, the effectors operate as a robust network that can tolerate substantial contractions. This was tested in vivo using the mouse pathogen *Citrobacter rodentium* (encoding 31 effectors). Sequential gene deletions showed that effector essentiality for infection was context dependent and that the network could tolerate 60% contraction while maintaining pathogenicity. Despite inducing very different colonic cytokine profiles (e.g., interleukin-22, interleukin-17, interferon-γ, or granulocyte-macrophage colony-stimulating factor), different networks induced protective immunity. Using data from >100 distinct mutant combinations, we built and trained a machine learning model able to predict colonization outcomes, which were confirmed experimentally. Furthermore, reproducing the human-restricted enteropathogenic *E. coli* effector repertoire in *C. rodentium* was not sufficient for efficient colonization, which implicates effector networks in host adaptation. These results unveil the extreme robustness of both T3SS effector networks and host responses.

Bacterial pathogens undergo a continuous evolutionary arms race with their hosts, driven by the acquisition of virulence factors through horizontal gene transfer. These factors enable the pathogens to bypass the resident microbiota, colonize the host, subvert immune responses, hijack cellular processes, and replicate in vivo (*1*). In response to infection, mammalian hosts, aided by the endogenous microbiota, mount immune and colonization resistance responses that are tailored to each pathogen (*2*–*5*).

Most virulence factors are either expressed on the bacterial surface or are secreted across the cell wall; accordingly, secretion systems are at the heart of bacterial pathogenesis. Many Gram-negative pathogens utilize secretion systems to inject bacterial effectors into host cells (*6*). The pulmonary pathogen *Legionella pneumonia* uses a type IV secretion system (T4SS) to inject >300 effectors into protozoa or macrophages (*7*). However, a much smaller number of type III secretion system (T3SS) effectors are injected by enteric pathogens into phagocytic or intestinal epithelial cells (IECs) (*8*). The repertoire of T3SS effectors varies extensively from one pathogen to another but can also vary within a single pathotype (e.g., clinical *Escherichia coli* isolates encode between 21 and 50 effectors). The function of an effector can be dependent on the activity of another [e.g., actin polymerization triggered by Tir of enterohemorrhagic *E. coli* (EHEC) is dependent on TccP/EspF$_U$] (*9*, *10*). Moreover, some effectors have complementary functions [e.g., the *E. coli* effectors NleC and NleE inhibit p65 and TAB2 and TAB3 in the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signaling cascade, respectively] (*1*), whereas others have antagonistic activities [e.g., the *Salmonella* Rho guanosine triphosphatases (GTPases) GEF and GAP effectors SopE and SptP] (*11*).

Studies of bacterial effectors have mostly focused on comparisons of the outcomes of infection between wild-type (WT) and single-effector gene mutants, whereas two-hybrid systems and other protein-interaction methodologies have been used to identify the effectors' binding partners and substrates (*12*–*14*). Over recent years, these studies have generated a wealth of data about the activity of individual effectors and have revealed posttranslational modifications not found in eukaryotes (*15*–*18*). However, although fundamental, these approaches do not take into account the intracellular connectivity and interdependency of the effectors.

Citrobacter rodentium is a natural mouse pathogen and the etiologic agent of transmissible colonic crypt hyperplasia (CCH) (*19*). Given the largely conserved infection strategies, infection of mice with *C. rodentium* provides a robust in vivo model for studying virulence mechanisms of the human pathogens enteropathogenic *E. coli* (EPEC) and EHEC (*20*, *21*). *C. rodentium* causes self-limiting infection in C57BL/6 mice, where regeneration of the intestinal barrier functions and clearance are dependent on interleukin-22 (IL-22) and immunoglobulin G (IgG), respectively (*2*, *22*). Infection with *C. rodentium* triggers early tissue damage repair responses, characterized by rapid proliferation of colonic transit amplifying (TA) cells. This proliferation is fueled by shifting adenosine 5′-triphosphate (ATP) production from the mitochondria to aerobic glycolysis (the Warburg effect) and the simultaneous activation of processes involved in biogenesis, import, and efflux of cholesterol (*23*, *24*).

The ability of *C. rodentium* and EPEC or EHEC to colonize the colonic mucosa is conferred by a pathogenicity island called the locus of enterocyte effacement (LEE) (*25*). The LEE encodes transcriptional regulators, the outer membrane adhesin intimin, a T3SS, translocators (EspA, EspB, and EspD), chaperones, and six effectors (Tir, Map, EspF, EspG, EspH, and EspZ) (*26*–*28*). In *C. rodentium,* 23 additional effector genes are located outside of the LEE, mainly in genomic insertions and cryptic prophages (*29*). After translocation, Tir is integrated into the IECs' plasma membrane, where it serves as a receptor for intimin. This process enables intimate bacterial attachment and the formation of actin-rich pedestal-like structures (*30*). Map and EspF target the host cell mitochondria and affect the integrity of the tight junctions (TJs) between cells (*31*, *32*). Single mutants lacking intimin or Tir (*33*); NleB, an arginine glycosyltransferase that modifies the death domain–containing proteins such as FADD (Fas-associated death domain protein) or RIPK1 (receptor-interacting protein kinase 1) (*16*); NleA (also known as EspI), which binds Sec24 and affects endosomal trafficking (*34*); or EspZ, which serves as the T3SS gate keeper protecting cells from effector overdose cytotoxicity (*26*, *35*), are highly attenuated in vivo (*16*, *36*, *37*). Single deletion of other effector genes has little effect on colonization; the known functions assigned to each of the *C. rodentium* effectors are shown in table S1. We therefore hypothesized that, rather than operating individually, the effectors form robust networks that hijack host signaling pathways in a coordinated manner.

[1]Centre for Molecular Microbiology and Infection, Department of Life Sciences, Imperial College, London, UK. [2]Functional Proteomics Group, Chester Beatty Laboratories, Institute of Cancer Research, London, UK. [3]Laboratorio de Inteligencia Artificial, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Madrid, Spain. [4]The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [5]Centro Nacional de Biotecnología (CNB-CSIC), Department of Microbial Biotechnology, Madrid, Spain.
*These authors contributed equally to this work. †Present address: Department of Molecular Evolution, Centro de Astrobiología, Instituto Nacional de Técnica Aeroespacial-Consejo Superior de Investigaciones Científicas (INTA-CSIC), Madrid, Spain. ‡Present address: Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA.
§Corresponding author. Email: g.frankel@imperial.ac.uk (G.F.); jyoti.choudhary@icr.ac.uk (J.S.C.)

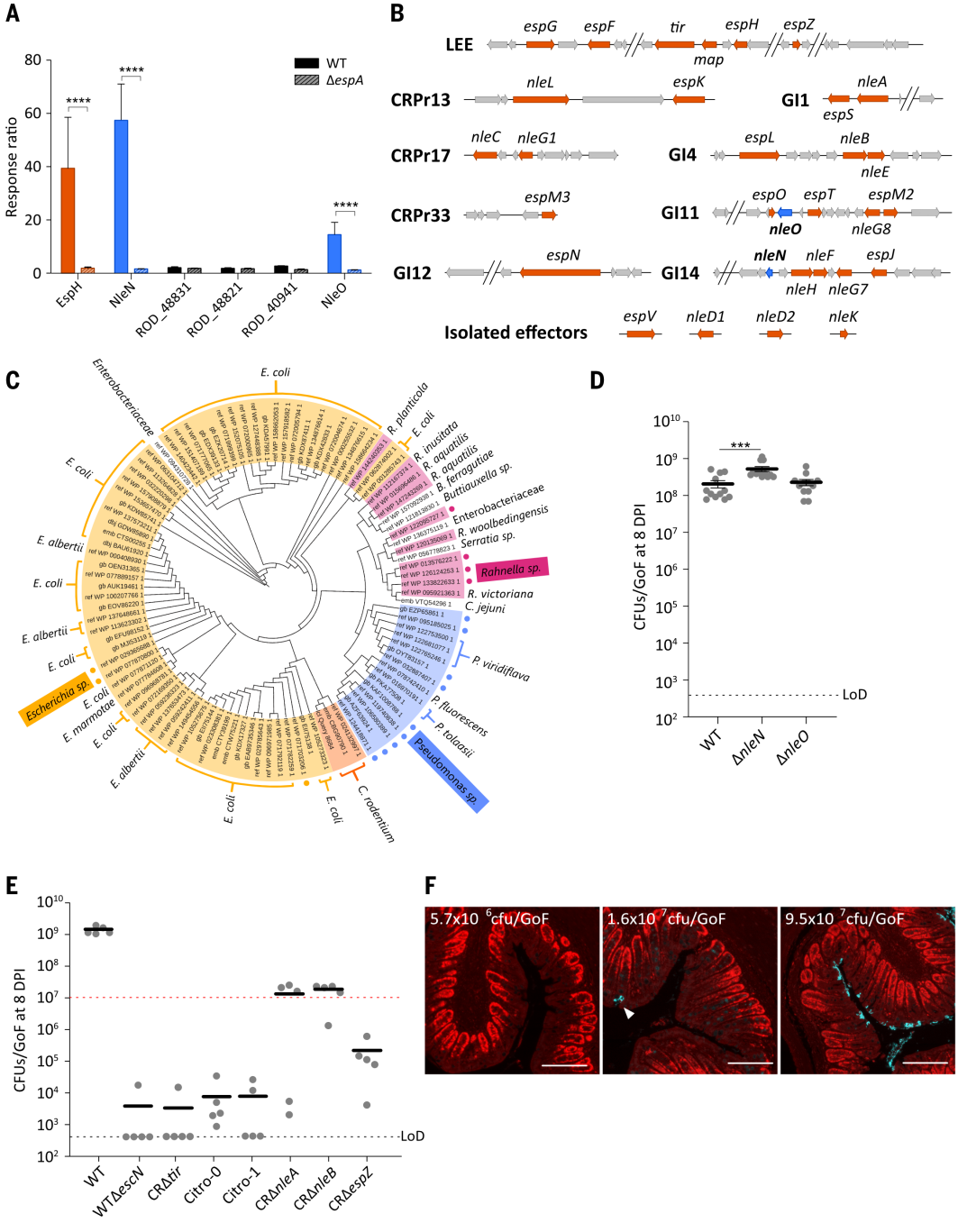


Fig. 1. NleN and NleO are newly identified *C. rodentium* T3SS effectors. (**A**) ROD_48841 (NleN) and ROD_40891 (NleO)–TEM-1 fusion proteins are translocated by WT but not *C. rodentium* Δ*espA*. Response ratio is 450/520-nm emission ratio, adjusted for background fluorescence from uninfected cells. Results show means + SEM from four biological repeats, analyzed by multiple *t* tests corrected with the two-stage step-up method of the Benjamini, Krieger, and Yekutieli approach for FDR. ****$P < 0.0001$. (**B**) Schematic showing the genomic location of the known and newly identified *C. rodentium* effectors (orange and blue, respectively). (**C**) Cladogram created with the first 100 BLASTp hits using ROD_40891 (NleO) as the query sequence. Leaf names are highlighted according to genus: *Escherichia* (yellow), *Citrobacter* (orange), *Pseudomonas* (blue), and *Rahnella* (magenta). Genera with fewer than three leaves are not highlighted. Species names are labeled where possible. Dots indicate unclassified species of the color-matched genus. (**D**) CFU/GoF at 8 DPI with WT and the indicated *C. rodentium* mutants. Graph shows means ± SEM from three biological replicates, where statistical significance was determined by one-way ANOVA and Dunnett's multiple comparison test (values compared with WT infection). ***$P < 0.001$. (**E**) Mean numbers of CFU/GoF at 8 DPI with the indicated *C. rodentium* strains ($n = 5$ per group). (**F**) Representative immunostaining images of *C. rodentium* (cyan) and PCNA (red) on colonic sections with the indicated CFU/GoF. Scale bars, 100 μm. LEE, locus for enterocyte effacement; GI, genomic island; CRPr, prophage elements; LoD, limit of detection.

We used the *C. rodentium* model to test the effector-network hypothesis. To this end, we derived two mutants, each generated through a distinct sequence of progressive effector gene deletions. These as well as the intermediate mutants (>100 different combinations) were tested in vivo. The phenotypic data were used to train a machine learning (ML) model that inferred the contribution of different effectors to infection and predicted colonization outcomes. We also generated a *C. rodentium* strain replicating the effector composition of the human-restricted pathogen EPEC strain E2348/69. We found that the function of the effectors was highly cooperative, forming a robust network that could withstand substantial contractions while maintaining pathogenicity. Moreover, the composition of the network contributed to host adaptation. Equally, the host was able to fight infections with a single bacterial pathotype that injected vastly different effector networks.

The comprehensive *C. rodentium* T3SS effector repertoire

C. rodentium expresses six LEE-encoded and 23 non–LEE-encoded effectors. Because our aim was to probe the robustness of the entire effector network, we first used ML classification algorithms to search for potentially unidentified effectors (the search criteria and databases used are listed in the Materials and methods section). Briefly, in the learning phase, we searched for 66 features that distinguish known effectors

from known noneffectors. Every open reading frame (ORF) was assigned a score that reflects how likely it is to encode an effector. The proteins used to train the algorithm, the set of feature values given to each ORF, and the results of the ML classifier are presented in tables S2, S3, and S4, respectively. We tested five candidates from predictions showing confidence scores ≥0.9 and with a minimum length of 80 amino acids using the β-lactamase (TEM-1) translocation assay (fig. S1A). Infection of 3T3 fibroblast cells with WT *C. rodentium* or *C. rodentium* Δ*espA* (T3SS deficient), expressing the putative effector–TEM-1 fusions or EspH–TEM-1 as a positive effector control, revealed that candidates ROD_48841 (renamed NleN) and ROD_40891 (renamed NleO) were translocated from WT *C. rodentium* but not from Δ*espA* (Fig. 1A).

NleN is a 9.5-kDa effector encoded in the same genomic island as NleH, NleF, NleG7, and EspJ (Fig. 1B). NleN-like sequences, ranging in size between 85 and 613 amino acids, are present in several hypothetical proteins from human and cattle pathogenic *E. coli* isolates (including O26:H11, O177:25, and O45:H2), *Escherichia albertii*, *Salmonella enterica* subspecies *salamae*, *Shigella boydii*, and *Shigella dysenteriae* (fig. S1B). NleO is a 21.9-kDa effector encoded in the same island as EspO, EspT, EspM2, and NleG8 (Fig. 1B). NleO homologs (>95% identity) are found in multiple *E. coli* isolates, *E. albertii*, and *Pseudomonas* spp. (Fig. 1C). Oral infection of C57BL/6 mice with *C. rodentium* Δ*nleN* or Δ*nleO* revealed that both mutants efficiently colonized at 8 days postinfection (DPI) (which was heightened in the Δ*nleN*) (Fig. 1D) and induced CCH (which was lower in the Δ*nleO* compared with WT) (fig. S1C). Thus, *C. rodentium* encodes at least 31 effectors, which include the newly identified NleN and NleO.

Shrinking the network reveals context-dependent effector essentiality and function

Once we established the *C. rodentium* effector repertoire, we constructed a set of plasmids that allowed us to generate scarless deletions of each of the effector genes, either individually or as a group. This was used to generate a comprehensive collection of 31 single effector–deletion mutants, which were used for testing and validation throughout the study (table S5).

We first assessed whether Tir, which is necessary for intimate bacterial attachment and pedestal formation in cultured cells (*30*), was sufficient for *C. rodentium* colonization in vivo. Given the recently observed altered susceptibility to infection of mice from different breeding facilities (*38*, *39*), we first confirmed that *C. rodentium* Δ*tir* (CRΔ*tir*) was attenuated during infection of our commercially sourced mice, and we verified that it colonized similarly to CRΔ*escN* (T3SS deficient) (Fig. 1E). In a reciprocal experiment, we generated *C. rodentium* strains missing all the effectors genes except *tir* (Citro-1) or missing all 31 effector genes including *tir* (Citro-0) as a control, both of which encode a functional T3SS and intimin. Infections of the mouse colorectal cell line CMT-93 confirmed that Citro-1, but not Citro-0, induced formation of actin pedestals (fig. S2A). However, neither Citro-0 nor Citro-1 were able to infect mice (Fig. 1E), which suggests that intimin-Tir interactions, which have been shown in many studies to play a key role in infection and immunity (*26*, *40*–*42*), are necessary but not sufficient for colonization, and other effectors are needed. An EPEC-1 strain, which similarly expresses a functional T3SS, intimin, and Tir as the sole effector, is unable to infect human intestinal biopsies ex vivo (*43*).

Aiming to contract the effector network, we set 10^7 colony-forming units per gram of feces (CFU/GoF) as a shedding threshold for infected mice; this was based on the observation that *C. rodentium* could not be detected by immunofluorescence in colonic sections of mice shedding $<10^7$ CFU/GoF (Fig. 1F). We first confirmed that, consistent with previously published data, Δ*nleA*, Δ*nleB*, and Δ*espZ* mutants are attenuated—Δ*nleA* and Δ*nleB* were shed around the 10^7 CFU/GoF cutoff, whereas Δ*espZ* was shed at 10^5 CFU/GoF (Fig. 1E). Accordingly, we concluded that Tir, NleA, NleB, and EspZ are essential in the context of the WT effector network.

We next stress-tested the network by sequentially deleting effector genes from WT *C. rodentium* followed by mouse infection and quantification of fecal shedding (Fig. 2, A and B). To maintain an unbiased approach, we initially mutated effector genes in a random order, generating strains named CR1, CR2, etc., according to the number of deletion rounds performed (Fig. 2B). Groups of four to five mice were infected with each mutant, and fecal *C. rodentium* shedding was enumerated on 3, 6, 7, and 8 DPI (8 DPI is shown in Fig. 2A; 3, 6, and 7 DPI are shown in fig. S2B). We observed no major defects in shedding after sequential deletion of the effector genes *espK*, *nleL*, *espM3*, *espV*, *espS*, *espL*, *nleG1*, and *espG* (CR7) (Fig. 2, A and B). However, deletion of *espF* from CR7 (CR7Δ*espF*) decreased bacterial shedding below the 10^7 CFU/GoF threshold (Fig. 2A).

To determine where in the deletion process leading to CR7Δ*espF* EspF became essential, we constructed WTΔ*espF*, CR1Δ*espF*, CR2Δ*espF*, CR3Δ*espF*, CR4Δ*espF*, CR5Δ*espF*, and CR6Δ*espF* mutants, as well as a Δ*espF*/Δ*espG* double mutant (the last two deletions in CR7Δ*espF*). Infection with all these strains showed similar levels of shedding as that observed in WT (fig. S2C), indicating that effector essentiality is context dependent and that the decreased shedding of CR7Δ*espF* is caused by a combinatorial effect of deleting at least three effectors. As such, we identified EspF as a conditionally essential effector; grouped EspF with the essential effectors Tir, NleA, NleB, and EspZ (Fig. 2B); and continued the sequential deletions from CR7.

By deleting *map*, *espH*, *espN*, *nleD1*, and *nleD2* from CR7, we generated CR12 (missing 13 effector genes) (Fig. 2B), which was shed at $>10^8$ CFU/GoF (Fig. 2A). Although deleting either *nleC* or *nleE* from CR12 resulted in shedding at the threshold (fig. S2D), deletion of both *nleC* and *nleE* from CR12 resulted in shedding below the threshold (Fig. 2A). Notably, deletion of both *nleC* and *nleE* from WT *C. rodentium* (named CRi2; see below) did not affect shedding (Fig. 2, C and D), which supports the concept of context-dependent effector essentiality. Because NleC and NleE have been shown to target the NF-κB signaling cascade in vitro (table S1), we investigated whether this was their physiological function in vivo. For this, we treated CR12Δ*nleC*/*E*-infected mice with the NF-κB inhibitor BMS-345541 (*43*), using CR12 as a control. BMS-345541 treatment did not affect the outcome of infection with CR12 and rescued colonization of CR12Δ*nleC*/*E* (Fig. 2E), providing mechanistic insight into the role of these effectors in vivo, which could not be elucidated in the past owing to the absence of a phenotype during infection with WTΔ*nleC*/*E*.

Because deletion of either *nleC* or *nleE* affected shedding, they were retained in CR12 and we continued with the deletion of *nleK*, *espJ*, *nleG7*, *nleF*, *nleH*, and *nleN*, generating CR14 (missing 19 effector genes), which encodes only 12 of the 31 *C. rodentium* effectors (Fig. 2B). Although CR14 was shed at ~10^8 CFU/GoF, further individual deletions of *nleO*, *nleG8*, *espM2*, *espO*, or *espT* from CR14 showed erratic levels of shedding, below the 10^7 CFU/GoF cutoff (Fig. 2A). Therefore, by deleting the selected 19 genes in CR14, we reached the robustness limit of this effector network, which was unable to sustain further contractions.

A network lacking the major anti-inflammatory effectors supports colonization

The unbiased approach to contract the network that led to CR14 showed context-dependent effector essentiality. Thus, the starting point of effector gene deletion, and the subsequent order of network contraction, is pivotal in reaching an endpoint. Accordingly, we decided to conduct a second round of progressive deletions from an alternative starting point. Because deleting both *nleC* and *nleE* from CR12 (Fig. 2B)—but not WT (CRi2; Fig. 2C)—markedly affected shedding, we started the second round of deletions from CRi2.

While embarking on the second round of network contraction, we decided to take a targeted approach and sequentially deleted

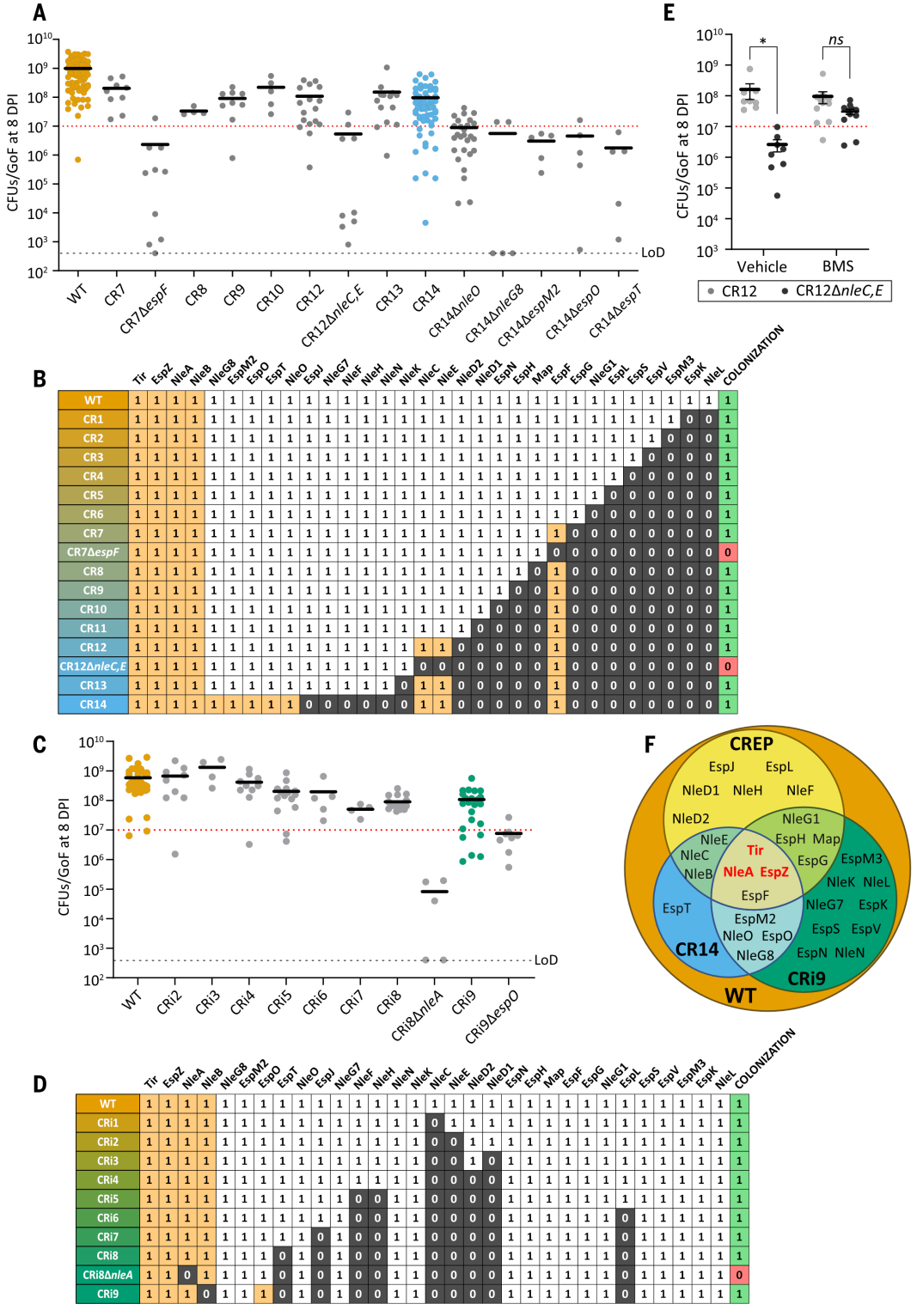


Fig. 2. *C. rodentium* withstands deletion of 19 unrelated and 10 immuno-modulatory effectors. (**A**) CFU/GoF at 8 DPI with WT and the indicated mutants (n = 5 to 77 per group). (**B**) A table illustrating the rounds of sequential deletion toward generation of CR14. The 1 and 0 labels (black backgrounds) indicate the presence or absence of an effector, respectively, and orange backgrounds indicate essential effectors. The colonization column indicates more than (1) or less than (0) 10^7 CFU/GoF at 8 DPI. (**C**) CFU/GoF at 8 DPI with WT and the indicated *C. rodentium* mutants (n = 4 to 41). (**D**) A table illustrating the rounds of sequential deletion toward generation of CRi9 [similar to that in (B)]. (**E**) Mice were infected with the indicated strains, and 10 mg/kg of BMS-345541 (BMS) or vehicle were administered by oral gavage at 3 DPI. CFU/GoF were enumerated at 8 DPI. Results show means ± SEM from two biological replicates, and significance was determined by two-way ANOVA and corrected with the two-stage step-up method of the Benjamini, Krieger, and Yekutieli approach for FDR. *P < 0.01; ns, not significant. (**F**) Venn diagram representing the effectors present in WT, CR14, CRi9, and CREP (*C. rodentium* encoding the EPEC 2348/69 effector repertoire). The master effectors are indicated in red. In (A), (C), and (E), the colonization threshold of 10^7 CFU/GoF is indicated by a dotted red line. Each data point represents a single mouse.

effector genes implicated in subversion of host immunity (strains named CRi) from CRi2 (Fig. 2D). Deletion of *nleC*, *nleE*, *nleD1*, *nleD2*, *nleF*, *nleH*, *espL*, *espJ*, and *espT* resulted in CRi8, which was shed at >10^8 CFU/GoF (Fig. 2C). Because NleA and NleB are also reported to affect host immunity (table S1), the effect of deleting *nleB* or *nleA* from CRi8 was tested. Whereas deleting *nleA* from CRi8 resulted in shedding at ~10^5 CFU/GoF, deletion of *nleB* (CRi9) did not affect shedding (Fig. 2C). Therefore, although essential for efficient colonization in the context of the WT (Fig. 1D), NleB is dispensable in the context of the newly established effector network in CRi8. Notably, deletion of *espO*—which affects IL-22 signaling in IECs (*44*)—from CRi9 decreased shedding to <10^7 CFU/GoF (Fig. 2C). EspO thus exhibits context-dependent effector essentiality in CRi9, which we considered the endpoint for deletion of effectors affecting host inflammation.

We confirmed that CR14 and CRi9 formed typical actin-rich pedestals after infection of

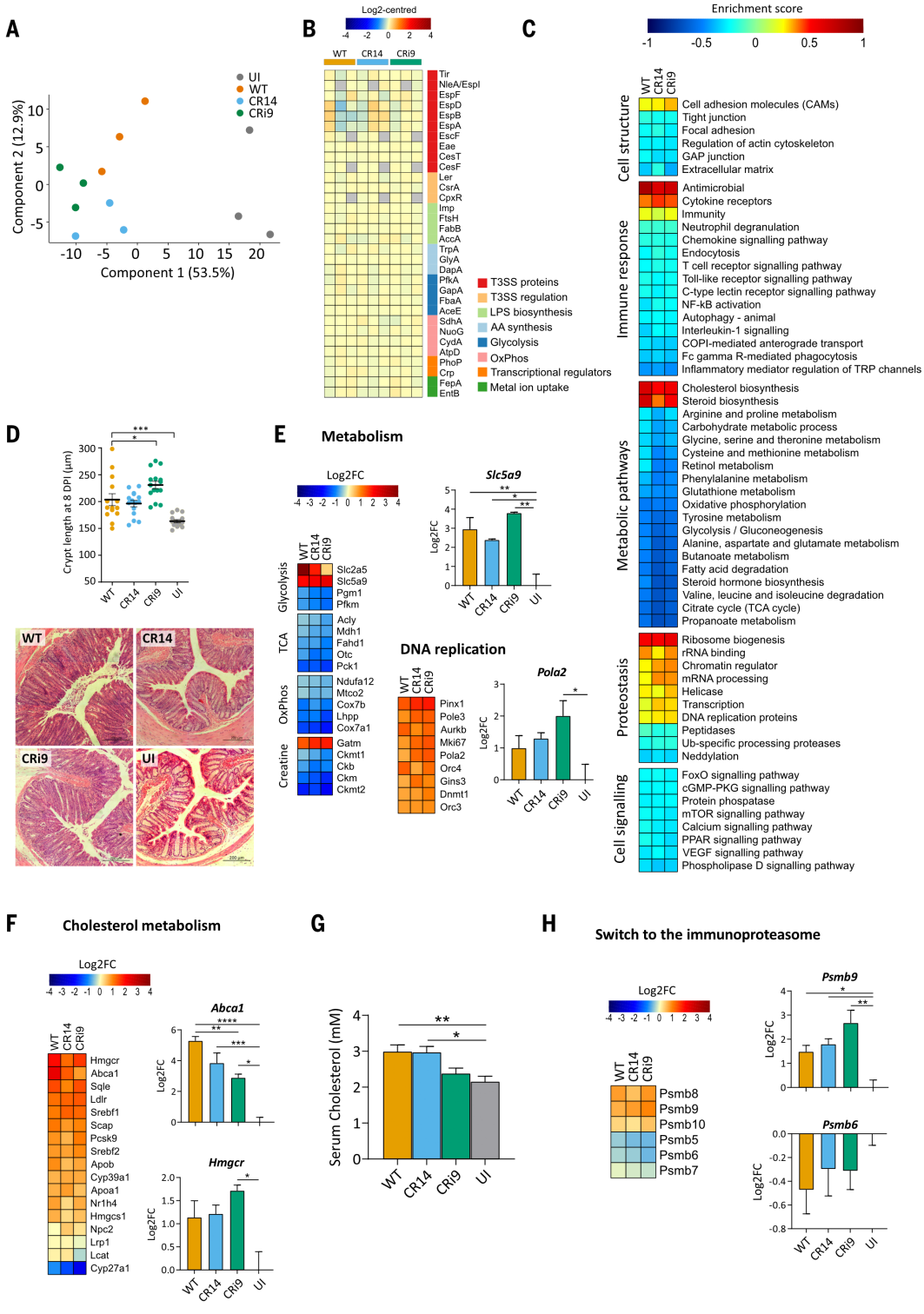


Fig. 3. CR14 and CRi9 activate a conserved molecular signature of infection. (**A**) PCA based on proteins with quantitative values from IECs of mice infected with WT, CR14, or CRi9 or those that were mock infected (UI). (**B**) Heatmap representation of the abundance of the indicated *C. rodentium* proteins involved in major metabolic pathways and virulence, detected in IEC lysates. (**C**) Heatmap of the major IECs' pathways similarly altered after infection with either WT, CR14, or CRi9. (**D**) Representative H&E-stained sections (bottom) and crypt-length measurements (top) from the same groups of mice at 8 DPI (scale bars, 200 µm). Each dot represents the mean per mouse, and lines show means ± SEM. (**E** and **F**) Abundance heatmaps of selected proteins involved in metabolic processes and DNA replication (E) and cholesterol metabolism (F). qPCR of *Slc5a9* and *Pola2* (E) and *Abca1* and *Hmgcr* (F) revealed changes in mRNA levels in IECs extracted from the same mice. (**G**) Serum cholesterol values in animals infected with the indicated strains. Graphs show means + SEM. (**H**) Heatmap of selected proteasomal subunits of the constitutive- and immune-proteasomes. Validations were by qPCR of *Psmb6* (constitutive proteasome) and *Psmb9* (immunoproteasome). All RT-qPCR graphs show means + SEM of $\log_2$FCs in expression levels compared with mean expression levels in UI mice. In (D) to (H), statistical significance was determined by one-way ANOVA and Tukey's multiple comparison test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

CMT-93 cells (fig. S2A) and sequenced their genomes, alongside the parental strain ICC169, both before infection and after isolation from feces at 8 DPI. Comparison of the genomes confirmed that the only differences compared with WT were in the intended deletions in CR14 and CRi9, which share eight effectors (Fig. 2F), and verified genomic stability because no further mutations occurred during infection (table S6).

Contraction of the effector network does not affect gene expression or bacterial physiology in vivo

Because *C. rodentium* binds IECs at the apex of the colonic crypt, where it injects effectors, we aimed to compare the IEC responses to infection with WT, CR14, and CRi9 at the molecular level. Groups of mice making up three biological repeats were infected, and the colons from those exceeding the 10^7 CFU/GoF threshold were extracted at 8 DPI (fig. S3A), followed by histology, immunostaining, enzyme-linked immunosorbent assay (ELISA), cytokine

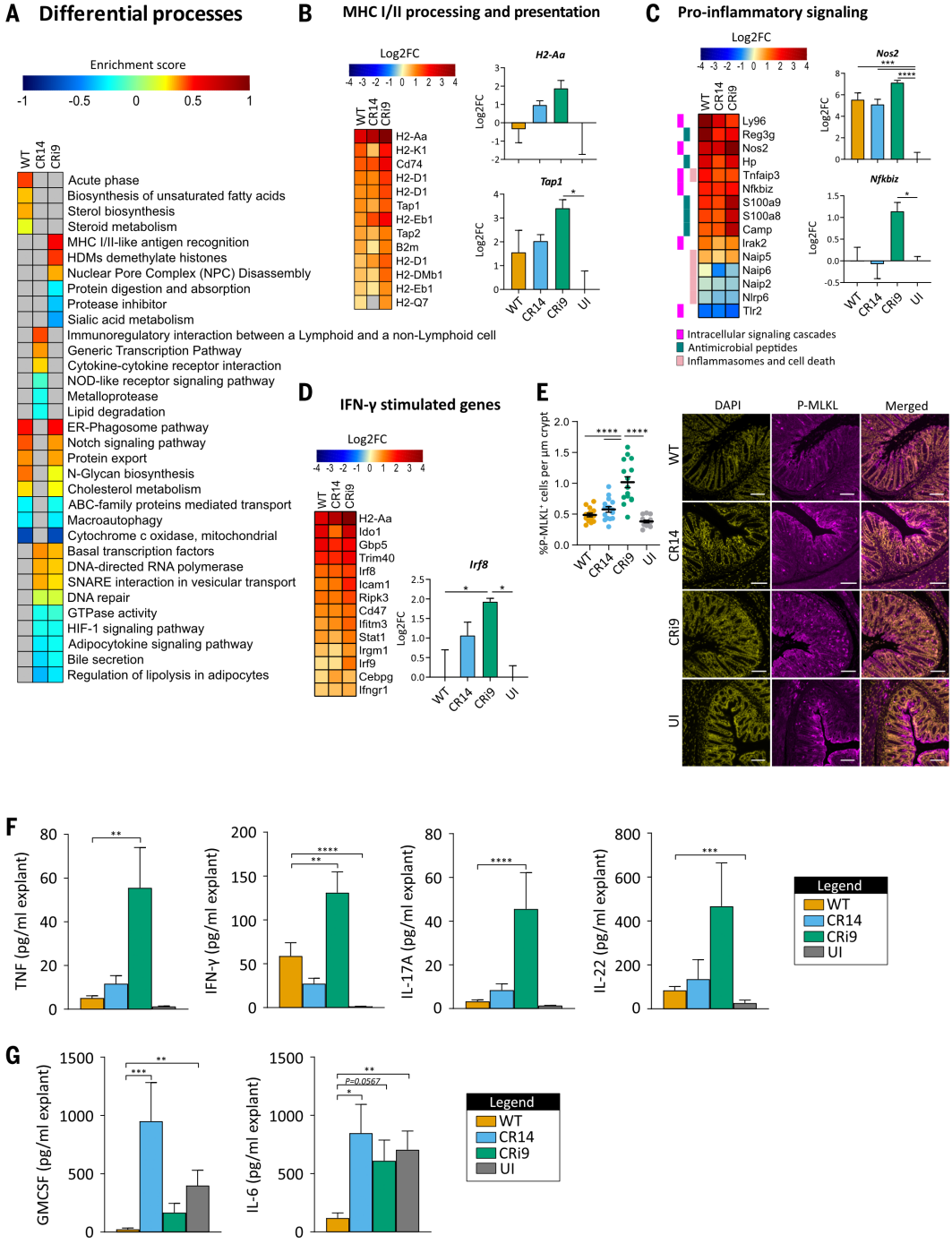


Fig. 4. CRi9 and CR14 trigger distinct inflammatory responses. (**A**) Heatmap of the major pathways differentially regulated in IECs in mice infected with CR14 or CRi9 compared with WT. (**B** to **D**) Heatmaps of selected proteins involved in MHC class I or II antigen processing and presentation (B), proinflammatory signaling (C), and in the IFN-γ–stimulated response (D). RT-qPCR of *H2-Aa* and *Tap1* (B), *Nos2* and *Nfkbiz* (C), and *Irf8* (D) reveal changes in mRNA levels. Shown are means + SEM of three biological repeats of $\log_2$FCs in expression levels compared with mean expression levels in UI mice. Statistical analysis by one-way ANOVA with Tukey's multiple comparison test. (**E**) Quantification (left) and representative images (right) of phospho-MLKL–positive cells in the distal colon at 8 DPI after infection with the indicated strains. Percentage of MLKL-positive cells = [(number of phospho-MLKL–positive cells counted from the colonic crypts) / (length of the crypt from which they were counted)] × 100. Each point represents the mean value per mouse, and lines show means ± SEM from three biological replicates. (**F** and **G**) Cytokine levels measured from colonic explants of the same mice at 8 DPI. TNF, IFN-γ, IL-17A, and IL-22 show increased values in CRi9 (F), and GM-CSF and IL-6 are increased in CR14-infected colons (G). Graphs show means + SEM of three biological repeats, with five mice per group. [(E) to (G)] Statistical analysis was by one-way ANOVA with Tukey's (E) or Dunnett's multiple comparison test (values compared with WT infection) [(F) and (G)]. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

profiling, immune cell infiltration, and IEC proteomics (fig. S3B).

Protein extracts from the IECs within the same infection group were pooled at a 1:1 ratio and analyzed by isobaric labeling proteomics in two batches (fig. S3C). A total of 7760 host proteins (table S7) [false discovery rate (FDR) < 5%] and 671 *C. rodentium* proteins were quantified (table S8). For the host proteins, the median coefficient of variation (CV) of technical and biological replicates between the sample batches was 4 and 7.4%, respectively, suggesting high reproducibility. Moreover, principal components analysis (PCA) revealed that biological replicates clustered without batch effects, demonstrating a clear separation between infected and uninfected samples (Fig. 3A).

Because intimately attached *C. rodentium* copurifies with the IECs, we used the proteome to determine whether the deletions introduced into CR14 and CRi9 affected the abundance of other bacterial virulence factors, regulators, or lipopolysaccharide (LPS) biosynthesis and metabolic proteins. This analysis revealed no changes in the abundance of proteins involved in these processes, which suggests that the deletions of effector genes have no major effect on bacterial protein abundance or physiology in vivo (Fig. 3B).

We next used quantitative polymerase chain reaction (qPCR) to analyze whether the multiple deletions in CR14 and CRi9 affected the expression of selected translocator and effector genes in vivo, using bacterial RNA associated with lysates of infected IECs as a template. This revealed similar expression of *espB*, *tir*, *nleA*, *espZ*, and *espF* in WT, CR14, and CRi9—consistent with the deletion profiles, expression of *nleN* and *map* was specifically detected

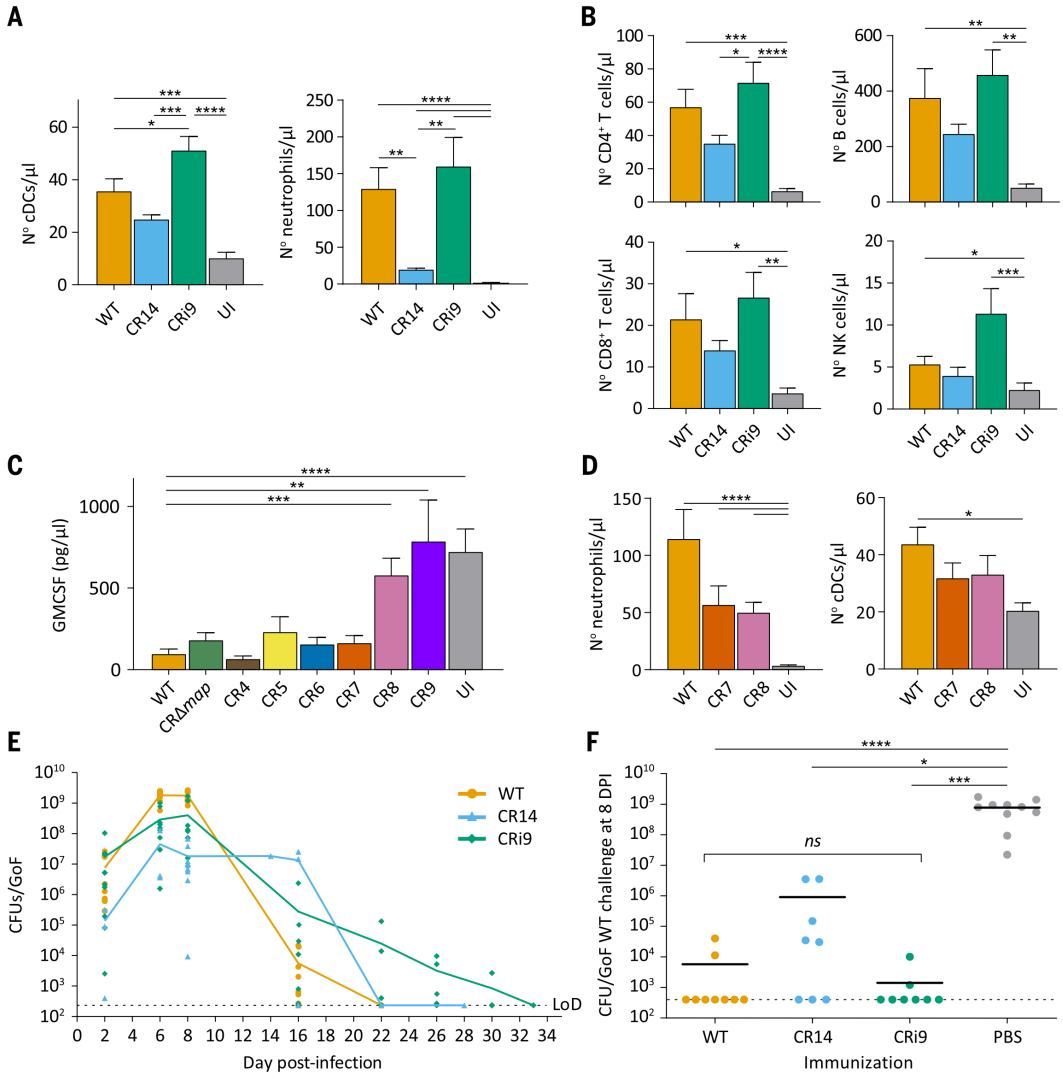


Fig. 5. Effector network compositions govern specific infection outcomes. (**A** and **B**) Immune cell recruitment to the colonic tissue of either mice infected with WT, CR14, or CRi9 or UI mice at 8 DPI. Total numbers of cDCs and neutrophils (A) and CD4+ or CD8+ T cells, NK cells, and B cells (B) per microliter of colon homogenate. Shown are means + SEM (n = 10). (**C**) GM-CSF levels measured from colonic explants of mice infected with the indicated CR14 intermediates or UI mice at 8 DPI. Graph shows means + SEM (n = 5 to 10). Statistical analysis by one-way ANOVA with Dunnett's multiple comparison post hoc test (values compared with WT infection). (**D**) Total numbers of neutrophils and cDCs recruited to the colonic tissue of mice infected with the indicated mutants at 8 DPI (number of cells per microliter of colon homogenate). Shown are means + SEM of n = 10 mice. In (A), (B), and (D), statistical significance was determined by one-way ANOVA with Tukey's multiple comparison test. (**E**) Temporal fecal bacterial shedding in mice infected with WT, CR14, and CRi9. Lines represent the mean bacterial load. (**F**) Fecal bacterial shedding 8 days after reinfection of immunized mice with WT *C. rodentium*. Shown are means of n = 8 to 10 mice. Significance was determined by nonparametric statistical analysis by Kruskall-Wallis test. Mice that did not reach 10^6 CFU/GoF during the immunization step were excluded from the experiment. In (E) and (F), each data point represents CFU/GoF in individual mice. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

in WT and CRi9, whereas expression of *nleB* and *nleC* was specifically detected in WT and CR14 (fig. S3D). Thus, the sequential deletions did not affect the expression of the remaining effector genes in CR14 and CRi9.

WT, CR14, and CRi9 activate similar metabolic and protein degradation processes

Despite expressing vastly different effector networks, WT, CR14, and CRi9 are shed at >10^7 CFU/GoF. We were therefore interested in determining how the different effector networks affected host responses. Using the IEC proteomics data, we calculated $\log_2$ fold-change ($\log_2$FC) over the uninfected control and performed one-sample t tests using all infected samples as replicates. We found that 14% (1095 proteins) of the proteome was differentially regulated (P < 0.05; $\log_2$FC > 0.5, either up- or down-regulated), with 744 proteins increasing and 351 decreasing across all conditions. Enrichment analysis classified the changes in protein abundances into the following biological processes: metabolic pathways, proteostasis, cell signaling, immune response, and cell structure (Fig. 3C).

In line with previous studies (*23*, *24*), we observed an overall down-regulation of pathways associated with energy metabolism, mainly within the mitochondria, whereas pathways associated with cell proliferation—DNA replication, transcription, mRNA processing, and ribosome biogenesis—increased (Fig. 3C). Consistently, histological analysis and immunostaining revealed CCH (Fig. 3D) and expansion of the PCNA$^+$ proliferative zone (fig. S4A), which were greater in CRi9. Because expansion of TA cells is associated with diminishing abundance of differentiated cells, we quantified differentiation markers by qPCR, which revealed reduced abundance of *Slc26a3* and *Reg4* mRNA—markers of absorptive epithelial cells and deep crypt secretory cells, respectively (fig. S4B). Additionally, the expansion of the proliferative zone coincided with up-regulation of the glucose transporter SLC5A9, fructose transporter SLC2A5, and glycine amidinotransferase (GATM) (Fig. 3E), which mediate monosaccharide influx and creatine biogenesis, respectively, and a shift to aerobic glycolysis and energy dissemination via phospho-creatine.

Slc2a5 is regulated by the transcription factor liver X receptors (LXRs) (*45*), which are generally associated with the up-regulation of genes involved in cholesterol efflux. Consistent with this, the abundance of the LXR-regulated cholesterol transporter ABCA1 also increased upon infection with WT, CR14, and CRi9, although to a lesser extent during CRi9 infection. (Fig. 3F). In line with the lower abundance of ABCA1, which is involved in reverse cholesterol transport (*46*), the level of serum cholesterol was lower in CRi9 compared with WT and

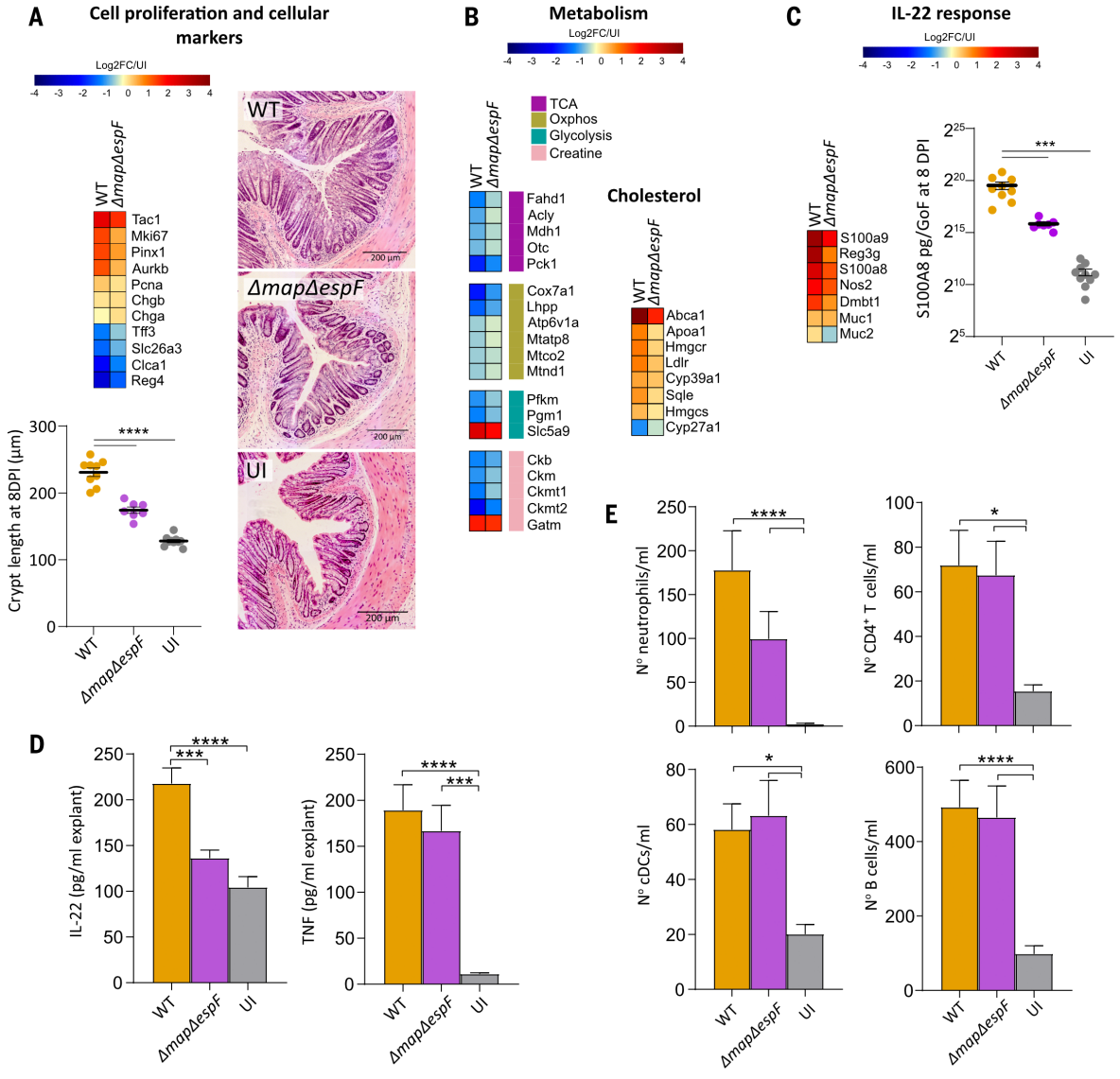


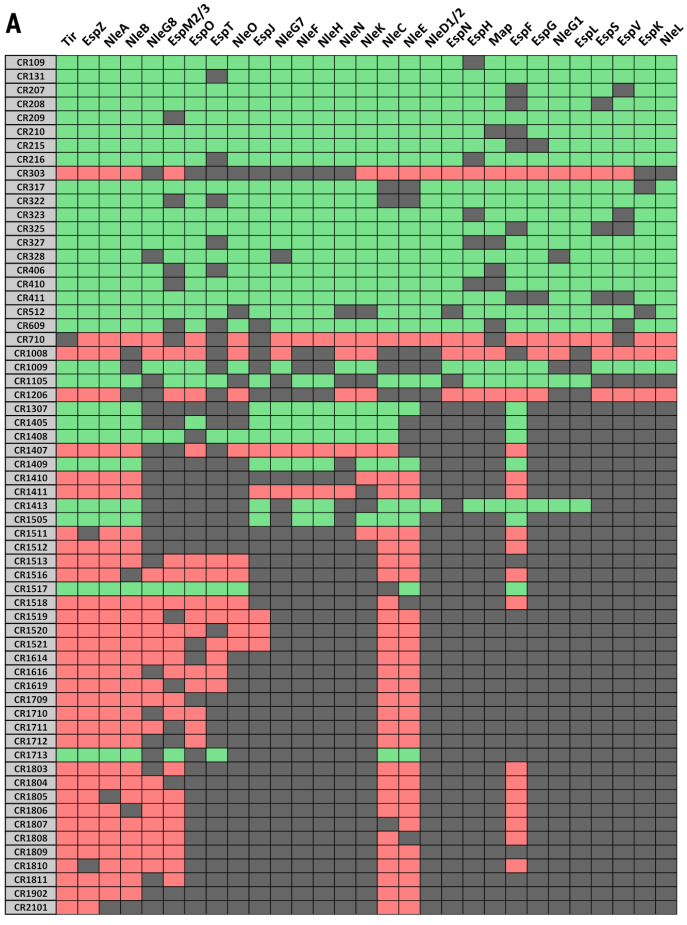
Fig. 6. Map and EspF are required for induction of IL-22 responses. (**A**) Heatmap representing the abundance of specific cell markers and proteins involved in cell proliferation (top left) in IEC samples from mice infected with WT *C. rodentium* or CR*Δmap/ΔespF* [normalized to uninfected (UI) samples]. Representative H&E-stained sections (right) and crypt-length measurements (bottom left) from the same groups of mice at 8 DPI (scale bars, 200 µm). Each dot represents the mean per mouse, and lines show means ± SEM from two biological replicates. (**B**) Heatmaps illustrating the changes in abundance of proteins involved in cellular [tricarboxylic acid cycle (TCA), oxidative phosphorylation (Oxphos), glycolysis, and creatine pathway] and cholesterol metabolism in IEC samples after infection with the indicated strains. (**C**) Heatmap representation of the changes in the selected proteins involved in the IL-22 response at 8 DPI with the indicated strains (left) and S100A8 ELISA on homogenized stool samples collected at 8 DPI (right). Each dot represents a single mouse. (**D**) IL-22 and TNF levels measured by ELISA in colonic explants of mice infected with WT *C. rodentium* or CR*Δmap/ΔespF* at 8 DPI; IL-22 shows decreased values in CR*Δmap/ΔespF*-infected samples. (**E**) Immune cell recruitment to the colonic tissue at 8 DPI. Graphs show total numbers of neutrophils, conventional dendritic cells (cDCs), $CD4^+$ T cells, and B cells per microliter of colon homogenate. In (C) to (E), graphs show means ± SEM of two biological repeats. Statistical analysis was determined by one-way ANOVA with Tukey's multiple comparison test. *$P < 0.05$; ***$P < 0.001$; ****$P < 0.0001$.

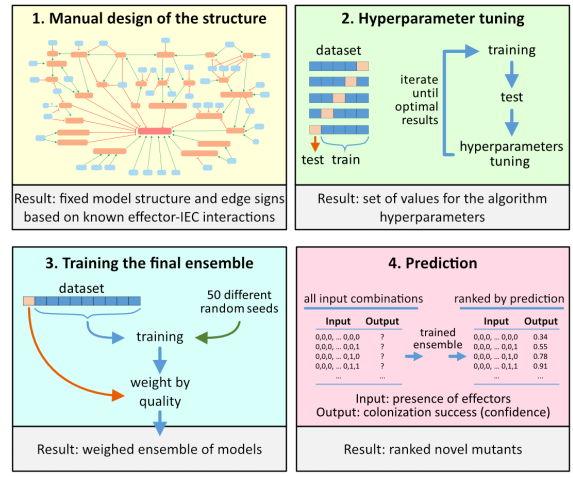
CR14 (Fig. 3G). Similarly, the SREBP2-regulated genes involved in cholesterol biosynthesis (HMGCR) and uptake (LDLR and PCSK9) (*47*) increased upon infection with WT, CR14, and CRi9 (Fig. 3F and fig. S4C). These changes were validated by qPCR (Fig. 3F and fig. S4C). Thus, modulation of LXR-mediated responses is differentially affected by the effector networks in WT, CR14, and CRi9.

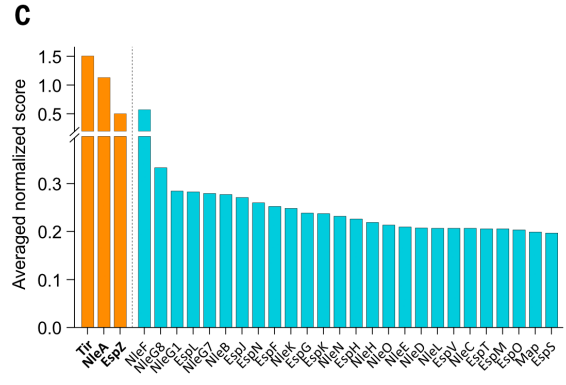
Notably, while the constitutively expressed proteasomal subunits PSMB5, PSMB6, and PSMB7 were down-regulated, the abundance of the immunoproteasomal subunits PSMB8, PSMB9, and PSMB10 was increased, which we also confirmed by qPCR (Fig. 3H). This suggests that IECs might participate in antigen presentation to infiltrating immune cells. Additionally, cellular adhesion molecules (CAMs) were positively enriched, whereas the TJ protein CLDN4 was down-regulated in IECs (fig. S4D). Thus, despite the vastly different effector networks, WT, CR14, and CRi9 trigger conserved molecular infection signatures.

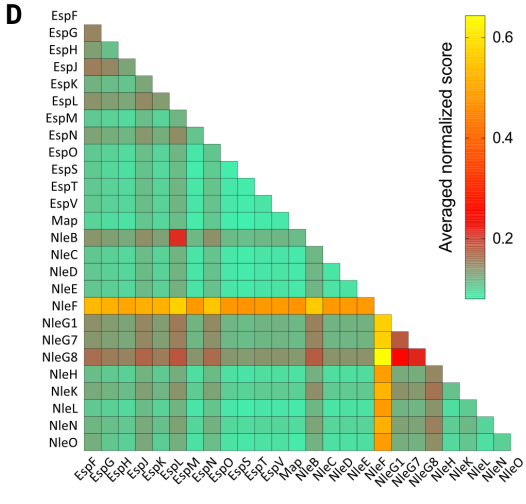
CRi9 triggers distinct inflammatory responses

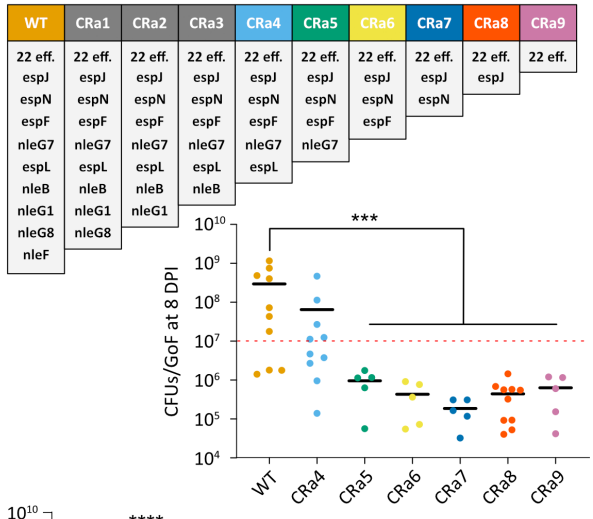
In line with the deletion of the anti-inflammatory effectors, infection with CRi9 resulted in the











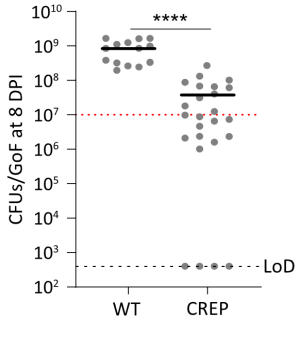


Fig. 7. Prediction of a specific essential effector combination. (**A**) A table illustrating the effector composition in additional *C. rodentium* mutants used to feed the ML model. Gray backgrounds indicate mutated effectors. Rows highlighted in green and red pinpoint mutants colonizing above or below the 10^7 CFU/GoF threshold at 8 DPI, respectively. (**B**) Steps used to train the ML model for finding noncolonizing mutants. The structure was manually fixed according to published relationships (fig. S7A). Training consists of assigning quantitative weight to every connection via a genetic algorithm. Good values for the hyperparameters were found by fivefold cross-validation. Fifty models were then trained with 90% of the dataset. Their predictions were averaged by weighting the models in the remaining 10% of the dataset and ranked by confidence. (**C**) Bar chart representing the importance score given to each effector by averaging the predicted confidence in no colonization of all the mutants that include that deletion. In the calculation of nonessential effectors (blue), mutants with essential deletions were excluded and the results were normalized in [0,1]. The score of essential effectors Tir, NleA, and EspZ (orange) takes into account all possible deletions; they ranked among the four highest in this classification. (**D**) Heatmap showing the importance of effector pairs that illustrates the codependency and interactions between pairs of nonessential effectors predicted by the model. Most pairings show low values, indicating additive relationships (green). Pairs including NleF showed high values due to its high individual score (yellow); NleB and EspL and pairings involving two of the NleG proteins show high scores (multiplicative relationships; red). (**E**) Schematic illustrating the *C. rodentium* mutants generated to test the ML model and bacterial load at 8 DPI (strains shown in fig. S7C). Graph shows means from n = 5 to 10 mice; statistical significance compared with WT colonization levels was determined by one-way ANOVA and Dunnett's multiple comparison test. ***P < 0.001. (**F**) CFU/GoF at 8 DPI with WT and CREP. Graph shows means ± SEM of three biological replicates; significance was calculated by an unpaired Mann-Whitney U test. ****P < 0.0001. Colonization threshold of 10^7 CFU/GoF is indicated by a dotted red line.

largest increase in abundance of major histocompatibility complex (MHC) class II proteins (e.g., H2-Aa), Tap1 (involved in antigen presentation on MHC class I), and antimicrobial and proinflammatory proteins, including iNOS, calprotectin, CAMP, REG3G, and NFKBIZ (Fig. 4, A to C). Induction of calprotectin, which sequesters the trace elements Mn and Zn (*48*), was validated by fecal S100A8 ELISA (fig. S4E), and increased gene expression of *Nos2* and *Reg3γ* was validated by qPCR. Notably, *Nfkbiz* was only transcriptionally elevated in CRi9 (Fig. 4C and fig. S4F). The abundance of multiple cytokines, cytokine receptors, and their downstream targets was also differentially regulated, with CRi9 exhibiting higher levels of Tnfrsf1b (TNFR2), CCL28 [chemokine (C-C motif) ligand 28], LtbR (lymphotoxin beta receptor), and IL6st (interleukin-6 signal transducer), all validated by qPCR (fig. S4G), and interferon-γ (IFN-γ)–regulated proteins (Fig. 4D), where *Irf8*, as well as *Icam1* and *H2-Aa*, were validated by qPCR (Fig. 4, B and D, and fig. S4D). Moreover, the lack of NleB and EspL in CRi9 resulted in significantly higher mixed lineage kinase domain–like protein (MLKL) phosphorylation, consistent with their role in the inhibition RIP kinases (Fig. 4E). Thus, deletion of the major anti-inflammatory effectors from the network in CRi9 results in changes to the abundance of specific inflammatory signatures in IECs compared with WT. By contrast, removal of a subset of the anti-inflammatory effectors from the network in CR14 (e.g., NleD1, NleD2, NleF, NleH, EspJ, and EspL) resulted in moderate changes to the abundance of proteins in the same processes.

CR14 and CRi9 trigger distinct mucosal immune response but also protective immunity

Considering the fact that WT, CR14, and CRi9 trigger distinct immune signaling in IECs, we analyzed how this was reflected in the levels of 13 cytokines [tumor necrosis factor (TNF), IFN-γ, IL-22, IL-17A, IFN-β, CXCL1, IL-23, IL-6, CXCL5, granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-27, IL-12p70, and IL-21] secreted from colonic explants or found in serum at 8 DPI. All cytokines except IL-21 and IL-12p70 were detectable in serum but did not show any significant differences after infection with CRi9 or CR14 when compared with the other treatment groups (fig. S5A). By contrast, we observed reduced secretion of CXCL1 after infection with WT but not with CR14 or CRi9, which suggests that the effector(s) involved in down-regulation of CXCL1 are absent in both mutants (fig. S5B). In line with changes to protein abundances in IECs (Fig. 4, A to D, and fig. S4, D, F, and G), infection with CRi9 resulted in significantly higher secretion of the leukocyte-secreted cytokines TNF, IL-17A, IL-22, and IFN-γ from colonic explants compared with WT and CR14 (Fig. 4F).

To determine the effect of perturbation of host immunity in the context of their respective effector networks, we infected IFN-γ knockout (KO) (*Ifng*$^{-/-}$) and littermate control mice (*Ifng*$^{+/+}$) with WT *C. rodentium* and CRi9, which were shed at comparable levels (fig. S5C). As expected, we detected elevated levels of secreted IFN-γ only in *Ifng*$^{+/+}$ mice infected with either WT *C. rodentium* or CRi9 (fig. S5D). Although infection of *Ifng*$^{-/-}$ mice with WT *C. rodentium* did not reveal substantial changes compared with littermate controls at 8 DPI, overlaying the CRi9 effector combination on a perturbed immune signaling network (i.e., *Ifng*$^{-/-}$) resulted in increased secretion of IL-17A and IL-22 coupled with an upward trend in secretion of IL-23 and a downward trend in TNF (fig. S5D). This coincided with significantly elevated levels of conventional dendritic cells (cDCs) and increased detection of CD4$^+$ T cells (P = 0.052) in CRi9-infected *Ifng*$^{-/-}$ colons (fig. S5E). Thus, in the absence of IFN-γ, an effector network lacking the anti-inflammatory effectors induces a shift from the combined T helper 1 (T_H1)–like (characterized in part by IFN-γ and TNF) and T_H17/T_H22 responses seen in WT *C. rodentium* infection (*49*) toward a dominant T_H17/T_H22-driven response.

GM-CSF is an immunoregulatory cytokine involved in the maturation of immune cell precursors, with a protective role in inflammatory bowel disease (IBD) and colitis (*50*, *51*), whereas IL-6 is a pleiotropic cytokine best known as a promoter of inflammation (*52*). The cytokine array revealed that WT *C. rodentium* infection results in inhibition of GM-CSF and IL-6 secretion from colonic explants, below the physiological levels seen in mock-infected mice (Fig. 4G). Notably, we observed significantly higher levels of GM-CSF and IL-6 in CR14– compared with WT-infected mice (Fig. 4G). This suggests that the CR14 effector network is missing effectors that specifically inhibit secretion of GM-CSF and IL-6. Although GM-CSF is mainly secreted by immune cells, it can also be secreted from nonhematopoietic cells (*51*). Thus, we performed qPCR to determine whether the reduced secretion of GM-CSF is reflected by the level of *Csf2* transcripts in IECs. This revealed that transcription of *Csf2* was equally inhibited after infection with all three strains (fig. S5F), indicating that after infection with CR14, GM-CSF is secreted from other cell types.

To determine whether differential secretion of cytokines affects immune cell infiltration, we analyzed the colons of mice infected with WT, CR14, and CRi9 by flow cytometry. WT and CRi9 recruited similar numbers of neutrophils, monocytes-macrophages, eosinophils, natural killer (NK) cells, B cells, and T cells, with a tendency toward greater counts in CRi9 that becomes significant in the overall numbers of cDCs (Fig. 5, A and B, and fig. S5, G and H). By contrast, despite the high levels of secreted GM-CSF, CR14 recruited significantly fewer myeloid cells to the colon, which was more pronounced for monocytes-macrophages and neutrophils (Fig. 5A and fig. S5G).

We investigated which of the 19 effectors missing in CR14 were involved in blocking the secretion of GM-CSF and IL-6 and the recruitment of myeloid cells by infecting mice with all of the intermediates between CR4 and CR9 (Fig. 2B and fig. S6A). Analysis of colonic

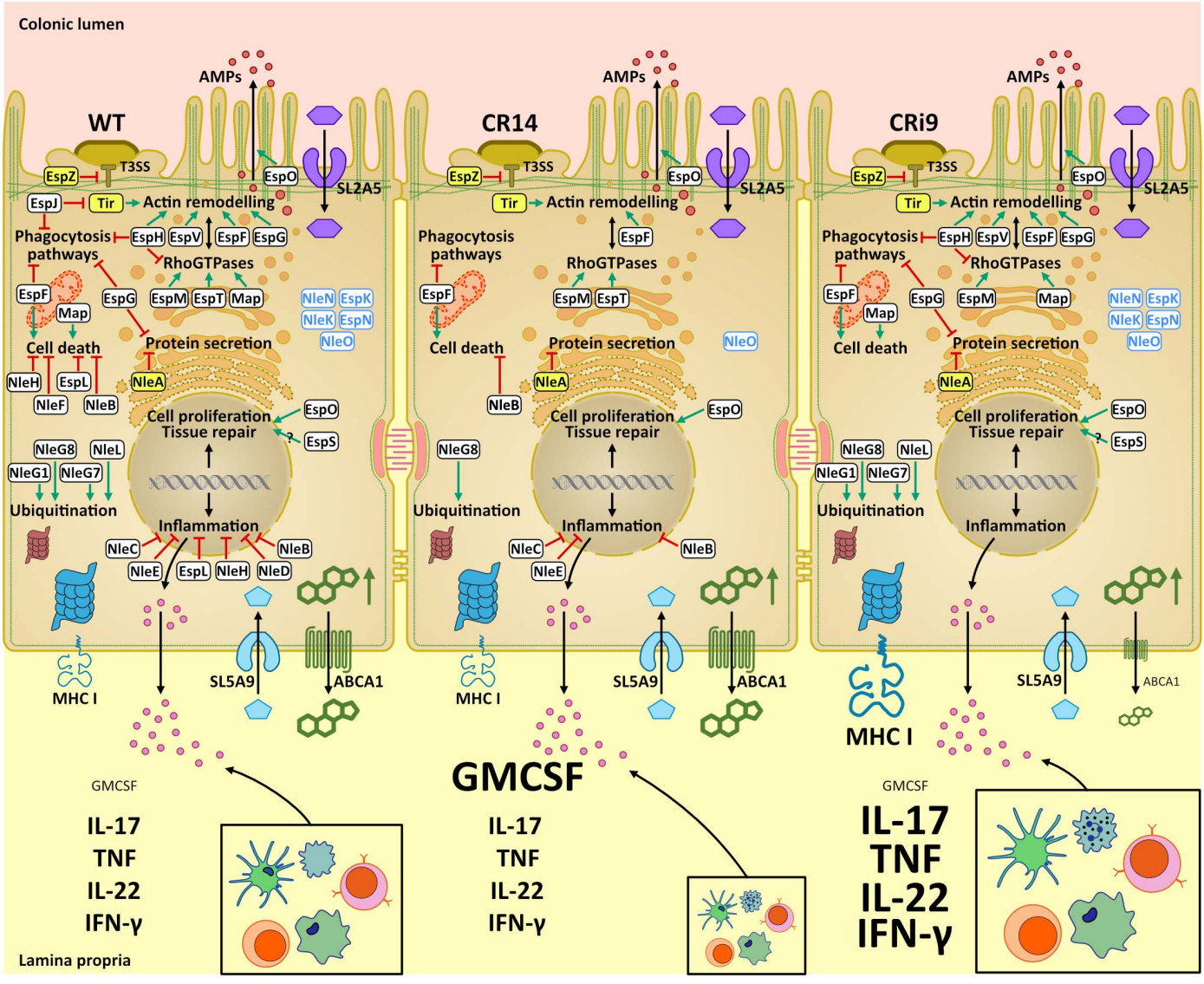


Fig. 8. Schematic representing the alternative *C. rodentium* effector networks and IEC responses to infection with WT, CR14, and CRi9. The master effectors Tir, EspZ, and NleA are highlighted in yellow, and green arrows and red lines symbolize stimulation and inhibition, respectively (the effector functions are detailed in table S1). Shown are both conserved effector-composition–independent changes [e.g., disruption of the mitochondria, up-regulation of glucose and fructose import (SLC5A9 and SLC2A5, respectively), cholesterol efflux (ABCA1), MHC I expression, and a switch from the constitutive proteasome (in maroon) to the immunoproteasome (in blue)] and effector-composition–dependent changes (e.g., alterations in immune responses, represented by cytokine secretion in the colonic environment and immune cell recruitment). Differences in size represent the magnitude of the change. AMPs, antimicrobial peptides.

explants revealed variable secretion trends of multiple cytokines, including CXCL1 and IFN-γ (fig. S6, B and C), which underlines the specific contribution to immune responses by different effector combinations. However, low levels of GM-CSF and IL-6 were observed in explants from mice infected with CR4, CR5, CR6, or CR7, whereas high levels were seen from mice infected with CR8 or CR9 (Fig. 5C and fig. S6B). Given that CR8 is CR7Δ*map*, we hypothesized that Map may be responsible for this phenotype. However, infection with the single Δ*map* mutant induced low levels of GM-CSF and IL-6 secretion (Fig. 5C and fig. S6B). This suggests that Map, in conjunction with other effectors, affects GM-CSF and IL-6 secretion from immune cells, which emphasizes the synergy and interdependency between effectors that leads to context-dependent immune modulation. Notably, analysis of myeloid (Fig. 5D and fig. S6D) and lymphoid cell (fig. S6E) recruitment to the colon revealed similar low levels in mice infected with either CR7 (low GM-CSF or IL-6) and CR8 (high GM-CSF or IL-6), thereby uncoupling colonic GM-CSF and IL-6 secretion from recruitment of the major immune cell types.

Considering the notable differences in immune responses triggered by CR14 and CRi9, we assessed their ability to elicit protective immunity from a subsequent challenge with WT *C. rodentium*. To this end, mice were infected with either WT (positive control), CR14, or CRi9, and mock infection was used as a negative control. WT and CR14 were similarly cleared between 16 and 22 DPI. By contrast, clearance of CRi9 was delayed (30 to 34 DPI) (Fig. 5E), consistent with the deletion of the immune-modulatory effectors (Figs. 4 and 5). Five weeks after infection, once the original infection had been cleared, the mice were reinfected with WT and the number of fecal *C. rodentium* was enumerated 8 days after reinfection. Although mock-infected mice developed typical infection, immunization with WT,

CR14, or CRi9 protected mice from reinfection (Fig. 5F). Thus, although pathogens use multiple effectors to erect roadblocks on the path of protective immunity, the host can navigate around these barricades, which demonstrates the inherent plasticity and robustness of antimicrobial responses.

Map and EspF activate IL-22 responses

EspF and Map, which disrupt TJs and the mitochondria (*53*), affect virulence and immune responses in the context of CR7—EspF shows context-dependent essentiality, whereas Map affects secretion of GM-CSF and IL-6. To determine their combined contribution to infection, we generated a *map*/*espF* double mutant (CRΔ*map*/Δ*espF*), which colonized mice similarly to WT, CR14, or CRi9 (fig. S6F). Proteomics analyses of purified IECs revealed that infection with CRΔ*map*/Δ*espF* triggered the conserved molecular infection signatures (fig. S6G and table S9) and that IFN-γ–mediated responses were comparable to those seen in WT-infected mice (fig. S6H).

Unexpectedly, infection with CRΔ*map*/Δ*espF* triggered significantly lower CCH, which correlated with relatively higher abundances of the goblet and deep crypt secretory (DCS) cell markers ClcA1 and Reg4 and lower abundances of the proliferation markers Ki67 and PCNA compared with WT (Fig. 6A). These changes corresponded to moderate changes in the abundance of proteins involved in cholesterol metabolism and bioenergetics (Fig. 6B) and a lower abundance of IL-22–regulated proteins, including calprotectin, which was also confirmed by ELISA (Fig. 6C). Consistent with this, we found significantly reduced secretion of IL-22 from colonic explants of CRΔ*map*/Δ*espF*-infected mice, compared with WT. By contrast, the levels of TNF and recruitment of neutrophils, cDCs, B cells, and $CD4^+$ T cells were not significantly different (Fig. 6, D and E), which underlines that Map and EspF specifically affect the IL-22 response. These results illustrate how a relatively small contraction of the effector network can markedly and specifically affect host responses to infection. Moreover, Map and EspF synergistically activate a feedback loop through which their injection into IECs modulates cytokine secretion from immune cells. This in turn affects the behavior of the epithelial monolayer. Because Map and EspF disrupt the TJs, this could be a key factor in triggering barrier regeneration responses.

ML identified a previously unidentified essential effector combination

Over the course of this study, we generated and characterized >100 different *C. rodentium* mutants in vivo (Fig. 7A and table S10). We used this dataset to train a ML model with the aim of predicting colonization efficiencies of *C. rodentium* bearing previously uncharacterized combinations of effector deletions. Because a dataset of this size hinders the use of common ML algorithms, which tend to overfit, we developed and applied a hybrid algorithm that incorporates domain knowledge (the known functions of the effectors; table S1) in addition to experimental data. This was depicted as a simplified interaction network that relates the effector targets and their known intracellular functions to colonization success (fig. S7A). The dataset served to assign quantitative weight to every interaction, making a model where the effects of the inputs to a node are considered to be additive. To adjust these weights, we developed a methodology combining artificial neural networks (ANN) with a genetic algorithm (Fig. 7B, fig. S7B, and Materials and methods). Hence, the trained model uses combinations of the 29 effectors (EspM2/3 and NleD1/2 were considered as one) as inputs and binary colonization success (using the 10^7 CFU/GoF shedding threshold) as the output (Fig. 7A and table S10).

To evaluate the influence of each effector in colonization, we used the trained model to score the 16,489,545 possible mutant combinations lacking up to nine effectors and rank them according to their expected colonization success—those with lower success being ranked first. From the raw predictions, we extracted the importance of both individual (Fig. 7C) and groups of effectors. The influence of single effectors in the ranked mutants varied among different runs (owing to the ANN nature of the model and the small dataset), but general patterns could be observed. As expected, the essential effectors Tir, NleA, and EspZ were among the four highest scores, together with NleF, which stood out among the nonessential effectors. The remaining effectors showed similar individual scores, implying that their exact positions may vary among runs, but they could be divided into two groups, consistent with those ranking higher (EspF, NleG8, NleK, NleN, EspH, EspN, EspL, NleG1, NleB, EspJ, NleG7, EspK, NleH, and EspG) or lower (NleE, NleO, NleL, EspV, NleD, EspM, EspT, Map, EspO, and EspS).

To examine the codependency of effectors, we analyzed the predicted influence of mutant pairs (Fig. 7D). Many combinations were essentially additive in relation to individual mutations, but some showed unexpected high scores: (i) NleB with EspL; (ii) any pair of NleG proteins, especially NleG8 with NleG1; and (iii) EspF with EspH, EspG, or EspJ. The effectors involved in these pairs share common functions, which suggests that the model is capable of predicting outcomes.

Additionally, we selected the best colonization failure predictions of five (CRa5: *nleF*, *nleG8*, *nleG1*, *nleB*, and *espL*) and nine (CRa9: *nleF*, *nleG8*, *nleG1*, *nleB*, *espL*, *nleG7*, *espF*, *espN*, and *espJ)* mutations, while discarding those supersets of colonization-inefficient mutants already in the dataset [external table (*54*)]. We generated the strains and their intermediates (CRa1 to CRa9), infected groups of five mice, and measured bacterial shedding at 8 DPI (Fig. 7E). Strains CRa5 to CRa9 did not colonize efficiently, as predicted by the model. Nonetheless, CRa4 (lacking *nleF*, *nleG8*, *nleG1*, and *nleB*), which ranks in second place among the four-effector mutants, shed above the 10^7 CFU/GoF threshold, revealing context-dependent EspL essentiality (Fig. 7E). Notably, of the 13 *Shigella flexneri* effectors that are functionally conserved with *C. rodentium*, four are deleted in CRa5 (table S11).

Thus, *C. rodentium* has acquired a collection of sophisticated effectors that can generate an enormous number of effector combinations with sufficient robustness and flexibility, which could also potentially contribute to host specificity. To test this hypothesis, we constructed a *C. rodentium* strain expressing the EPEC E2348/69 effector repertoire (CREP) (Fig. 2F and table S11). Infection of mice with CREP resulted in colonization at the 10^7 CFU/GoF borderline, which implicates effector networks in host adaptation (Fig. 7F).

Discussion

The aim of this study was to test the hypothesis that T3SS effectors form robust networks inside mammalian host cells. Using *C. rodentium* as a model, we first identified two additional effectors, NleN and NleO. From the complete set of 31 effectors, we generated >100 effector mutant combinations. We characterized the effect of contraction on the functionality of the effector network in vivo and established a model to predict host-pathogen interaction outcomes. This approach for interrogating biological networks exposed codependencies and context-dependent relationships in a native biological setting, while showcasing the robustness of T3SS effector networks (refer to the model in Fig. 8). The fact that the vastly different immune responses generated by WT, CR14, and CRi9 induced protective immunity illustrates the inherent flexibility and adaptability of the host in fighting infections by pathogens that are constantly evolving and expressing diverse effector networks. The robustness of host immune responses to gut infections is illustrated by the subtle differences in cytokine secretion and immune cell infiltration seen once the WT, or the perturbed effector network in CRi9, are overlaid on IFN-γ KO mice.

We used two distinct pathways to contract and test the robustness of an effector network. Unexpectedly, the initial approach revealed that contracting the network by 60%, through deletion of *nleL*, *espK*, *espM3*, *espV*, *espS*, *espL*, *nleG1*, *espG* (CR7), *espH*, *map*, *espN*, *nleD1* and *nleD2* (CR12), *nleK*, *espJ*, *nleG7*, *nleF*, or *nleH* and *nleN* (CR14), did not substantially affect

virulence. However, CR14 was the limit of the effector network robustness because it did not tolerate further contractions. Notably, a network missing 10 key effectors modulating immune responses (CRi9) also enabled colonization. While performing the second round of contractions, we confirmed that NleC and NleE target NF-κB in vivo. Deleting *espF* from CR7, as well as *nleC* and *nleE* from CR12 or *espO* from CRi9, but not from WT, resulted in shedding below the threshold. Thus, the outcome of infection may depend on a fine balance between effectors with diverse activities within a given network.

The phenomena of context-dependent effector essentiality and network robustness becomes prominent by analyzing the intricate functions of the multiple effectors affecting cell survival pathways. As was seen in *Yersinia* (*55*, *56*), it might be expected that inhibition of the prosurvival NF-κB by NleC and NleE during WT *C. rodentium* infection would lead to cell death. However, because no IEC death is seen during *C. rodentium* infection (*57*), inhibition of prosurvival processes seems to be mitigated by NleF, NleB, and EspL, which block activation of cell death pathways (*1*, *19*). In agreement with this hypothesis, whereas removal of NleC and NleE was tolerated in WT (expressing the cell death inhibitor effectors), it was not tolerated in CR12, which is missing EspL and NleF. Moreover, deletion of both NleB and EspL was not tolerated in CRa5, which expresses the prosurvival effector inhibitors. Thus, using our approach, we could infer these types of effector relationships from in vivo colonization phenotypes. More data and further permutations are needed for this kind of analysis to be applied to other effector combinations.

Studying individual effectors in vivo provides a simplified interpretation of their function, devoid of effector interdependency or leveraging capability, with the exception of master effectors, which form essential nodes within all networks. In *C. rodentium*, the master effectors include Tir, EspZ, and NleA (also known as EspI) (Fig. 8), which, in this order, are the three most efficiently translocated effectors (*57*). Although Tir (*30*) and EspZ (*35*) target the plasma membrane, the translation and translocation of NleA, which targets COPII vesicles (*34*), is coupled to the injection of Tir (*58*). These essential effectors—Tir, EspZ, and NleA—are common between *C. rodentium* and EPEC; however, reducing the *C. rodentium* effector network to the size of the E2348/69 network did not enable efficient colonization, which suggests that effector networks are involved in host adaptation. Applying the effector network principle to other pathogens could explain the processes leading to host specification of *Salmonella enterica* serovar Typhi through progressive effector pseudogenization (*59*).

Within the differentially regulated pathways, the increased abundance of NFKBIZ (also known as IκBζ), which is known to associate with the p50 subunit of NF-κB and has been associated with IBD (*60*–*62*), could play a role in the regulation of proinflammatory mediators (e.g., IL-6). Whereas several cytokines and Toll-like receptor (TLR) ligands, including IFN-γ and LPS, induce *Nfkbiz* expression, IL-17 stabilizes its intrinsically unstable mRNA (*60*, *63*, *64*). The higher levels of *Nfkbiz* transcripts seen exclusively in CRi9, which triggers heightened secretion of both IFN-γ and IL-17 and expression of proinflammatory genes, validate the functionality of the rewired cytokine signaling circuit. In particular, CR14, CRi9, and CRΔ*map*/Δ*espF* triggered the secretion of distinct cytokines and the recruitment of immune cell populations. The absence of the immuno-modulatory effectors in CRi9 affected local secretion of TNF, IL-17A, IL-22, and IFN-γ by lymphocytes. The elevated levels of these cytokines did not significantly affect recruitment of myeloid cells to the mucosa (although the number cDCs increased), which suggests that the levels of these cytokines in mice infected with WT *C. rodentium* are at saturation or that myeloid cell recruitment peaks before 8 DPI. Although we did not detect great differences in T cell and NK cell recruitment between WT and CR14, their numbers increased in mice infected with CRi9, which could be the source of the elevated cytokines.

How modulation of signaling by effectors in IECs translates to changes in cytokine secretion from immune cells remains unclear. The use of CRΔ*map*/Δ*espF* could provide a partial answer. Map and EspF disrupt the mitochondria (*53*) and TJs (*65*), which induces localized oxygenation of the microenvironment (*23*) and penetration of luminal metabolites to the submucosa (*31*, *65*). Because single *map* and *espF* mutants trigger typical CCH, we hypothesize that they trigger a synergistic signal that activates DCs, leading to secretion of IL-23 and resulting in IL-22 release from ILC3 or T cells (*19*). Notably, unlike WT, IL-22 KO mice do not succumb to infection with Δ*espF* (*31*), which suggests that IL-22 is needed to repair the EspF-induced damage to TJs. Another clue to this question could be found in the differential activation of LXR by CRi9. In addition to their roles in cholesterol efflux and lipid metabolism, LXRs are considered anti-inflammatory transcription factors, repressing the expression of inflammatory genes such as *Nos2*, and regulators of innate and adaptive immune responses (*46*, *66*). Although most of the studies have been performed in macrophages and, more recently, other immune cells such as T cells (*67*), it is likely that reduced LXR activation in IECs contributes to the exacerbated immune response triggered by CRi9.

Although infection with WT resulted in reduced levels of GM-CSF compared with mock-infected control mice, its secretion was increased ~40-fold in CR14. Unexpectedly, this phenotype coincided with significantly lower recruitment of neutrophils and monocytes-macrophages to the colonic mucosa. Testing our intermediate mutants showed that deletion of *map* from CR7, but not a single *map* mutant strain, was responsible for the reduced secretion of GM-CSF. Because qPCR revealed that infection with WT, CRi9, and CR14 leads to similarly lower levels of *Csf2* expression in IECs, the source of secreted GM-CSF is likely resident or recruited immune cells. We used CR7 and CR8 to determine whether the level of GM-CSF directly affects immune cell recruitment. The fact that both CR7 (low GM-CSF) and CR8 (high GM-CSF) triggered low infiltration of myeloid cells and lymphocytes into the colonic tissue dissociates GM-CSF secretion from cell recruitment.

We used our database in conjunction with the published activities of the effectors to train a ML model that predicts colonization success of previously uncharacterized mutant combinations. The prior biological knowledge mapped into the architecture of the neural network and the fixed and sparsely connected structure made the model easier to train despite the small dataset. Our algorithm differs from that used by Eetemadi and Tagkopoulos (*68*) because it only requires a single macroscopic output (colonization) instead of intracellular genetic measures. The predictive power of this approach is demonstrated by the validated discovery of the combination of effector gene mutations of CRa5 (*nleF*, *nleG8*, *nleG1*, *nleB*, and *espL*), which failed to colonize. Because the deletions in CRa5 include effector genes deleted from CR14 (*nleG8*, *nleG1*, *espL*, and *nleF*) and CRi9 (*nleB*, *espL*, and *nleF*), which are able to colonize, it would not have been an obvious combination of deletions to test without the aid of the model. CRa5, which is unable to colonize, lacks the best-ranked nonessential effector NleF and the effector pairs NleG8-NleG1 and NleB-EspL, thus supporting the predictions. The fact that CRa4 did colonize further supports the importance of the NleB-EspL pairing. Thus, the model has been much more efficient in finding a small set of mutations that reduce the colonization success compared with the empirical process of producing CR14 and CRi9. We anticipate that by extending the dataset with further mutant combinations, adding new nodes to the graph, and further resolving relationships such as temporal progression of infection and immune responses, we will be able to fine-tune the algorithm to better predict how different effector combinations will translate into infection outcomes and pathogen-host interactions in vivo.

Materials and methods

Bacterial cultures

Bacterial strains used in this work are listed in table S5. Strains were grown at 37°C on

Lysogeny broth (LB) agar plates (1.5% w/v), in liquid LB, or Dulbecco's modified Eagle's medium (DMEM). Nalidixic acid (Nal, 50 μg/ml), gentamicin (Gm, 10 μg/ml), and streptomycin (Sm, 50 μg/ml) were added for plasmid or strain selection, as required. For infection of CMT-93, *C. rodentium* strains were grown for 8 hours at 37°C (200 rpm) in a flask with 10 ml of liquid LB, inoculated in capped Falcon tubes (BD Biosciences) with 5 ml DMEM, and incubated overnight (o/n) at 37°C in a CO_2 incubator (static) for the induction of the T3SS.

DNA constructs, oligonucleotides, and generation of mutants

Plasmids used in this study are listed in table S5. Strains *E. coli* DH10BT1R or CC118λpir (*69*) were used as host for the cloning and propagation of plasmids. The homology regions (HR) for the deletion of effector genes were designed using Benchling software to be flanking each target gene, taking the 300 base pairs (bp) upstream HR just before the start codon and the 300 bp downstream HR just after the stop codon. HRs were synthesized through DNA synthesis (Thermo) and cloned into the SacI-SphI sites in pSEVA612S (*70*). PCR reactions were performed with Taq 2x Master Mix (NEB) for standard amplifications in screenings. All DNA constructs were fully sequenced (Eurofins).

Genes were deleted via tri-parental conjugation. Briefly, 20 μl of the helper strain *E. coli* 1047 pRK2013 were incubated with 20 μl of the donor strain (*E. coli* CC118-λpir pSEVA612S) for 2 hours at 37°C on LB agar. Next, 40 μl of the receiver strain (indicated *C. rodentium* strain with pACBSR) was added and the plate incubated for 4 hours at 37°C. Conjugants were selected on LB+Gm+Sm. To remove Gm resistance, conjugants were grown on 5 ml of liquid LB+Sm, supplemented with 0.4% L-arabinose, for 8 hours to induce expression of the I-SceI endonuclease from pACBSR, and streaked on LB+Sm plates. To remove pACBSR, strains were passaged several times in liquid LB, and bacteria sensitive to Sm were selected. Deletion mutants were screened by PCR for confirmation of the deletion using primers listed in table S5. All plasmids and deletion mutants were confirmed by sequencing (Eurofins).

ML for effector prediction

To predict T3SS effectors, we developed ML classification algorithms, similar to the ones we have previously described (*71*–*75*). The positive training data included 28 ORFs that were known to encode T3SS effectors. The negative training data included 1568 ORFs that had high sequence similarity (E value of 0 and minimum coverage of 70%) to ORFs in the nonpathogenic *E. coli* K12 genome (accession no. NC_000913.3). The full lists of positive and negative training sets are available in table S2. In total, 66 features were used for this ML, including homology (to *C. rodentium* effectors, to T3SS effectors of other mammalian pathogenic bacteria, to the host proteome, and to proteomes of 13 related *Citrobacter* strains without T3SS and five other bacteria with T3SS), composition [amino acid composition, guanine + cytosine (GC) content], and location in the genome (i.e., distance from known T3SS effector genes). Features were considered on the basis of their ability to discriminate effectors from noneffectors.

Homology features were extracted using BLASTp, such that for each dataset two features were obtained: the number of homologs in the datasets and the bit score of the best hit. These features were extracted for each ORF against three datasets: (i) effectors in *C. rodentium*; (ii) effectors in other bacteria; and (iii) the host proteome. The third dataset was used because effectors tend to encode eukaryotic motifs, which allow them to interact with host proteins.

Features related to the evolutionary conservation of putative effectors were also computed. (i) For each ORF, we calculated whether a homologous sequence exists in *Citrobacter* strains that do not encode T3SS. Specifically, this feature records the number of proteomes among the following 13 strains in which a homolog is found: *C. amalonaticus* FDAARGOS 122, *C. braakii* FDAARGOS 253, *C. europaeus* 67A, *C. farmer* AUSMDU00008141, *C. freundii* CRCB-101, *C. gillenii* UMG736, *C. koseri* ATCC BAA-895, *C. murliniae* P080C, *C. pasteurii* UMH17, *C. portucalensis* FDAARGOS 738, *C. sedlakii* MGYG-HGUT-01700, *C. werkmanii* BF-6, and *C. youngae* NCTC13709. It was calculated using BLASTp when a homolog was considered if the E value was ≤0.01. (ii) We ran BLASTp against the proteomes of five bacteria that encode a T3SS (specifically, *Shigella flexneri* 5b, *Shigella dysenteriae*, *Salmonella enterica* Typhimurium, *E. coli* O157:H7, and *E. coli* O127:H6) and looked for homologs in each. For each ORF we computed a binary feature for each of these datasets—1 if such a homolog was found in this bacterium and 0 otherwise. The values of these features for the effectors, representing their evolutionary conservation, are available in table S3.

GC content was calculated for each ORF and was used as a feature because some effectors tend to have atypical low GC content compared with noneffectors (suggesting that these effectors were obtained through lateral gene transfer) (*71*, *76*, *77*). Twenty features, which are the frequencies of each amino acid across the entire ORF, were also computed. The same 20 features were also computed for the N terminus of the protein, as the secretion signal of type 3 effectors resides in the N terminus. Biochemical characteristics of the N terminus, such as hydrophobicity and hydrophilicity, were calculated using AA index (*78*). Moreover, a feature that reflects the similarity of the amino acid profile of a given ORF to either effectors or noneffectors was computed. Specifically, a profile of amino acids frequencies was computed for each ORF and compared with the average amino-acid frequency profile of effectors and of noneffectors. Two Euclidean distances were then computed: between the amino acid profile vector of the given ORF and the average profile of effectors and noneffectors, respectively. The difference between these distances reflects the similarity to one of the groups compared with the other and was used as a feature.

Another group of features used were the genome organization features: effector-encoding genes tend to reside physically close to other effector genes in the genome, compared with noneffectors (*71*, *79*). Hence, for each ORF we computed two groups of genome-organization features: (i) distance from closest effector (i.e., number of ORFs separating the ORF for which this feature is computed from the closest effector-encoding ORF) and (ii) number of effector-encoding genes in the proximity of K ORFs, for K = 5, 10, 15, 20, 25, or 30. The complete list of features and their values for all the ORFs is given in table S3. Features were extracted using in-house Python scripts. The result of each ML run is a score for each ORF, reflecting its likelihood to encode a T3SS effector. We evaluated several classification algorithms: random forest (*80*), support vector machine (*81*), K nearest neighbors (KNN), linear discriminate analysis (LDA), logistic regression (LR) (as described in (*82*), and Voting, which predicts by averaging over all other ML algorithms. For each run, feature selection was performed. The ML algorithms and feature selection were based on the Scikit-learn module in Python (*83*). The area under the precision-recall curve (AUPRC) score over 10-fold cross-validation was used as a measure of the classifier performance. By weighting the predictions of the above algorithms, ML predictions in this study were based on the Voting algorithm as it gave the highest AUPRC score (a perfect 1.0).

TEM translocation assay

Swiss 3T3 cells were seeded in triplicate in a black-walled 96-well plate at 7×10^4 cells per well in DMEM (4500 mg of glucose per liter) supplemented with glutaMAX and 10% fetal bovine serum (FBS) and incubated o/n at 37°C with 5% (v/v) CO_2. WT *C. rodentium* was grown to saturation at 37°C with 200 rpm agitation for 8 hours before subculturing 1:500 for 16 hours in DMEM with 1000 mg of glucose per liter at 37°C, 5% (v/v) CO2, static to induce expression of the T3SS. Cells were infected with the induced culture at a multiplicity of infection (MOI) of 100, confirmed by retrospective plating. The plate was centrifuged at 1000 rpm for 5 min to synchronize infection. Cells were

cultured at 37°C with 5% (v/v) CO_2 for 2 hours before effector–TEM-1 expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated for a further 1.5 hours. Cells were washed in Hanks' balanced salt solution (HBSS) and incubated with freshly prepared HBSS supplemented with 6 mM probenecid (Sigma-Aldrich), 20 mM HEPES and 1 x CCF2/AM (ThermoFisher Scientific) (table S12, key resources table) in the dark for 1.5 hours. Cells were again washed in HBSS and fluorescence was measured using a Fluostar Optima (BMG Labtech, United Kingdom) at 450 and 540 nm. Graphs display the response ratio, which is defined as the 450/540 nm emission ratio, adjusted for background fluorescence from uninfected cells.

Phylogenetic analysis

Full-length effector protein queries were submitted to NCBI BLAST. The NleO cladogram was created with the first 100 BLASTp hits using ROD_40891 (NleO) as the query sequence. The tree was generated using NGPhylogeny.fr using default settings (*84*), and visualized using iTOL (*85*) (table S12, key resources table). For the NleN multiple sequence alignment (MSA), the primary sequences from relevant *C. rodentium* ROD_48841 (NleN) homologs in enteric pathogens were identified by PSI-BLAST and aligned using Clustal Omega v1.2.4 (*86*), with the default settings for proteins; JalView (*87*) was used for visualization.

C. rodentium *infections of CMT-93 cells*

The mouse cell line CMT-93 was grown in DMEM supplemented with 10% fetal bovine serum and 2 mM glutamine at 37°C in a 5% CO_2 incubator. For infection, cells were seeded out on glass coverslips in 24-well plates at a concentration of 10^5 cells/well. One hundred microliters of the o/n cultures were added to fresh medium in each well (MOI of 100). Plates were centrifuged at 1000 rpm for 5 min at room temperature (RT) and incubated for 4 hours at 37°C in a 5% CO_2 atmosphere. The infection was stopped with three washes with sterile phosphate-buffered saline (PBS). Cells were fixed with 4% (w/v) paraformaldehyde in PBS for 20 min at RT and permeabilized in 0.1% (v/v) of Triton X100 in PBS for 10 min. All antibodies were diluted in PBS with 1% bovine serum albumin (BSA), and rabbit polyclonal anti–*C. rodentium* (1:500) was used for detection. After incubation for 1 hour at RT, coverslips were washed three times with PBS, and incubated for 45 min with secondary antibody, Alexa647-conjugated anti-rabbit IgG, and Tetramethylrhodamine (TRITC)–conjugated phalloidin (1:500) and 4′,6-diamidino-2-phenylindole (DAPI) (1:500) to label F-actin and DNA, respectively (table S12, key resources table). Coverslips were washed three times with PBS after incubation, mounted in ProLong Gold antifade reagent (Thermo), and analyzed with a Zeiss AxioVision Z1 microscope (Zeiss) and ImageJ software.

Mouse infections

All animal experiments complied with the Animals Scientific Procedures Act 1986 and U.K. Home Office guidelines and were approved by the local ethical review committee. Experiments were designed in agreement with the ARRIVE guidelines (*88*) for the reporting and execution of animal experiments, including sample randomization and blinding. Mouse experiments were performed with five mice per group. Pathogen-free female 18 to 20 g C57BL/6 mice (Charles River Laboratories) and *Ifng* KO (and littermate controls) (The Jackson Laboratory) were housed in high-efficiency particulate air (HEPA)–filtered cages with sterile bedding and given food and water ad libitum. For infections, the respective *C. rodentium* strains were grown o/n in 15 ml of LB (50 ml capped Falcon tubes) at 37°C and 170 rpm. The next day, the tubes were centrifuged, and the pellet was resuspended in 1.5 ml of sterile PBS. Mice were infected by oral gavage with 200 μl of this solution (10^9 CFUs). Infections were followed by plating mouse stools at 3, 6, 7, and 8 DPI, and mice were culled at 8 DPI, unless otherwise indicated. Colons were extracted for downstream processing.

In the experiments shown in Fig. 2E, mice were infected with the indicated *C. rodentium* strains as described above. On day 3 after infection, mice were given 10 mg/kg of the IKK inhibitor BMS-345541, or vehicle alone, via oral gavage. BMS-345541 solution was prepared immediately before administration and dissolved in 2% DMSO, 30% PEG 300, and 2% Tween80 in water. For the experiment described in Figs. 3, 4, and 5; figs. S3 and S4; and fig. S5, A and B, 15 mice were infected with strains CR14 and CRi9, whereas five mice were infected with ICC169 and five mice were mock-infected with PBS. Five mice from each group showing a similar infection pattern were selected for further analysis (for reproducibility). At 8 DPI, the colon was extracted, and the distal 0.5 cm was used for histological analysis. The next 4 cm were divided in two, longitudinally. The first half was dedicated for extraction of IECs, and the second half was used for explant analysis. For the experiments described in Fig. 5, E and F, mice infections were followed until total clearance of the bacteria, defined as three consecutive *C. rodentium* negative stool samples; mice were then reinfected with WT strain ICC169 to follow bacterial shedding for an additional 8 days. In the experiments using the CRΔ*map*/Δ*espF* shown in Fig. 6 (and fig. S6, E to G), only results from mice colonized above the threshold (10^7 CFU/GoF) were analyzed (for reproducibility).

Histological analysis and immunostaining

The distal 0.5 cm of the colon was fixed in 4% paraformaldehyde (PFA) for 2.5 hours and then submersed in 70% ethanol. Fixed tissues were processed, paraffin embedded, and sectioned at 5 μm. The sections were then stained with either hematoxylin and eosin (H&E) or processed for immunofluorescence. Crypt hyperplasia was determined by measuring the lengths of at least 10 well-oriented crypts from a H&E-stained section of each mouse. All histological sections were evaluated blindly, and the mean of each mouse was plotted for statistical analysis. For immunofluorescence, sections were dewaxed by submersion in Histo-Clear solution (VWR) twice for 10 min, 100% ethanol twice for 10 min, 95% ethanol twice for 3 min, 80% ethanol once for 3 min, and PBS-0.1% Tween 20-0.1% saponin (PBS-TS) twice for 3 min. Subsequently, sections were heated for 30 min in demasking solution (0.3% trisodium citrate-0.05% Tween 20 in distilled H_2O). Once cooled, slides were first blocked in PBS-TS supplemented with 10% normal donkey serum (NDS) for 20 min in a humid chamber, before being incubated with primary antibody diluted in PBS-TS with 10% NDS for 1 hour. Sections were incubated with anti-PCNA antibody (1:500) and anti-intimin (1:50) or phospho-MLKL (P-MLKL) (1:50). To take into account CCH, the mean percentage of MLKL-P positive (MLKL-P^+) cells per micrometer of crypt for each mouse was calculated by dividing the total number of MLKL-P-positive cells counted in each of the well-oriented crypts by the total crypt length of the crypt (in micrometers) from which they were counted. Slides were rinsed twice for 10 min each time in PBS-TS, followed by incubation with secondary antibodies (1:100) and DAPI (1:1000) (details about the reagents used and antibodies can be found in table S12). Washing steps were repeated before slides were mounted with ProLong Gold antifade mountant (Thermo Fisher Scientific). Images were acquired using a Zeiss AxioVision Z1 microscope, using an AxioCam MRm camera, and processed using Zen 2.3 Blue Version (Carl Zeiss MicroImaging GmbH, Germany).

ELISA for fecal samples

Weighed mouse stool samples were homogenized in PBS with 0.1% Tween 20 (PBST) using a vortex machine for ~15 min at RT, where 1 ml of PBST was added per 0.1 g of stool. Samples were centrifuged at 16,000 rpm for 10 min to remove fecal debris, and the supernatant was stored at −80°C. The concentration of fecal S100A8 was determined with the DuoSet mouse S100A8 ELISA (R&D Systems; table S12). The assay was performed following the manufacturer's recommendations. Readings were obtained using a FLUOstar Omega microplate reader (BMG biotech).

Extraction of IECs for proteomics

At 8 DPI, a 4-cm segment of terminal colon was cut longitudinally and half of it was placed

in 4 ml of enterocyte dissociation buffer (1× HBSS without Mg and Ca with 10 mM HEPES, 1 mM EDTA, and 0.5% of β-mercaptoethanol), and incubated at 37°C with shaking for 45 min. The remaining tissue was removed and stored at −80°C and detached IECs were collected by centrifugation (2000 × g for 10 min), followed by three washes with ice-cold PBS; cells were centrifuged at 4°C after the first spin. IEC pellets were stored at −80°C until proteomics analysis.

Sample preparation and TMT labeling

For the final analysis, we included mice in which satisfactory colonization level after the infection was reached (>1 × 10^7 CFU/GoF by 8 DPI). IEC pellets isolated from five mice per condition per biological replicate were solubilized using probe sonication in lysis buffer [100 mM triethylammonium bicarbonate (TEAB), 1% sodium deoxycholate (SDC), 10% isopropanol, 50 mM NaCl] supplemented with Halt Protease and Phosphatase inhibitor cocktail (Thermo Scientific), boiled for 5 min at 90°C, and sonicated once more. Protein concentration was measured with Coomassie Plus assay (Thermo Scientific) following manufacturer's protocol. One hundred micrograms of protein, with an equal contribution from each individual mouse within the group, were reduced with 5 mM tris-2-carboxyethyl phosphine (TCEP) for 1 hour, followed by alkylation with 10 mM iodoacetamide (IAA) for 30 min, then digested by adding trypsin (Pierce) at final concentration 75 ng/μl to each sample and incubating the samples for 18 hours at RT. Peptides were labeled with tandem mass tag (TMT) multiplex reagent (Thermo Scientific) following manufacturer's protocol. SDC was precipitated with formic acid (FA) at final concentration of 2% (v/v) and centrifugation for 5 min at 10,0000 rpm. Supernatant containing TMT-labeled peptides was dried with a centrifugal vacuum concentrator.

High-pH reversed-phase peptide fractionation and liquid chromatography–mass spectrometry analysis

Before mass spectrometry (MS) analysis, TMT-labeled peptides underwent high-pH reversed-phase (RP) fractionation using the Waters XBridge C18 column (2.1 × 150 mm, 3.5 μm) on a Dionex Ultimate 3000 high-performance liquid chromatography (HPLC) system. Mobile phase A was 0.1% ammonium hydroxide, whereas mobile phase B was 100% acetonitrile and 0.1% ammonium hydroxide. The peptides were separated with a gradient elution at 200 μl/min including the following steps: isocratic for 5 min at 5% phase B, gradient for 40 min to 35% phase B, gradient to 80% phase B in 5 min, isocratic for 5 min, and re-equilibrated to 5% phase B. Fractions were collected in a 96-well plate every 42 s to a total of 65 fractions, then concatenated into 28 fractions, dried, and reconstituted in 50 μl 0.1% formic acid. The samples were analyzed on an Orbitrap Fusion and Orbitrap Fusion Lumos Trybrid mass spectrometers. A Dionex Ultimate 3000 system and and mass spectrometer (Thermo Scientific) were used for data acquisition. From each fraction 10 μl was injected onto a C18 trapping column (Acclaim PepMap 100, 100 μm × 2 cm, 5 μm, 100 Å) at a 10 μl/min flow rate. The samples were subjected to a low-pH gradient elution on a nanocapillary reversed phase column (Acclaim PepMap C18, 75 μm × 50 cm, 2 μm, 100 Å) at 45°C. Mobile phases A and B were 0.1% formic acid and 80% acetonitrile, 0.1% formic acid, respectively. The separation was performed at 300 nl/min flow rate and 90 min gradient from 5 to 38% phase B followed by 10 min up to 95% phase B, isocratic for 5 min at 95% B, re-equilibrated to 5% phase B in 5 min, and isocratic for 10 min at 5% phase B. MS1 scans were collected with mass resolution of 120,000, automatic gain control of 4 × 10^5, and injection time of 50 ms. Precursor ions (Top Speed mode, 3 s) were fragmented with collision-induced dissociation (CID) with a quadrupole isolation width of 0.7 Th (Thomson unit). Collision energy was set at 35%, with AGC at 1 × 10^4 and IT at 50 ms. Quantification was obtained at the MS3 level with higher-energy collisional dissociation (HCD) fragmentation of the top seven most abundant collision-induced dissociation (CID) fragments isolated with synchronous precursor selection (SPS). Quadrupole isolation width was set at 0.7 Th, collision energy was applied at 65%, and the AGC setting was at 1 × 10^5 with IT at 105 ms. The HCD MS3 spectra were acquired for the range 100 to 500 m/z with a resolution of 50,000. Targeted precursors were dynamically excluded for further fragmentation for 45 s with 7 ppm mass tolerance.

Protein identification and TMT-based quantification

Tandem mass spectrometry (MS/MS) spectra were searched against UniProt annotated reference proteomes of *Mus musculus* and *C. rodentium* in Proteome Discoverer v2.3 and v2.4 (Thermo Scientific) with the SEQUEST-HT engine using following parameters: 20 ppm precursor mass tolerance and 0.5 Da fragment ion mass tolerance, the static modifications were specified as carbamidomethylation of Cys residues and TMT modification of peptide N-term and Lys residues, and dynamic modifications were specified as oxidation of Met and deamidation of Asp and Glu residues. The search was restricted to fully tryptic peptides with maximum of two missed-cleavages. Based on the *q* value and decoy database search confidence of peptide discovery was estimated at 1% FDR with percolator node. The reporter ion quantifier node included a TMT-10-plex quantification method with an integration window tolerance of 15 ppm and integration method based on the most confident centroid peak at MS3 level. Protein groups with only unique peptides were used for quantification.

Proteomics data normalization, analysis, and visualization

To compare the results between sample batches, the datasets were normalized in various steps. For proteins with at least 60% of the TMT channels present, the remaining missing values were replaced with the minimum value in the TMT batch. For mouse proteins, normalization factor for each condition was calculated by the sum of all raw abundances in each TMT channel and dividing it by the maximum sum within given TMT-plex. To correct for the bacterial load variation in the different strains of *C. rodentium*, the median abundances of the *Citrobacter* proteins were computed per TMT channel and regressed out from each *Citrobacter* protein across samples. For %CV calculations and PCA, each biological group was centered to its mean within their respective TMT-batch. For further analysis, the normalized data were used to compute $\log_2$ ratios versus the uninfected samples.

For further analysis, the technical replicates of the same biological repeat included in both TMT batches were averaged. Gene ontology (GO) terms slim name, InterPro name, KEGG name, Keywords name, Pfam name, Reactome name, and SMART name enrichment analysis was performed with one-dimensional (1D)–annotation enrichment method (*89*) in the Perseus software (*90*).The data were visualized using the online Phantasus tool (*91*) (table S12, key resources table). Terms with significant positive enrichment were addressed as up-regulated and those with negative enrichment as down-regulated. All terms were filtered for a Benjamini-Hochberg FDR of <0.05. Infection signature and differential expression *P* values were calculated using one-sample *t* test and two-sample *t* test, respectively, in the Perseus proteomic tool. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*92*) partner repository.

RNA isolation and reverse transcription quantitative PCR (RT-qPCR)

RNA from lysed IECs and intimately attached *C. rodentium* was isolated using the RNeasy minikit (Qiagen) following the manufacturer's instructions, and RNA concentration was determined using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific). One microgram of RNA was treated with RQ1 RNase-free DNase for 30 min (for mouse transcript analysis) or 1 hour (for *C. rodentium* transcript analysis) at 37°C (Promega), and subsequently, cDNA was synthesized using a Moloney murine leukemia virus (MMLV) reverse transcription kit (Promega) following the manufacturer's

protocol. As a negative control to check for the presence of remaining DNA, a reaction mixture without the MMLV reverse transcriptase was also included (NRT). qPCR was performed using PowerUp SYBR green master mix (ThermoFischer Scientific) and the corresponding primer pairs to the gene being tested (table S12, key resources table); the assay was run on a StepOnePlus System (Applied Biosystems) and results were analyzed using the StepOne software (Applied Biosystems). Relative gene expression levels were analyzed by using the $2^{-\Delta CT}$ (where *CT* is threshold cycle) method for bacterial genes and $2^{-\Delta\Delta CT}$ for mammalian genes. The specificity of the primers against bacterial RNA was tested (i) using the NRT sample as a negative control and (ii) performing the qPCR in uninfected IEC samples, where the *CT* values obtained were undetectable or very high.

Cholesterol assay

Total serum cholesterol (cholesterol esters and free cholesterol) was quantified using a colorimetric reaction, as per manufacturer recommendations (Cell Biolabs) (table S12, key resources table). The colorimetric signal was analyzed using a FLUOstar Omega microplate reader (BMG biotech) in the 540 to 570 nm range.

Explant and serum cytokines and chemokines

Blood from mice was extracted after cervical dislocation and allowed to coagulate at RT for at least 3 hours. Blood was then centrifuged at 2000 rpm for 15 min to pellet the blood clots, and the supernatant (serum) was collected and immediately stored at −80°C for later use.

For explants colon tissue without feces was excised, weighed and washed with RPMI medium with 100 μg/ml of streptomycin and 100 U/ml of penicillin. Tissues were cultured in RPMI medium for 2 hours to remove residual cytokines and chemokines present on the cell surface as a result of the tissue perturbation and excision. To assess immune activity of the colonic tissue, the explants were then washed and placed in fresh complete RPMI media (0.1 ml of media per 10 mg of tissue) and cultured for 24 hours at 37°C, 5% CO_2. The supernatant was collected, centrifuged 10 min at 15,000 rpm to remove cellular debris and immediately stored at −80°C for later use.

Explant and serum cytokine levels were assessed from the corresponding samples via LEGENDplex (custom-made mouse panel 13-plex; BioLegend) kit per manufacturer's instructions. The panel included TNF, IFN-γ, IL-22, IL-17A, IFN-β, CXCL1, IL-23, IL-6, CXCL5, GM-CSF, IL-27, IL-12p70, and IL-21. Cytokine levels were acquired using a FACSCalibur flow cytometer (BD Biosciences), and analyses were performed using LEGENDplex data analysis software (BioLegend).

For each cytokine, a biological repeat was considered valid only if at least 90% of the samples were above the limit of detection; any values below the detection limit within a repeat such as this were shown as lowest value detectable by the assay. In serum, IL-21 was only consistently within the detection range in one out of three biological repeats, and thus values are not shown; IL-12p70 was always below the detection limit. In explant supernatants, we only consistently detected TNF, IFN-γ, IL-22, IL-17A, and CXCL1; CXCL5, IL-21, and IL-27 were only above the detection limit in one in three repeats (the rest below the limit of detection). Notably, IL-23 is detectable at 4 DPI in explants supernatants after *C. rodentium* WT infection, but not at 8 DPI, the time point investigated in this study.

The IL-22 and TNF concentration in explants shown in Fig. 6D was determined using a Mouse IL-22 DuoSet ELISA and a TNF-α Mouse Uncoated ELISA Kit following the manufacturer's instructions (table S12).

Preparation of cells from colon tissue and flow cytometry

Colon was extracted, opened longitudinally to remove feces, and washed in HBSS (Gibco). Tissue was then chopped into small pieces and transferred into a c-Mac (Miltenyi Biotec) tube containing RPMI supplemented with 10% FBS, penicillin and streptomycin solution (final concentration of 100 U and 100 μg/ml, respectively), HEPES (10mM), and Sodium Pyruvate (1mM), all from Sigma (hereafter called complete RPMI). Complete RPMI supplemented with Liberase TM (0.13 mg/ml final, Roche) and DNaseI (10 μg/ml; Sigma-Aldrich) was added, to reach a total of 4 ml of media per tube (table S12). For the IFN-γ KO experiments, IECs were first removed from the tissue as described above. The tissue was then washed in RPMI to remove remaining EDTA from IEC extraction buffer and then sliced and transferred into a c-Mac tube containing 4 ml of RPMI supplemented with DNaseI and Liberase TM.

A first mild homogenization step was carried out using a gentleMACS dissociator (Miltenyi Biotec), followed by enzymatic digestion for 40 min on a shaker at 37°C. A second homogenization step with the gentleMACS dissociator was performed after addition of 10 mM EDTA to the samples. The cells were filtered through a 70-μm cell strainer, spun down and resuspended in 600 μl of complete RPMI. From this point forward, steps were carried out on ice to preserve the cells.

For staining, ~5×10^5 cells (~100 μl) from each sample were added per well in a 96-well V-bottom plate. Dead cells were routinely excluded with Zombie Aqua Fixable Dead Cell Stain (ThermoFisher Scientific). Single cell suspensions were incubated with Fc block (Biolegend) in FACS buffer (1 BSA, 2mM EDTA), followed by staining with the conjugated antibodies indicated in the key resources table in FACS buffer, all at 4°C in the dark; FMOs were always included as controls. Briefly, neutrophils were defined as $CD11b^+$ $Ly6G^+$ cells; monocytes/macrophages as $CD11b^+$ $F4/80^+$ $CD64^+$ cells; conventional dendritic cells cDCs as $CD11b^+$ $CD64^-$ $CD11c^+$ cells; eosinophils as $CD11b^+$ $CD11c^-$ $siglecF^+$. NK cells were identified as $CD3^-$ $NK1.1^+$ $NKp46^+$; B cells as $CD45^+$ $CD3^-$ $B220^+$ and T cells as $CD45^+$ $B220^-$ $CD3^+$ (either $CD4^+$ or $CD8^+$) (gating strategy shown in fig. S8) Cells were then washed and fixed for 20 min with 1% paraformaldehyde in PBS; after fixing the cells were kept in the dark at 4°C until analysis. Single stain controls for compensation were prepared using VersaComp beads (Beckman Coulter) and ArC Amine Reactive Compensation beads (ThermoFisher Scientific) and flow cytometry analysis on 20,000 live cells was performed on a BD LSRFortessa cell analyzer (BD Biosciences). Data were analyzed using FlowJo software (Tree Star); absolute numbers of cells in the sample were calculated using CountBright absolute beads (ThermoFisher Scientific).

ML algorithm for predicting colonization success

In a preliminary exploration with Weka software (*93*), which included rules, trees, and Bayesian classifiers, the resulting trained models relied on untrue relationships between the deletions in the dataset and were inconsistent with our knowledge (e.g., setting as essential some effectors known to be not essential).

We thus restricted a connectionist discriminative classifier to prior knowledge by representing the known interactions between the effectors and the mouse IEC intracellular processes (table S1) as a directed acyclic graph (fig. S7A). The graph resembles an abstraction of the real protein interaction network, with the edges going from the node that influences to the influenced one, being the 29 effectors the input nodes (no parents) and colonization the single output (no children). We omitted nodes with a single parent—unless it implied getting an effector directly connected to the output—and cut any feedback.

The quantitative trainable part was inspired by feed-forward ANN (*94*). The value of the effectors (1 if present, 0 if absent) is transmitted through the network toward the output node by performing a calculation with two steps at every node: (i) a linear combination over the parent nodes and (ii) a nonlinear activation function. The weights of the linear combination are the trainable parameters of the model. Their sign was restricted in those cases where the sign of the interaction (positive or negative) was known. As (i) only includes first order terms, only additive noncooperative interactions were modeled, assuming they

account for the biggest part of the influence. We selected a saturating exponential with a [0,1] range as the activation function for internal nodes. It is biologically meaningful and simple (needs no additional parameters) and, to our knowledge, has not been used before for this aim

$$f(x) = \begin{cases} 0 \text{ if } x < 0 \\ 1 - e^{-x} \text{ if } x \geq 0 \end{cases}$$

A sigmoid is applied at the output node. The weights were adjusted applying a real-coded steady-state multipopulation genetic algorithm (fig. S7B) with binary cross-entropy loss as fitness function. We used proportional roulette-wheel selection of seven parents; a morphological crossover operator (MMX) (*95*) and stochastic mutation to generate two offspring that deterministically replace the two worst solutions. Several isolated populations were randomly and uniformly initialized and then evolved in parallel, allowing migration at certain time points. We preset some of the hyperparameters (*95*) and manually tuned the rest by a stratified fivefold cross-validation (Fig. 7B).

To train the model, we prioritized precision by applying 80% of the fitness cost to the cases with positive colonization. We created an ensemble with 50 trained models (suboptimal but diverse, all with weighted accuracy >90% in validation set) and averaged their predictions to rank combinations of mutations from one to nine effectors.

To offset the influence of the training parameters in the range of values for the colonization outputs, they were normalized in [0,1] (1 corresponds to the lowest output). Then, we calculated a score for each effector, as well as for groups of two and three effectors, by averaging the mutants that include those deletions. We repeated this process within each number of deletions (one to nine) and then averaged to compensate for the different number of mutants present in each category.

Statistics

No statistical methods were used to determine sample size. For ELISA and RT-qPCR, two technical replicates were used to estimate experimental mean. Statistical significance between more than two normally distributed groups was assessed by a one-way analysis of variance (ANOVA). Statistical analysis of proteomics data are detailed in the section above (Proteomics data normalization, analysis, and visualization). When data were not normally distributed (based on D'Agostino-Pearson or Shapiro-Wilk normality tests), a logarithm transformation was applied and were then analyzed by parametric statistical methods (ANOVA or student *t* test). When normality was not achieved by transformation, a non-parametric test was applied (Kruskal-Wallis test followed by Dunn's test). One-way ANOVA was used to analyze microscopy (CCH and % PCNA), ELISA and LEGENDplex results, cell recruitment, and RT-qPCRs; when more than three comparisons were made, *P* values were adjusted for multiple comparisons as indicated in the figure legends (Dunnett's or Tukey's post hoc test). Statistical analysis was only performed with datasets where each group analyzed had two or more biological repeats. After the establishment of a colonization threshold of 10^7 CFU/GoF (Fig. 1F), colonization data are mainly used to define successful colonization; colonization levels between CR14, CRi9, and CRΔ*map*/Δ*espF* are comparable (no significant different by one-way ANOVA, corrected for multiple comparisons by Tukey's post hoc test). Data plotting and statistical analysis were performed using Prism 9.0.0 (GraphPad software Inc). Statistical details of experiments are described in the figure legends. A *P* value <0.05 is considered statistically significant.

REFERENCES AND NOTES

1. A. R. Shenoy, R. C. D. Furniss, P. J. Goddard, A. Clements, in *Escherichia coli, a Versatile Pathogen*, G. Frankel, E. Z. Ron, Eds. (Springer, 2018), pp. 73–115.
2. N. Kamada *et al*., Humoral Immunity in the Gut Selectively Targets Phenotypically Virulent Attaching-and-Effacing Bacteria for Intraluminal Elimination. *Cell Host Microbe* **17**, 617–627 (2015). doi: 10.1016/j.chom.2015.04.001; pmid: 25936799
3. A. Jacobson *et al*., A Gut Commensal-Produced Metabolite Mediates Colonization Resistance to *Salmonella* Infection. *Cell Host Microbe* **24**, 296–307.e7 (2018). doi: 10.1016/j.chom.2018.07.002; pmid: 30057174
4. A. J. Bäumler, V. Sperandio, Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016). doi: 10.1038/nature18849; pmid: 27383983
5. C. Mullineaux-Sanders, J. Suez, E. Elinav, G. Frankel, Sieving through gut models of colonization resistance. *Nat. Microbiol.* **3**, 132–140 (2018). doi: 10.1038/s41564-017-0095-1; pmid: 29358683
6. T. R. D. Costa *et al*., Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015). doi: 10.1038/nrmicro3456; pmid: 25978706
7. A. Best, Y. Abu Kwaik, Evolution of the Arsenal of *Legionella pneumophila* Effectors To Modulate Protist Hosts. *mBio* **9**, e01313-18 (2018). doi: 10.1128/mBio.01313-18; pmid: 30301851
8. L. Pinaud, P. J. Sansonetti, A. Phalipon, Host Cell Targeting by Enteropathogenic Bacteria T3SS Effectors. *Trends Microbiol.* **26**, 266–283 (2018). doi: 10.1016/j.tim.2018.01.010; pmid: 29477730
9. J. Garmendia *et al*., TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell. Microbiol.* **6**, 1167–1183 (2004). doi: 10.1111/j.1462-5822.2004.00459.x; pmid: 15527496
10. K. G. Campellone, D. Robbins, J. M. Leong, F. Esp, EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Dev. Cell* **7**, 217–228 (2004). doi: 10.1016/j.devcel.2004.07.004; pmid: 15296718
11. T. Kubori, J. E. Galán, Temporal regulation of *Salmonella* virulence effector function by proteasome-dependent protein degradation. *Cell* **115**, 333–342 (2003). doi: 10.1016/S0092-8674(03)00849-3; pmid: 14636560
12. A. Mousnier *et al*., A new method to determine *in vivo* interactomes reveals binding of the *Legionella pneumophila* effector PieE to multiple rab GTPases. *mBio* **5**, e01148-14 (2014). doi: 10.1128/mBio.01148-14; pmid: 25118235
13. M. A. Pallett *et al*., Bacterial virulence factor inhibits caspase-4/11 activation in intestinal epithelial cells. *Mucosal Immunol.* **10**, 602–612 (2017). doi: 10.1038/mi.2016.77; pmid: 27624779
14. S. L. Slater, G. Frankel, Advances and Challenges in Studying Type III Secretion Effectors of Attaching and Effacing Pathogens. *Front. Cell. Infect. Microbiol.* **10**, 337 (2020). doi: 10.3389/fcimb.2020.00337; pmid: 32733819
15. J. Ding *et al*., Structural and Functional Insights into Host Death Domains Inactivation by the Bacterial Arginine GlcNAcyltransferase Effector. *Mol. Cell* **74**, 922–935.e6 (2019). doi: 10.1016/j.molcel.2019.03.028; pmid: 30979585
16. J. S. Pearson *et al*., A type III effector antagonizes death receptor signalling during bacterial gut infection. *Nature* **501**, 247–251 (2013). doi: 10.1038/nature12524; pmid: 24025841
17. L. Zhang *et al*., Cysteine methylation disrupts ubiquitin-chain sensing in NF-κB activation. *Nature* **481**, 204–208 (2011). doi: 10.1038/nature10690; pmid: 22158122
18. J. Qiu *et al*., Ubiquitination independent of E1 and E2 enzymes by bacterial effectors. *Nature* **533**, 120–124 (2016). doi: 10.1038/nature17657; pmid: 27049943
19. C. Mullineaux-Sanders *et al*., *Citrobacter rodentium*–host–microbiota interactions: Immunity, bioenergetics and metabolism. *Nat. Rev. Microbiol.* **17**, 701–715 (2019). doi: 10.1038/s41579-019-0252-z; pmid: 31541196
20. R. Mundy, T. T. MacDonald, G. Dougan, G. Frankel, S. Wiles, *Citrobacter rodentium* of mice and man. *Cell. Microbiol.* **7**, 1697–1706 (2005). doi: 10.1111/j.1462-5822.2005.00625.x; pmid: 16309456
21. J. W. Collins *et al*., *Citrobacter rodentium*: Infection, inflammation and the microbiota. *Nat. Rev. Microbiol.* **12**, 612–623 (2014). doi: 10.1038/nrmicro3315; pmid: 25088150
22. N. Satoh-Takayama *et al*., Microbial flora drives interleukin 22 production in intestinal NKp46+ cells that provide innate mucosal immune defense. *Immunity* **29**, 958–970 (2008). doi: 10.1016/j.immuni.2008.11.001; pmid: 19084435
23. C. N. Berger *et al*., *Citrobacter rodentium* Subverts ATP Flux and Cholesterol Homeostasis in Intestinal Epithelial Cells *In Vivo*. *Cell Metab.* **26**, 738–752.e6 (2017). doi: 10.1016/j.cmet.2017.09.003; pmid: 28988824
24. E. G. D. Hopkins, T. I. Roumeliotis, C. Mullineaux-Sanders, J. S. Choudhary, G. Frankel, Intestinal Epithelial Cells and the Microbiome Undergo Swift Reprogramming at the Inception of Colonic *Citrobacter rodentium* Infection. *mBio* **10**, e00062-19 (2019). doi: 10.1128/mBio.00062-19; pmid: 30940698
25. T. K. McDaniel, K. G. Jarvis, M. S. Donnenberg, J. B. Kaper, A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1664–1668 (1995). doi: 10.1073/pnas.92.5.1664; pmid: 7878036
26. W. Deng *et al*., Dissecting virulence: Systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3597–3602 (2004). doi: 10.1073/pnas.0400326101; pmid: 14988506
27. J. Garmendia, G. Frankel, V. F. Crepin, Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: Translocation, translocation, translocation. *Infect. Immun.* **73**, 2573–2585 (2005). doi: 10.1128/IAI.73.5.2573-2585.2005; pmid: 15845459
28. S. L. Slater, A. M. Sågfors, D. J. Pollard, D. Ruano-Gallego, G. Frankel, in *Escherichia coli, a Versatile Pathogen*, G. Frankel, E. Z. Ron, Eds. (Springer, 2018), pp. 51–72.
29. N. K. Petty *et al*., The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J. Bacteriol.* **192**, 525–538 (2010). doi: 10.1128/JB.01144-09; pmid: 19897651
30. B. Kenny *et al*., Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**, 511–520 (1997). doi: 10.1016/S0092-8674(00)80437-7; pmid: 9390560
31. X. Xia *et al*., EspF is crucial for *Citrobacter rodentium*-induced tight junction disruption and lethality in immunocompromised animals. *PLOS Pathog.* **15**, e1007898 (2019). doi: 10.1371/journal.ppat.1007898; pmid: 31251784
32. P. Dean, B. Kenny, Intestinal barrier dysfunction by enteropathogenic *Escherichia coli* is mediated by two effector molecules and a bacterial surface protein. *Mol. Microbiol.* **54**, 665–675 (2004). doi: 10.1111/j.1365-2958.2004.04308.x; pmid: 15491358
33. V. F. Crepin *et al*., Dissecting the role of the Tir:Nck and Tir:IRTKS/IRSp53 signalling pathways *in vivo*. *Mol. Microbiol.* **75**, 308–323 (2010). doi: 10.1111/j.1365-2958.2009.06938.x; pmid: 19889090
34. J. Kim *et al*., The bacterial virulence factor NleA inhibits cellular protein secretion by disrupting mammalian COPII function. *Cell Host Microbe* **2**, 160–171 (2007). doi: 10.1016/j.chom.2007.07.010; pmid: 18005731

35. C. N. Berger *et al*., EspZ of enteropathogenic and enterohemorrhagic *Escherichia coli* regulates type III secretion system protein translocation. *mBio* **3**, e00317-12 (2012). doi: 10.1128/mBio.00317-12; pmid: 23033475
36. M. Kelly *et al*., Essential role of the type III secretion system effector NleB in colonization of mice by *Citrobacter rodentium*. *Infect. Immun.* **74**, 2328–2337 (2006). doi: 10.1128/IAI.74.4.2328-2337.2006; pmid: 16552063
37. R. Mundy *et al*., Identification of a novel *Citrobacter rodentium* type III secreted protein, EspI, and roles of this and other secreted proteins in infection. *Infect. Immun.* **72**, 2288–2302 (2004). doi: 10.1128/IAI.72.4.2288-2302.2004; pmid: 15039354
38. E. M. Velazquez *et al*., Endogenous Enterobacteriaceae underlie variation in susceptibility to *Salmonella* infection. *Nat. Microbiol.* **4**, 1057–1064 (2019). doi: 10.1038/s41564-019-0407-8; pmid: 30911125
39. L. Osbelt *et al*., Variations in microbiota composition of laboratory mice influence *Citrobacter rodentium* infection via variable short-chain fatty acid production. *PLOS Pathog.* **16**, e1008448 (2020). doi: 10.1371/journal.ppat.1008448; pmid: 32208465
40. D. B. Schauer, S. Falkow, The eae gene of Citrobacter freundii biotype 4280 is necessary for colonization in transmissible murine colonic hyperplasia. *Infect. Immun.* **61**, 4654–4661 (1993). doi: 10.1128/IAI.61.11.4654-4661.1993; pmid: 8406863
41. E. M. Mallick *et al*., Allele- and tir-independent functions of intimin in diverse animal infection models. *Front. Microbiol.* **3**, 11 (2012). doi: 10.3389/fmicb.2012.00011; pmid: 22347213
42. V. F. Crepin *et al*., Tir Triggers Expression of CXCL1 in Enterocytes and Neutrophil Recruitment during *Citrobacter rodentium* Infection. *Infect. Immun.* **83**, 3342–3354 (2015). doi: 10.1128/IAI.00291-15; pmid: 26077760
43. M. Cepeda-Molero *et al*., Attaching and effacing (A/E) lesion formation by enteropathogenic *E. coli* on human intestinal mucosa is dependent on non-LEE effectors. *PLOS Pathog.* **13**, e1006706 (2017). doi: 10.1371/journal.ppat.1006706; pmid: 29084270
44. C. N. Berger *et al*., The *Citrobacter rodentium* type III secretion system effector EspO affects mucosal damage repair and antimicrobial responses. *PLOS Pathog.* **14**, e1007406 (2018). doi: 10.1371/journal.ppat.1007406; pmid: 30365535
45. I. Zwarts *et al*., Identification of the fructose transporter GLUT5 (*SLC2A5*) as a novel target of nuclear receptor LXR. *Sci. Rep.* **9**, 9299 (2019). doi: 10.1038/s41598-019-45803-x; pmid: 31243309
46. B. Wang, P. Tontonoz, Liver X receptors in lipid signalling and membrane homeostasis. *Nat. Rev. Endocrinol.* **14**, 452–463 (2018). doi: 10.1038/s41574-018-0037-x; pmid: 29904174
47. H. Shimano, R. Sato, SREBP-regulated lipid metabolism: Convergent physiology – divergent pathophysiology. *Nat. Rev. Endocrinol.* **13**, 710–730 (2017). doi: 10.1038/nrendo.2017.91; pmid: 28849786
48. E. M. Zygiel, E. M. Nolan, Transition Metal Sequestration by the Host-Defense Protein Calprotectin. *Annu. Rev. Biochem.* **87**, 621–643 (2018). doi: 10.1146/annurev-biochem-062917-012312; pmid: 29925260
49. H. Ahlfors *et al*., IL-22 fate reporter reveals origin and control of IL-22 production in homeostasis and infection. *J. Immunol.* **193**, 4602–4613 (2014). doi: 10.4049/jimmunol.1401244; pmid: 25261485
50. M. Dougan, G. Dranoff, S. K. Dougan, GM-CSF, IL-3, and IL-5 Family of Cytokines: Regulators of Inflammation. *Immunity* **50**, 796–811 (2019). doi: 10.1016/j.immuni.2019.03.022; pmid: 30995500
51. L. Egea *et al*., GM-CSF produced by nonhematopoietic cells is required for early epithelial cell proliferation and repair of injured colonic mucosa. *J. Immunol.* **190**, 1702–1713 (2013). doi: 10.4049/jimmunol.1202368; pmid: 23325885
52. C. A. Hunter, S. A. Jones, IL-6 as a keystone cytokine in health and disease. *Nat. Immunol.* **16**, 448–457 (2015). doi: 10.1038/ni.3153; pmid: 25898198
53. C. Ma *et al*., *Citrobacter rodentium* infection causes both mitochondrial dysfunction and intestinal epithelial barrier disruption *in vivo*: Role of mitochondrial associated protein (Map). *Cell. Microbiol.* **8**, 1669–1686 (2006). doi: 10.1111/j.1462-5822.2006.00741.x; pmid: 16759225
54. J. Sanchez-Garrido, "External table_ML Model Effector Predictions" (2021); .doi: 10.14469/hpc/7791
55. D. M. Monack, J. Mecsas, D. Bouley, S. Falkow, *Yersinia*-induced apoptosis in vivo aids in the establishment of a systemic infection of mice. *J. Exp. Med.* **188**, 2127–2137 (1998). doi: 10.1084/jem.188.11.2127; pmid: 9841926
56. K. Orth *et al*., Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. *Science* **290**, 1594–1597 (2000). doi: 10.1126/science.290.5496.1594; pmid: 11090361
57. R. Barry *et al*., Faecal neutrophil elastase-antiprotease balance reflects colitis severity. *Mucosal Immunol.* **13**, 322–333 (2020). doi: 10.1038/s41385-019-0235-4; pmid: 31772324
58. N. Katsowich *et al*., Host cell attachment elicits posttranscriptional regulation in infecting enteropathogenic bacteria. *Science* **355**, 735–739 (2017). doi: 10.1126/science.aah4886; pmid: 28209897
59. R. Johnson, E. Mylona, G. Frankel, Typhoidal *Salmonella*: Distinctive virulence factors and pathogenesis. *Cell. Microbiol.* **20**, e12939 (2018). doi: 10.1111/cmi.12939; pmid: 30030897
60. M. Yamamoto *et al*., Regulation of Toll/IL-1-receptor-mediated gene expression by the inducible nuclear protein IκBζ. *Nature* **430**, 218–222 (2004). doi: 10.1038/nature02738; pmid: 15241416
61. N. Kakiuchi *et al*., Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **577**, 260–265 (2020). doi: 10.1038/s41586-019-1856-1; pmid: 31853061
62. K. Nanki *et al*., Somatic inflammatory gene mutations in human ulcerative colitis epithelium. *Nature* **577**, 254–259 (2020). doi: 10.1038/s41586-019-1844-5; pmid: 31853059
63. K. Okamoto *et al*., IκBζ regulates T_H17 development by cooperating with ROR nuclear receptors. *Nature* **464**, 1381–1385 (2010). doi: 10.1038/nature08922; pmid: 20383124
64. T. Ishiguro-Oonuma, K. Ochiai, K. Hashizume, M. Morimatsu, The role of IFN-γ in regulating Nfkbiz expression in epidermal keratinocytes. *Biomed. Res.* **36**, 103–107 (2015). doi: 10.2220/biomedres.36.103; pmid: 25876660
65. J. A. Guttman *et al*., Attaching and effacing pathogen-induced tight junction disruption in vivo. *Cell. Microbiol.* **8**, 634–645 (2006). doi: 10.1111/j.1462-5822.2005.00656.x; pmid: 16548889
66. N. J. Spann, C. K. Glass, Sterols and oxysterols in immune cell function. *Nat. Immunol.* **14**, 893–900 (2013). doi: 10.1038/ni.2681; pmid: 23959186
67. C. T. Chan *et al*., Liver X receptors are required for thymic resilience and T cell output. *J. Exp. Med.* **217**, e20200318 (2020). doi: 10.1084/jem.20200318; pmid: 32716519
68. A. Eetemadi, I. Tagkopoulos, Genetic Neural Networks: An artificial neural network architecture for capturing gene expression relationships. *Bioinformatics* **35**, 2226–2234 (2019). doi: 10.1093/bioinformatics/bty945; pmid: 30452523
69. M. Herrero, V. de Lorenzo, K. N. Timmis, Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J. Bacteriol.* **172**, 6557–6567 (1990). doi: 10.1128/JB.172.11.6557-6567.1990; pmid: 2172216
70. E. Martínez-García *et al*., SEVA 3.0: An update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Res.* **48**, D1164–D1170 (2020). doi: 10.1093/nar/gkz1024; pmid: 31740968
71. D. Burstein *et al*., Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLOS Pathog.* **5**, e1000508 (2009). doi: 10.1371/journal.ppat.1000508; pmid: 19593377
72. Z. Lifshitz *et al*., Identification of novel *Coxiella burnetii* Icm/Dot effectors and genetic analysis of their involvement in modulating a mitogen-activated protein kinase pathway. *Infect. Immun.* **82**, 3740–3752 (2014). doi: 10.1128/IAI.01729-14; pmid: 24958706
73. G. Nissan *et al*., Revealing the inventory of type III effectors in *Pantoea agglomerans* gall-forming pathovars using draft genome sequences and a machine-learning approach. *Mol. Plant Pathol.* **19**, 381–392 (2018). doi: 10.1111/mpp.12528; pmid: 28019708
74. D. Teper *et al*., Identification of novel *Xanthomonas euvesicatoria* type III effector proteins by a machine-learning approach. *Mol. Plant Pathol.* **17**, 398–411 (2016). doi: 10.1111/mpp.12288; pmid: 26104875
75. I. Jiménez-Guerrero *et al*., Show me your secret(ed) weapons: A multifaceted approach reveals a wide arsenal of type III-secreted effectors in the cucurbit pathogenic bacterium *Acidovorax citrulli* and novel effectors in the *Acidovorax* genus. *Mol. Plant Pathol.* **21**, 17–37 (2020). doi: 10.1111/mpp.12877; pmid: 31643123
76. T. Tobe *et al*., An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14941–14946 (2006). doi: 10.1073/pnas.0604891103; pmid: 16990433
77. L. Rohmer, D. S. Guttman, J. L. Dangl, Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. *Genetics* **167**, 1341–1360 (2004). doi: 10.1534/genetics.103.019638; pmid: 15280247
78. T. Kawashima *et al*., Double-stranded RNA of intestinal commensal but not pathogenic bacteria triggers production of protective interferon-β. *Immunity* **38**, 1187–1197 (2013). doi: 10.1016/j.immuni.2013.02.024; pmid: 23791646
79. M. Chaand *et al*., Type Three Secretion System Island-Encoded Proteins Required for Colonization by Non-O1/Non-O139 Serogroup *Vibrio cholerae*. *Infect. Immun.* **83**, 2862–2869 (2015). doi: 10.1128/IAI.03020-14; pmid: 25939511
80. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001). doi: 10.1023/A:1010933404324
81. C. J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **2**, 121–167 (1998). doi: 10.1023/A:1009715923555
82. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
83. F. Pedregosa *et al*., Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
84. F. Lemoine *et al*., NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019). doi: 10.1093/nar/gkz303; pmid: 31028399
85. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019). doi: 10.1093/nar/gkz239; pmid: 30931475
86. F. Sievers, D. G. Higgins, Clustal Omega. *Curr. Protoc. Bioinformatics* **48**, 3.13.11–3.13.16 (2014). doi: 10.1002/0471250953.bi0313s48; pmid: 25501942
87. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009). doi: 10.1093/bioinformatics/btp033; pmid: 19151095
88. C. Kilkenny, W. J. Browne, I. C. Cuthill, M. Emerson, D. G. Altman, Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLOS Biol.* **8**, e1000412 (2010). doi: 10.1371/journal.pbio.1000412; pmid: 20613859
89. J. Cox, M. Mann, 1D and 2D annotation enrichment: A statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13**, S12 (2012). doi: 10.1186/1471-2105-13-S16-S12; pmid: 23176165
90. S. Tyanova *et al*., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016). doi: 10.1038/nmeth.3901; pmid: 27348712
91. D. Zenkova, V. Kamenev, R. Sablina, M. Artyomov, A. Sergushichev, Phantasus: Visual and interactive gene expression analysis (2018); .doi: 10.18129/B9.bioc.phantasus
92. Y. Perez-Riverol *et al*., The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019). doi: 10.1093/nar/gky1106; pmid: 30395289
93. I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, ed. 4, 2016).
94. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015). doi: 10.1038/nature14539; pmid: 26017442
95. D. Barrios, A. Carrascal, D. Manrique, J. Ríos, Optimisation With Real-Coded Genetic Algorithms Based On Mathematical Morphology. *Int. J. Comput. Math.* **80**, 275–293 (2003). doi: 10.1080/0020716022000005546

ACKNOWLEDGMENTS

We thank E. Hopkins for her technical and practical help. **Funding:** The work of T.I.R. and J.S.C. was funded by the CRUK Centre grant with reference no. C309/A25144. N.W. was supported by PhD fellowships from the Manna Center Program for Food Safety and Security at Tel Aviv University and the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University. The work of E.N.-B. was funded by a PhD FPU grant from the Spanish government (FPU2017/04179). A.R.-P. was partially funded by project PID2019-106960GB-I00 (AEI/FEDER, Spain) and by InGEMICS-CM, FSE/FEDER, Comunidad de Madrid Project B2017/BMD-3691. L.Á.F. is supported by grant

BIO2017-89081-R (AEI/MICIU/FEDER, European Union). G.F. is supported by an MRC program grant (MR/R02671/), a Royal Society grant (IC160080), and a Wellcome Investigator Award (107057/Z/15/Z). **Author contributions:** D.R.-G., J.S.-G., Z.K., M.C.-M., T.I.R., I.G.-M., J.N.-B.C., S.L.S., and C.M.-S. performed the experiments and data analysis. Z.K., J.S.-G., T.I.R., and J.S.C. performed the MS and proteomics analysis. T.P. and N.W. performed bioinformatics analysis and ML to find previously uncharacterized effectors. E.N.-B. and A.R.-P. established and applied the ML model for predicting colonization success and the importance of effectors. L.Á.F., T.P., and E.N.-B. provided technical support and edited the paper. G.F., D.R.-G., J.S.-G., A.R.-P., L.Á.F., T.P., and T.I.R. participated in supervision. D.R.-G., J.S.-G., G.F., J.S.C., T.I.R., and Z.K. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD018740 and PXD023145. The algorithm generated in this study to train the model that predicts colonization success has been made available online at https://github.com/liaupm/GraphGANN, and the external table featuring the predictions of the model is available at (*54*). All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6534/eabc9531/suppl/DC1
Figs. S1 to S8
Tables S1 to S12
References (*96*–*130*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

22 May 2020; resubmitted 15 December 2020
Accepted 15 January 2021
10.1126/science.abc9531

Type III secretion system effectors form robust and flexible intracellular virulence networks

David Ruano-Gallego, Julia Sanchez-Garrido, Zuzanna Kozik, Elena Núñez-Berrueco, Massiel Cepeda-Molero, Caroline Mullineaux-Sanders, Jasmine Naemi-Baghshomali Clark, Sabrina L. Slater, Naama Wagner, Izabela Glegola-Madejska, Theodoros I. Roumeliotis, Tal Pupko, Luis Ángel Fernández, Alfonso Rodríguez-Patón, Jyoti S. Choudhary and Gad Frankel

Science **371** (6534), eabc9531.
DOI: 10.1126/science.abc9531

Decrypting pathogen effector networks

Many disease-causing bacteria use a molecular syringe to inject dozens of their proteins, called effectors, into intestinal cells, blocking key immune responses. Ruano-Gallego *et al.* used the mouse pathogen *Citrobacter rodentium* to model effector function in vivo. They found that effectors work together as a network, allowing the microbe great flexibility in maintaining pathogenicity. An artificial intelligence platform correctly predicted colonization outcomes of alternative networks from the in vivo data. However, the host was able to bypass the obstacles erected by different effector networks and activate complementary immune responses that cleared the pathogen and induced protective immunity.

Science, this issue p. eabc9531