



HAL
open science

A gut meta-interactome map reveals modulation of human immunity by microbiome effectors

Veronika Young, Bushra Dohai, Thomas Hitch, Patrick Hyden, Benjamin Weller, Niels van Heusden, Deeya Saha, Jaime Fernandez Macgregor, Sibusiso Maseko, Chung-Wen Lin, et al.

► To cite this version:

Veronika Young, Bushra Dohai, Thomas Hitch, Patrick Hyden, Benjamin Weller, et al.. A gut meta-interactome map reveals modulation of human immunity by microbiome effectors. 2023. hal-04240263

HAL Id: hal-04240263

<https://hal.science/hal-04240263>

Preprint submitted on 13 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 **TITLE**

2 **A gut meta-interactome map reveals modulation of human immunity by microbiome**
3 **effectors**

4

5 **AUTHORS**

6 Veronika Young^{1,15}, Bushra Dohai^{1,15}, Thomas C. A. Hitch², Patrick Hyden³, Benjamin Weller¹,
7 Niels S. van Heusden⁴, Deeya Saha⁵, Jaime Fernandez Macgregor^{5,6}, Sibusiso B. Maseko⁷,
8 Chung-Wen Lin¹, Mégane Boujeant⁵, Sébastien A. Choteau⁵, Franziska Ober⁸, Patrick
9 Schwehn¹, Simin Rothballer¹, Melina Altmann¹, Stefan Altmann¹, Alexandra Strobel¹, Michael
10 Rothballer¹, Marie Tofaute⁸, Matthias Heinig^{9,10}, Thomas Clavel², Jean-Claude Twizere^{7,11,12},
11 Renaud Vincentelli⁶, Marianne Boes⁴, Daniel Krappmann⁸, Claudia Falter¹, Thomas Rattei³,
12 Christine Brun^{5,13}, Andreas Zanzoni⁵, Pascal Falter-Braun^{1,14}✉

13

14 **AFFILIATIONS**

15 ¹*Institute of Network Biology (INET), Molecular Targets and Therapeutics Center (MTTC),*
16 *Helmholtz Munich; Neuherberg, Germany.*

17 ²*Functional Microbiome Research Group, Institute of Medical Microbiology, University Hospital*
18 *of RWTH Aachen; Aachen, Germany.*

19 ³*Department of Microbiology and Ecosystem Science, Research Network: Chemistry meets*
20 *Microbiology, University of Vienna; Vienna, Austria.*

21 ⁴*Center for Translational Immunology, University Medical Center Utrecht; Utrecht, The*
22 *Netherlands.*

23 ⁵*Aix Marseille Univ, INSERM, TAGC, Turing Center for Living Systems; Marseille, France.*

24 ⁶*Aix Marseille Univ, CNRS, AFMB, Turing Center for Living Systems; Marseille, France.*

25 ⁷*Laboratory of Viral Interactomes, GIGA Institute, University of Liège; Liège, Belgium.*

26 ⁸*Research Unit Signaling and Translation, Group Signaling and Immunity, Molecular Targets*
27 *and Therapeutics Center (MTTC), Helmholtz Munich; Neuherberg, Germany.*

28 ⁹*Institute of Computational Biology (ICB), Computational Health Center; Neuherberg,*
29 *Germany.*

30 ¹⁰*Department of Computer Science, TUM School of Computation, Information and Technology,*
31 *Technical University of Munich; Garching, Germany.*

32 ¹¹*TERRA Teaching and Research Centre, University of Liège; Gembloux, Belgium.*

33 ¹²*Laboratory of Algal Synthetic and Systems Biology, Division of Science, New York University*
34 *Abu Dhabi; Abu Dhabi, United Arab Emirates.*

35 ¹³*CNRS; Marseille, France.*

36 ¹⁴*Microbe-Host Interactions, Faculty of Biology, Ludwig-Maximilians-Universität München;*
37 *Planegg-Martinsried, Germany.*

38 ¹⁵*These authors contributed equally*

39 ✉*email: pascal.falter-braun@helmholtz-munich.de*

40

41 **Correspondence and requests for materials** should be addressed to Pascal Falter-Braun.

42

43 **KEYWORDS**

44 microbiome, type-3-secretion system, virulence effectors, complex diseases, network biology,

45 protein-protein interactions, immune signaling, interactome

46 **SUMMARY**

47 The molecular mechanisms by which the gut microbiome influences human health remain
48 largely unknown. Pseudomonadota is the third most abundant phylum in normal gut
49 microbiomes. Several pathogens in this phylum can inject so-called virulence effector proteins
50 into host cells. We report the identification of intact type 3 secretion systems (T3SS) in 5 - 20%
51 of commensal Pseudomonadota in normal human gut microbiomes. To understand their
52 functions, we experimentally generated a high-quality protein-protein meta-interactome map
53 consisting of 1,263 interactions between 289 bacterial effectors and 430 human proteins.
54 Effector targets are enriched for metabolic and immune functions and for genetic variation of
55 microbiome-influenced traits including autoimmune diseases. We demonstrate that effectors
56 modulate NF- κ B signaling, cytokine secretion, and adhesion molecule expression. Finally,
57 effectors are enriched in metagenomes of Crohn's disease, but not ulcerative colitis patients
58 pointing toward complex contributions to the etiology of inflammatory bowel diseases. Our
59 results suggest that effector-host protein interactions are an important regulatory layer by
60 which the microbiome impacts human health.

61 MAIN

62 The host-associated microbiota influences human health in complex host genetics-dependent
63 ways^{1,2}. Especially intestinal microbes positively and negatively affect the risk for several
64 complex diseases ranging from inflammatory bowel disease (IBD)¹ and asthma³ to metabolic⁴
65 and neurodegenerative diseases⁵. Members of the bacterial phylum Pseudomonadota
66 (previously: Proteobacteria⁶) are prevalent in the human gut microbiome and their occurrence
67 is influenced by dietary ingredients such as fat and artificial sweeteners⁷. Unique features of
68 this phylum are the type-3, type-4, and type-6 secretion systems (TxSS) that enable the
69 injection of bacterial proteins directly into the host cytosol. The presence of T3SS has been
70 classically associated with pathogen virulence⁸. In the plant kingdom, however, important
71 mutualistic microbes also communicate with the host via effector proteins to establish
72 cohabitation and elicit host-beneficial effects⁹. We therefore wondered if commensal
73 Pseudomonadota in the healthy human gut microbiome possess host-directed secretion
74 systems.

75 T3SS are common in the normal human gut microbiome

76 Because of the higher quality and completeness of genome assemblies from cultured strains
77 compared to metagenome-assembled genomes (MAGs), we first evaluated Pseudomonadota
78 strains from gut and stool samples that were collected, among others, by the human
79 microbiome project and were available from culture collections. Using EffectiveDB¹⁰, a widely
80 used tool for secretion system identification, we detected complete T3SS in 44 of the 77
81 reference strain genomes (Extended Data Table 1). To expand the scope, we analyzed
82 genomes of 4,752 distinct strains, representing all major phyla from the human gut that had
83 been isolated by the human gastrointestinal bacteria genome collection (HBC)¹¹, and the
84 Unified Human Gastrointestinal Genome (UHGG) collection^{12,13}. Of the 2,272 Gram-negative
85 strains, 478 (21%) had complete T3SS (Fig. 1a); similar proportions have T4SS (527) and
86 T6SS (719), both of which can also deliver effectors into host cells but also have other functions
87 (Extended Data Fig. 1 and Extended Data Table 1)¹⁴. Together 729 of the 2,272 Gram-negative
88 strains, *i.e.*, 34%, have at least one host-directed secretion system. Because culturing can bias
89 the relative proportions of taxa, we sought to confirm the presence of T3SS in commensal
90 microbiota using metagenome datasets. From 16,179 Pseudomonadota MAG bins with high
91 or intermediate genome quality^{15,16}, 770, *i.e.*, 5%, encoded complete T3SS (Fig. 1a and
92 Extended Data Table 1). Notably, we only identified T3SS in Gammaproteobacteria, whereas
93 no secretion systems were found in the Beta- or Epsilonproteobacteria in the datasets, except
94 for a few *Helicobacter* strains. It is unclear if gut commensal strains in these orders lack T3SS,
95 or if the systems differ from those of the better-characterized Gammaproteobacteria and they
96 were missed by the algorithm. Across the analyses, T3SS were identified in strains of multiple

97 genera and were especially common among *Escherichia* (Fig. 1b and Extended Data Table
98 1). Notably, a recent *in vivo* profiling study of human digestive tracts using *in situ* sampling
99 found *Escherichia* as the genus that was most significantly enriched in intestinal over stool
100 samples¹⁷. Of the T3SS-positive (T3SS⁺) species, 24 matched representatives in two cohorts
101 of a dataset provided by the Weizmann Institute of Science (WIS cohorts)¹⁸. 59.4% of
102 individuals in the Israeli cohort and 47.1% in the Dutch cohort had potentially T3SS⁺ species
103 in their gut microbiome, with relative abundances of 0.80% and 0.48%, respectively (Fig. 1c).
104 The most common T3SS⁺ species in both cohorts was *Escherichia coli*, appearing within 54%
105 and 45% of individuals, respectively. Overall, T3SS⁺ strains constitute a substantial proportion
106 of commensal Pseudomonadota and are common in normal human gut microbiomes. We
107 therefore aimed to understand the functions of T3SS-delivered effector proteins of commensal
108 strains.

109 **Commensal effectors are unrelated to known pathogen effectors**

110 To identify gut microbiome-encoded effectors we used a combination of three complementary
111 machine learning models¹⁹⁻²¹ and considered 3,002 effector candidates from the 44 reference
112 strains that were most confidently predicted by all tools (Extended Data Table 2). In addition,
113 we identified 186 putative effectors in the 770 T3SS⁺ MAGs (Extended Data Table 2). As T3SS
114 and substrate effectors are best known for their role in supporting a pathogenic lifestyle, we
115 investigated if the commensal bacterial effectors share sequence similarity with 1,638 known
116 T3SS effectors from pathogens²². Only 17 of 3,002 (0.5%) effectors from strains and 6 of 186
117 (3%) from MAGs, respectively, showed extended high sequence similarity ($\geq 90\%$ sequence
118 similarity across $\geq 90\%$ length) to known pathogen effectors; lowering the thresholds to 50%
119 similarity across 75% length only marginally increased the numbers to 34 (1%) and 7 (4%),
120 respectively (Fig. 1d and Extended Data Table 2). The largest number of commensal effectors
121 with similarity to pathogenic effectors were found in the genomes of *Escherichia albertii* (12
122 effectors with 67% to 98% identity) and *Yersinia enterocolitica* (10 effectors at $> 98\%$ identity).
123 The fact that all such pathogen-similar commensal effectors were found in different species,
124 of which some even belong to a different order than the respective pathogen, suggests that
125 non-pathogenic microbes participate in the horizontal gene transfer of effectors^{23,24}. This is
126 supported by the observation that only a few pathogen-similar effectors were found among the
127 approximately 20 - 80 effectors of each strain. Of the six pathogen-similar effectors found in
128 MAGs, all but one matched the identified family of the pathogen from which they were initially
129 identified (Extended Data Fig. 2 and Extended Data Table 2). Plausibly, these effectors
130 originate from pathogens, or their relatives that were likely present in some samples. Jointly,
131 the data show that effector complements of commensal bacteria are distinct from those of
132 pathogens, thereby suggesting functions outside of the pathogen lifestyle.

133 **A microbiome-host protein-protein meta-interactome map**

134 To investigate the functions of commensal effectors, we cloned effector ORFs for experimental
135 studies from 18 bacterial strains with diverse effector complements (Fig. 1e and Extended Data
136 Fig. 1). We successfully PCR-cloned 786 ORFs for the 1,300 encoded effectors (60.2%) and
137 173 of 186 effector ORFs from MAG bins (meta-effectors) following chemical synthesis (Fig.
138 2a). Thus, 959 sequence-verified full-length effector ORFs were assembled as the human
139 microbiome effector ORFeome (HuMEOme_v1) (Extended Data Table 2). With these, we
140 conducted binary interactome (contactome) network mapping against the human
141 ORFeome9.1 collection encoding 18,000 human gene products using a stringent multi-assay
142 mapping pipeline²⁵. In the main screen by yeast-2-hybrid (Y2H), we identified 1,071
143 interactions constituting the human-microbiome meta-interactome main dataset (HuMMI_{MAIN})
144 (Fig. 2b,c). To assess sampling sensitivity²⁶, i.e., saturation of the screen, we conducted three
145 additional repeats of 290 randomly picked effectors and 1,440 human proteins, which yielded
146 39 verifiable interactions constituting the HuMMI repeat subset (HuMMI_{RPT}). The saturation
147 curve indicates that the single main screen has a sampling sensitivity of ~32% (Fig. 2d). Last,
148 to address how effector sequence similarity affects their interaction profiles we conducted a
149 homolog screen. Effectors were grouped if they shared $\geq 30\%$ sequence identity (Extended
150 Data Table 2) and all effectors of one group were experimentally tested against the union of
151 their human interactors. The resulting dataset (HuMMI_{HOM}) contains 398 verified interactions,
152 of which 179 were not found in the other screens. Altogether, HuMMI contains 1,263 unique
153 verified interactions between 289 effectors and 430 human proteins (Fig. 2b,c and Extended
154 Data Table 3).

155 To assess data quality, we assembled a positive control set of 67 well-documented manually
156 curated binary interactions of bacterial (pathogen-) effectors with human proteins from the
157 literature (bacterial human literature binary multiple – bhLit_BM-v1, Extended Data Table 3)
158 and a corresponding negative control set of random bacterial and human protein pairs
159 (bacterial host random reference set - bhRRS-v1). Benchmarking our Y2H assay in a single
160 orientation with these and with the established human positive reference set (hsPRS-v2) and
161 hsRRS-v2 indicated an assay sensitivity of ~13% and 17.5%, respectively, which is consistent
162 with previous observations^{27,28} (Fig. 2e and Extended Data Table 3). No negative control pair
163 in either reference set scored positive, demonstrating the reliability of our system. In addition,
164 we assessed the biophysical quality of HuMMI using the yeast nanoluciferase-2-hybrid assay
165 (yN2H), which we benchmarked using the same four reference sets²⁵. Notably, the retest rates
166 of all sets involving bacterial proteins were lower than those of the human hsPRS-v2 and
167 hsRRS-v2 across most of the scoring spectrum (Extended Data Fig. 2). Partly, this could be
168 due to the nature of hsPRS-v2 pairs, which consist of very well-documented interaction pairs,

169 which may have been selected for good detectability. In addition, the fact that the RRS sets
170 exhibit the same overall trend indicates that interactions with prokaryotic proteins are more
171 challenging to reproduce in this eukaryotic assay system, which reinforces the necessity for
172 bacterial protein-specific reference sets (Fig. 2f, Extended Data Fig. 2, and Extended Data
173 Table 3). At thresholds where the control sets were well separated, the retest rate of 173
174 randomly selected HuMMI interactions was statistically indistinguishable from the positive
175 control sets, and significantly different from those of the negative controls (Fig. 2f, Extended
176 Data Fig. 2, and Extended Data Table 3), indicating that the biophysical quality of our dataset
177 is comparable to those of well-documented interactions in the curated literature.

178 The degree distribution of $\text{HuMMI}_{\text{MAIN}}$ shows that numerous human proteins are targeted by
179 multiple effectors (Fig. 2g and Extended Data Table 3), often from different species. Indeed,
180 sampling analysis demonstrates that commensal effectors significantly converge on fewer host
181 proteins than expected from a random process (Fig. 2h), thus suggesting selection for
182 interactions with these targets. We had previously observed convergence of effectors from
183 phylogenetically diverse pathogenic microbes on common proteins of their plant host^{29,30}. In
184 that system, we demonstrated with infection assays on genetic null mutant plant lines that the
185 extent of convergence correlates with the importance of the respective host proteins for the
186 outcome of the microbe-host interaction²⁹. We therefore identified the human host proteins
187 onto which commensal effectors converge. To this end, we sampled random effector targets
188 for each strain and analyzed the distribution of repeatedly targeted proteins (Fig. 2i). While
189 host proteins interacting with effectors from two strains are expected at high frequency by
190 chance, targeting by four bacterial strains is unlikely to emerge by chance (Fig. 2i and
191 Extended Data Table 3). Thus, the 60 human proteins targeted by effectors from four or even
192 more commensal strains are subject to effector convergence and may be of general
193 importance for human microbe-host interactions. Together with our recently published plant-
194 symbiont interaction data³¹, these data suggest that convergence has evolved as a universal
195 feature of effector-host interactions independent of the microbial lifestyle and kingdom of the
196 host organism.

197 **Sequence features mediating effector-host interactions**

198 The function of unknown proteins can often be inferred from better-studied orthologues, but
199 convergence could also result from high sequence similarity among effectors. We therefore
200 compared sequence- to interaction-similarity as a proxy for their function in host cells (Fig. 3a).
201 Within the systematically retested $\text{HuMMI}_{\text{HOM}}$ clusters, both are poorly correlated and
202 sequence similarity merely defines the upper limit for interaction similarity but does not imply
203 it. This is illustrated by cluster 3, in which all seven effectors share over 90% mutual sequence

204 similarity while their pairwise interaction profile similarities range from identical to
205 complementary (Fig. 3b and Extended Data Table 3).

206 Using HuMMI_{MAIN} we also investigated if effectors without substantial sequence similarity share
207 interaction similarity, which might indicate shared functions. In fact, clustering effectors by their
208 pairwise interaction similarity identified substantial overlap outside the homology clusters
209 (Extended Data Fig. 3), indicating that dissimilar effectors may have similar functions in the
210 host. Both analyses indicate that effector function as measured by protein-interaction profiles
211 is largely independent of overall sequence similarity.

212 Looking for structural correlates for interaction specificity, we wondered whether domain-
213 domain or domain-short linear motif (SLiM) interfaces mediating the interactions can be
214 identified (Fig. 3c). Using experimentally identified interaction templates³², a putative interface
215 was found for 52 interactions in the HuMMI_{MAIN} screen (Extended Data Table 4). Of these, 43
216 interactions matched motif-domain templates passing one (Fig. 3d), and 22 passing two
217 stringency criteria (Extended Data Fig. 3). Among the former, 23 interactions involve PDZ
218 domains in the human protein, which recognize PDZ-binding motifs (PBM) in the C-terminus
219 of interacting bacterial proteins. PDZ domain-containing proteins commonly mediate cell-cell
220 adhesion, cellular protein trafficking, tissue integrity, as well as neuronal and immune
221 signaling³³. To experimentally validate these interfaces, individual and tandem PDZ domains
222 from 13 human proteins and C-terminal peptides from 16 interacting bacterial effectors were
223 tested via Holdup, a quantitative chromatographic *in vitro* interaction assay^{34,35}. For 16 of 23
224 Y2H pairs (70%) at least one PDZ-peptide interaction was identified, all with affinities between
225 1 and 200 μ M (Fig. 3e and Extended Data Table 4). In three instances two PDZ domains
226 arranged in tandem were required to detect the interaction by Holdup, indicating that some
227 Y2H pairs might have been missed because not all PDZ combinations of the proteins were
228 tested. For human proteins with multiple PDZ domains, often different domains were the target
229 for different effectors demonstrating both specificity and functional specialization of the
230 effectors (Fig. 3e).

231 Because of their functioning in immune signaling and cell shape, PDZ domains are frequently
232 targeted by viruses³⁶. This opens the possibility that bacterial effectors and viral proteins
233 compete for PDZ-binding and thus mutually influence their respective impact on the host. To
234 gather support for this possibility, we identified viruses that can cause infections in the digestive
235 tract, namely SARS-CoV-2³⁷, HPV16 and 18, which have a high prevalence in human guts and
236 have been linked to colorectal cancer³⁸, and norovirus, a globally common cause of
237 gastroenteritis and diarrhea³⁹. We selected two hitherto unpublished interactions of Norovirus
238 VP2 C-terminal peptide with DLG1 (domain 2) and MAGI1 (domain 4), and previously
239 observed interactions between the C-terminal peptides of SARS-CoV-2 E with SHANK3, and

240 of HPV16 and 18 E6 with the PDZ domains of PICK1 and MAGI4 (domain 1), respectively³⁴.
241 Indeed, in fluorescent polarization assays the viral PBM peptides competed with those of the
242 effectors Vfu_12, met_32, met_31, and met_46 (Fig. 3f and Extended Data Fig. 3). Similarly,
243 the functionally well-characterized interaction of the C-terminus of HTLV1 Tax1 with DLG1⁴⁰
244 was competed off by the met_32 PBM peptide. Thus, viral and bacterial proteins may compete
245 in the intracellular environment for binding partners and hence for influence on human cell
246 function. Such competition could contribute to the previously observed mutual influence of
247 microbiome and viral infection on each other⁴¹.

248 Thus, while the overall sequence similarity of effectors does not correlate with their host-protein
249 interaction profiles, several interfaces mediating the interactions can be identified. How these
250 interactions compete with human and viral proteins to modulate the host network is an
251 important question for future studies.

252 **Effector-targeted functions and disease modules**

253 To explore the potential roles of commensal effectors in the host we analyzed the functions of
254 the targeted human proteins through gene ontology (GO) enrichment analysis (Fig. 4a,
255 Extended Data Fig. 4, and Extended Data Table 5). Redundant parent-child GO-term pairs
256 were grouped and are displayed by a representative term. Intriguingly, “response to muramyl-
257 dipeptide (MDP)”, a bacterial cell wall-derived peptide that can be perceived by human cells,
258 was among the most enriched functions, thus not only supporting the relevance of our
259 interactions but indicating that effectors modulate cellular responses to their detection.
260 Moreover, a key component of the MDP signaling pathway is NOD2, which is encoded by a
261 major susceptibility gene for Crohn’s disease (CD)⁴², an autoimmune disease with a strong
262 etiological microbiome contribution⁴³. In addition, several central immune signaling pathways
263 are enriched among the targets, namely the NF- κ B and the stress-activated protein kinase and
264 Jun-N-terminal kinase (SAPK/JNK) pathways, supporting the notion that modulation of
265 immune signaling is an important function of commensal effectors. Remarkably, five of the
266 significantly targeted convergence-proteins belong to the NF- κ B module (Extended Data Fig.
267 4), one of the evolutionarily oldest immune signaling pathways in animals that is already
268 present in sponges⁴⁴. This may reflect the long co-evolution between microbial effectors and
269 this ancient immune coordinator. Relating to human disease, anti-TNF biologicals, which
270 dampen NF- κ B-driven immunity, are an important therapeutic for diverse autoimmune
271 diseases including CD, psoriasis, and rheumatoid arthritis. Another highly enriched group of
272 five terms relates to collagen production, which suggests that effectors may modulate the
273 extracellular environment that hosts the microbes. Inflammation-independent fibrotic collagen
274 production is an important clinical feature of CD, and the gut microbiota has been found to be
275 a main driver⁴⁵. As several metabolism-related terms were identified, we also tested directly

276 whether enzymes in the Recon3D⁴⁶ model of human metabolism were targeted. Indeed, we
277 detected a significant enrichment of metabolic enzymes ($P = 0.0001$, Fisher's exact test) and
278 nominally significant targeting of bile acid and glycerophospholipid metabolism, and fatty acid
279 oxidation (Extended Data Table 5). Overall, however, despite the strong overall signal and
280 general targeting of fatty acid metabolism, no individual metabolic subsystem stood out as
281 being targeted by effectors from more than two strains or having more than two targeted
282 proteins.

283 From a network perspective, proteins encoded by disease-genes (disease proteins) constitute
284 nodes and form disease modules⁴⁷, whose functional perturbation promotes pathogenesis.
285 Importantly, viruses can contribute to non-infectious disease etiology by binding to and
286 similarly perturbing these disease proteins and modules⁴⁸. Therefore, we wondered if bacterial
287 effectors also target such network elements and may thereby influence human traits. We
288 started with "causal genes/proteins" identified from genome-wide-association studies (GWAS)
289 by the Open Targets initiative⁴⁹, and merged gene sets for traits identified as identical by their
290 experimental factor ontology (EFO) terms (Extended Data Table 5). We first investigated direct
291 effector targets. The strong enrichment of the "immunoglobulin isotype switching" trait among
292 these is intriguing as the evolutionarily older IgA antibodies are emerging as having an
293 important role in shaping the gut microbiome^{50,51}. Effector-targeted proteins are further
294 associated with diverse cancers and with diseases that have a strong immunological
295 component, including asthma, psoriasis, allergies, and systemic lupus erythematosus (Fig. 4b,
296 cutoff nominal $P = 0.05$, Fisher's exact test, Extended Data Table 5). While none of the
297 identified diseases is currently known as an ailment of the gut it has emerged that the gut
298 microbiome shapes immune homeostasis and contributes to lung and skin diseases like
299 asthma⁵² and psoriasis⁵³. In addition, some of the disease-associated genes encode
300 convergence proteins for effectors from multiple bacterial species (Fig. 2g). As such, it is
301 plausible that proteins like REL or TCF4 are similarly targeted by effectors from
302 Pseudomonadota in skin or lung microbiome communities and contribute to the identified
303 diseases. Moreover, 26% of the effectors in HuMMI are also detectable in skin microbiome
304 samples (Extended Data Table 5), indicating that commensal effectors are shared between
305 different ecological niches.

306 A partly complementary explanation emerges from our previous studies of human and plant
307 pathogen-host systems. In these evolutionary distant systems, we showed that genetic
308 variation affecting the severity of infection does not reside in genes encoding direct targets but
309 in interacting, i.e., neighboring proteins in the host network^{25,29}. We, therefore, explored the
310 network neighborhood of all effector-targets using short random walks in the human reference
311 interactome (HuRI)⁵⁴. We identified proteins that were significantly more often visited in HuRI

312 compared to degree-preserved randomly rewired networks, which we considered the
313 'neighborhood'. For each effector-targeted neighborhood, we assessed the enrichment of gene
314 products associated with diverse human traits using Open Targets causal genes. Nominally
315 significant associations were aggregated on a strain level and summarized for disease groups
316 (Fig. 4c and Extended Data Table 5). Intriguingly, most disease groups for which susceptibility-
317 gene products are enriched in the target neighborhoods represent traits that have been linked
318 to the gut microbiome⁵⁵. Apart from immunological traits, these include cardiovascular,
319 metabolic, and neurological traits as well as multiple cancers, including colorectal cancer.
320 Among the target neighborhoods for immunological diseases, we identified associations to CD
321 (nominal $P = 8.5 * 10^{-5}$, Fisher's exact test) and inflammatory bowel disease (nominal $P =$
322 0.0008 , Fisher's exact test) but not to ulcerative colitis (UC) (Fig. 4d and Extended Data Table
323 5). Neighborhoods harboring genetic susceptibility associated with psoriatic arthritis, asthma,
324 and allergies were also significantly targeted, which recapitulates the observations for direct
325 targets. Considering the importance of the microbiome for human metabolic disorders⁵⁵ it is
326 noteworthy that network modules important for HDL and LDL cholesterol levels (nominal $P =$
327 0.006 and $P = 0.008$, respectively, Fisher's exact test), and several diabetes traits were
328 significantly targeted albeit less recurrently than inflammatory diseases and cancers (Extended
329 Data Table 5). Together, these results suggest that commensal effectors modulate their host's
330 immune system and local metabolic and structural microenvironment. As genetic variation
331 affecting the targeted proteins and their network neighborhood is linked to several human
332 diseases, functional modulation of the same network neighborhoods by commensal effectors
333 may contribute to disease etiology. The fact that the risk for several of the identified diseases
334 is known to be modulated by the microbiome strengthens this hypothesis. We therefore
335 investigated if commensal effectors, indeed, perturb some of the identified pathways and
336 functions.

337 **Effector function in human cells and disease**

338 The NF- κ B signaling module is enriched among the convergence proteins and all targets of
339 commensal effectors (Fig. 4a and Extended Data Fig. 4). Because of its important role in many
340 diseases, we chose a cell-based dual-luciferase assay²⁵ to test whether commensal effectors
341 modulate NF- κ B pathway activity in human cells. Indeed, five of 26 commensal effectors
342 caused a significant increase in NF- κ B pathway activity in the absence of exogenous
343 stimulation suggesting pathway activation (Fig. 5a and Extended Data Table 6). Conversely,
344 three effectors significantly reduced relative transcriptional NF- κ B activity even in the presence
345 of strong TNF stimulation (Fig. 5b, Extended Data Fig. 5, and Extended Data Table 6). Since
346 some bacterial effectors also modulate NF- κ B-independent induction of the thymidine kinase
347 control promoter, we assessed the impact of selected effectors on endogenous expression of

348 NF- κ B controlled human adhesion factor ICAM1 and cytokine secretion. We focused these
349 experiments on two NF- κ B activating (Kpn_9, met_7) and two NF- κ B inhibiting (Pst_11,
350 Cyo_12) bacterial effectors. ICAM1/CD54 is a glycoprotein that mediates intercellular epithelial
351 adhesion and interactions with immune cells, specifically neutrophils. Epidemiologically,
352 ICAM1 has been linked to CD such that increased ICAM1 expression is associated with higher
353 disease risk⁵⁶ likely by facilitating recruitment and retention of inflammatory immune cells^{57,58}.
354 Interference with ICAM1-mediated neutrophil trafficking is currently being tested as a
355 therapeutic approach to treat CD⁵⁹. In colon carcinoma Caco-2 cells, expression of met_7
356 caused a significant increase of ICAM1 expression ($P = 0.05$, one-way ANOVA with Dunnett's
357 multiple hypothesis correction, Extended Data Table 6) following stimulation with a pro-
358 inflammatory cocktail. Expression of the inhibitory effectors Pst_11 and Cyo_12 did not
359 significantly alter the induction of ICAM1 cell surface expression (Fig. 5c). We also investigated
360 the effect of met_7 and Cyo_12 on cytokine secretion in unstimulated Caco-2 cells or following
361 pro-inflammatory stimulation. In basal conditions, Cyo_12 reduced the secretion of several
362 cytokines especially IL6 and IL8, whereas met_7 caused an increase in IL8 secretion in these
363 conditions (Fig. 5d and Extended Data Table 6). Following proinflammatory stimulation,
364 expression of Cyo_12 further reduced cytokine secretion. This effect was most pronounced for
365 IL8, but also significant for IL6 and the pro-inflammatory IL1beta, IL18, and IL23. These
366 cytokines are noteworthy as they are linked to IBD pathogenesis. IL23R has been associated
367 to CD, and IL6 and IL23 stimulate the differentiation of Th17 cells, which have emerged as key
368 players in CD^{60,61}. IL8 is overexpressed in colonic tissue of IBD patients and has been
369 suggested as a chemoattractant triggering neutrophil invasion^{62,63}. In contrast, no significant
370 impact of met_7 on cytokine secretion was detectable in the context of stimulation (Fig. 5e and
371 Extended Data Fig. 5). Thus, commensal effectors can both stimulate and dampen intracellular
372 immune signaling and this modulation can impact immune and tissue homeostasis via cell-cell
373 adhesion and cytokine secretion.

374 As we identified both genetic and functional links between commensal effectors and IBD-
375 related processes, we sought clinical evidence for a potential role of effectors in these
376 diseases. We hypothesized that a potential role of effectors in IBD etiology may be reflected
377 in altered effector prevalence in the microbiota of patients versus healthy controls. Analyzing
378 a large dataset with > 800 IBD patient-derived and > 300 healthy control-derived
379 metagenomes⁶⁴ we found 64 effectors that were significantly more prevalent in the
380 metagenomes of CD patients compared to healthy controls (Fig. 5f and Extended Data Table
381 6). In metagenomes of UC patients only three effectors had a significantly different prevalence,
382 and, intriguingly, these were less common compared to healthy controls (Extended Data Table
383 6). This trend was recapitulated when the prevalence distributions of all detected effectors

384 were analyzed. Whereas CD patients had a significantly higher load of effectors, the overall
385 effector prevalence was lower in UC patients compared to healthy subjects (Fig. 5g and
386 Extended Data Table 6). These opposing findings were unexpected as an increased
387 abundance of Pseudomonadota has been reported both for CD and UC patients⁶⁵. At the same
388 time, many clinical features such as affected tissues and response to anti-TNF therapy differ
389 between these two forms of IBD, rendering it plausible that effectors contribute differently to
390 their etiology. Whether commensal effectors indeed causally contribute to disease etiology or
391 acute flairs is an important question with potential therapeutic implications.

392 **Discussion**

393 The presence of T3SS in human commensal microbes has been noticed previously and was
394 speculated to mediate crosstalk between the intestinal microbiota and the human host^{66,67}.
395 Here, we provide evidence that, analogous to the plant kingdom^{31,68}, also in the human gut
396 T3SS and effectors function in commensal microbe-host interactions and modulate immune
397 signaling. Thus, effector secretion appears to be used universally by Pseudomonadota to
398 mediate interactions with multicellular eukaryotes independently of the lifestyle of the microbe.

399 Since, as we show, commensal effectors modulate immune signaling we hypothesized that
400 this may affect the manifestation of human diseases, especially those involving the immune
401 system. The influence of the microbiome on IBD etiology is well documented¹. Therefore, it is
402 noteworthy that IBD, especially CD, emerged in several of our analyses. Effectors target the
403 “response to the muramyl-dipeptide” pathway which includes NOD2, a major CD-associated
404 gene product⁶⁹. Further, effectors target and regulate the NF- κ B pathway, which is strongly
405 activated by TNF, a key therapeutic target in CD⁷⁰. Likewise, ICAM1 is a susceptibility gene
406 for CD whereby high expression increases disease risk⁵⁶. Secretion of IL6, IL8, and IL23 is
407 significantly altered by effectors, and all have previously been linked to CD^{61,63}. Thus,
408 commensal effectors regulate several IBD-relevant pathways and can thus influence the
409 establishment or maintenance of feedback loops during disease development⁷¹. This
410 conclusion is strengthened by the observation that effectors are enriched in metagenomes of
411 a CD patient cohort. Thus, multiple lines of evidence suggest that by modulating immune
412 signaling, commensal effectors contribute to the etiology of CD.

413 Likely other microbial habitats of the human body, such as skin or lung, also host T3SS+
414 strains, and we identified effectors in a skin metagenome. It will be important to investigate this
415 in the future to understand if those effectors have similar targets and effects on local cells.
416 ICAM1, e.g., is the entry receptor for rhinovirus A⁷², and an increased expression due to
417 microbial effectors could increase the risk for infections and thus to develop asthma^{73,74}.

418 The broader question of how effectors influence the pathogenesis of IBD and other diseases
419 will be important to address in further detailed studies. Our molecular data show that different
420 effectors can have opposing impacts on immune pathways, analogous to genetic variants.
421 Thus, host genetics and effectors jointly impact on the molecular networks, and pathogenic
422 developments emerge from the interplay of protective and disease enhancing factors. For CD
423 specifically, however, our analyses suggest that effectors promote disease development.

424 In summary, we demonstrate that bacterial effector proteins constitute a hitherto unrecognized
425 regulatory layer by which the commensal microbiota communicates with host cells and
426 modulates human physiology. We anticipate that our findings and resources will open new
427 research directions towards understanding the host-genetics dependent mechanisms by which
428 the microbiome influences human health and exploring the potential of effectors for therapy
429 and prevention.

430 **METHODS**

431 **Identification of T3SS+ strains in culture collections and MAGs**

432 To collect reference genomes for strains available from culture collections, three large culture
433 collections were queried for all Pseudomonadota strains: DSMZ via BacDive⁷⁵, ATCC
434 (atcc.org) and BEI (beiresources.org). The strain numbers were looked up in GenBank
435 (Release 229) from which 77 strains could be identified as perfect match.

436 MAGs that were at least 50% complete and less than 5% contaminated (as estimated by
437 CheckM⁷⁶ from two different meta-studies) were selected. 92,143 MAGs of Almeida et al.¹⁵ and
438 9,367 Pseudomonadota MAGs from Pasolli et al.¹⁶ were used as input for T3SS prediction
439 scaled via massive parallel computing. The computational predictions presented have been
440 achieved in part using the Vienna Scientific Cluster (VSC). The prediction performance of
441 EffectiveDB¹⁰ on incomplete and contaminated MAGs was assessed by 5-fold cross-validation
442 with 5 repeats using 0 - 100% completeness and 0 - 50% contamination in 5% steps of
443 simulated incompleteness/contamination, randomly sampling genes from test-set. In addition,
444 T3SS were predicted for 4,753 strains isolated by the human gastrointestinal bacteria genome
445 collection (HBC)¹¹, and the unified gastrointestinal genome (UHGG) collection^{12,13}. A
446 performance-improved re-implementation of the EffectiveDB classifier
447 (<https://github.com/univieCUBE/phenotrex>, trained on EggNOG 4 annotations⁷⁷) was used to
448 predict functional T3SS present in MAGs and genomes of isolated strains. Threshold for
449 positive prediction was defined as > 0.7 .

450 Protein sequences were predicted from 44 T3SS-positive reference strains and MAGs using
451 prodigal v2.6.3⁷⁶. Of 770 MAGs a total of 474,871 representative protein sequences were
452 identified using CD-HIT⁷⁸ (v4.8.1, parameters: ``-c 1.0``). The identical procedure was performed
453 for 44 genomes from culture collections resulting in 161,115 proteins. Machine-learning based
454 tools were used to predict T3SS signals (EffectiveT3 v.2.0.1 and DeepT3 2.0¹⁹) or effector
455 homology using pEffect²¹ to extract potential effector proteins. The results of all three tools
456 were combined using a 0 - 2 scoring scheme: 2 for perfect score (pEffect > 90 , EffectiveT3 $>$
457 0.9999 , DeepT3: both classifiers positive prediction), 1 for positive prediction as defined by
458 default settings (pEffect > 50 , EffectiveT3 > 0.95 , DeepT3: one classifier) and 0 for negative
459 prediction. Sequences with a sum score above 4 were regarded as potential effectors. Further,
460 all sequences without start/stop-codon or trans-membrane region containing proteins (> 0
461 regions; predicted with TMHMM version 2.0) were excluded. Proteins were clustered using
462 90% sequence identity threshold (CD-HIT parameters ``-c 0.9 -s 0.9``) to reduce redundancy.
463 Effector-clusters with great diversity regarding T3SE-prediction scores were removed from the
464 final set. Full data in Extended Data Table 1.

465 **Identification of effector similarities and homology groups**

466 Based on a mutual sequence identity of $\geq 30\%$ over 90% of the common sequence length
467 effectors were considered 'homologous' and included in the HuMMI_{HOM} experiment to
468 investigate the impact of sequence similarity on interaction similarity. Protein sequences were
469 analyzed by global alignment using Needleman Wunsch algorithm implemented in the emboss
470 package (Extended Data Table 2).

471 **Commensal vs pathogen effector similarity**

472 We gathered the sequences of 1,195 known pathogenic T3 effectors from the BastionHub
473 database⁷⁹ (August 29th, 2022). We assessed the similarity between commensal and
474 pathogenic effector sequences using BLAST (stand-alone, version 2.10⁸⁰). For each
475 commensal effector, the pathogen effector with the highest sequence similarity was considered
476 as best match. Subsequently, we computed the alignment coverage over the pathogenic
477 effector sequence. Full data in Extended Data Table 2.

478 **Cohort analyses**

479 Genomes of bacterial isolates from the human gut were gathered from multiple published
480 datasets¹¹⁻¹³. The presence of T3SS was predicted for each of these genomes as described
481 above. GTDB-Tk (v2.1)⁸¹ was used to assign the taxonomy to each of the genomes, and the
482 concatenated bac120 marker proteins from this were used to generate a phylogenomic tree of
483 the isolates, visualized in iTOL⁸². FastANI was used to match the T3SS positive genomes to
484 the WIS representative genomes of the human gut¹⁸ based on ANI values $> 95\%$ ⁸³. The relative
485 abundance of the 10 matching representative genomes was then identified across 3,096
486 Israeli, and 1,528 Dutch patients¹⁸.

487 **Effector cloning**

488 Bacterial strains from the ATCC collection were ordered from LGS Standard Standard (Wesel,
489 Germany) or ATCC in the US (Manassas, Virginia). Bacterial strains from the DSMZ collection
490 were obtained from the Leibniz-Institut DSMZ (Braunschweig, Germany) and strains from the
491 BEI collection were ordered at BEI resources (Manassas, Virginia, USA) (Extended Data Table
492 2). Effectors identified from MAGs and effectors for the PRS were ordered at Twist Bioscience
493 (San Francisco, CA, 660 USA). If no genomic DNA could be obtained strains were cultured
494 according to the manufacturer's protocol and DNA was extracted using the NucleosSpin
495 Plasmid (NoLid) Mini kit (Macherey-Nagel cat. No. 740499) with vortexing after addition of
496 BufferA2 and BufferA3. A nested PCR was performed to add Sfi sites, the DNA was purified
497 using magnetic beads (magtivio cat. no. MDKT00010075), followed by an Sfi digestion and
498 another clean-up with magnetic beads. Digested PCR products were cloned into pENTR223.1

499 using T4 DNA Ligase (ThermoFisher ca. no. EL0011). Plasmids were propagated in DH5 α *E.*
500 *coli* and the plasmid DNA was extracted using the pipetting Bio Robot Universal System
501 (Qiagen cat. no. 9001094) and the QIAprep 96 plus BioRobot kit (Qiagen cat. no. 962241).
502 ORFs were verified by Sanger Sequencing. Effectors were cloned into the Y2H destination
503 plasmid pDEST-DB (pPC97, Cen origin), the pDEST-N2H-N1 and -N2, or the mammalian
504 expression vector pMH-FLAG-HA by an LR reaction of the Gateway System. After propagation
505 in DH5 α *E. coli* and DNA extraction plasmids were transformed into *S. cerevisiae* Y8930
506 (MAT α mating type) as DB-X ORFs as described⁸⁴.

507 **Meta-interactome mapping**

508 A state-of-the-art high-quality Y2H screening pipeline was followed as previously
509 described^{25,85}. DB-X ORFs were tested for autoactivation by mating against AD-empty
510 plasmids in Y8800 (MAT α). 45 ORFs of the strains and 14 meta effectors tested positive and
511 were excluded from subsequent steps. The remaining 900 ORFs were individually mated
512 against pools of ~188 AD-Y human ORFs from the human ORFeome collection v9.1 including
513 17,472 ORFs⁸⁶. During primary screening, haploid AD-Y and DB-X yeast cultures were spotted
514 on top of each other and grown on yeast extract peptone dextrose (YEED) agar (1%) plates.
515 After incubation for 24 h, the clones were replica plated onto selective synthetic complete
516 media lacking leucine, tryptophan and histidine (SC-Leu-Trp-His) + 1 mM 3-AT (3-amino-1,2,4-
517 triazole) (3-AT plates) and replica cleaned after 24 h. 48 h later, three colonies were picked
518 per spot and grown for 72h in SC-Leu-Trp liquid medium. For the secondary phenotyping,
519 yeasts were spotted on SC-Leu-Trp plates and after incubation for 48 h replica plated and
520 cleaned on 3-AT-plates and SC-Leu-His + 1 mM 3-AT + 1 mg per litre cycloheximide plates to
521 identify spontaneous DB-X autoactivators. Clones growing on 3-AT plates, but not on
522 cycloheximide plates were picked into yeast lysis and processed to generate a library for pair
523 identification by Next Generation Sequencing using a modified KiloSeq procedure as
524 previously described²⁵. Identified DB-X and AD-Y pairs were mated individually during the
525 fourfold verification, replica plated and cleaned after 24 hours and picked after another 48 h
526 incubation. Growth scoring was performed using a custom dilated convolutional neural network
527 as described²⁵. Pairs scoring positive at least three out of the four repeats qualified as bona
528 fide Y2H interactors. The AD-Y and DB-X constructs were identified once more by NGS. All
529 interaction data are in Extended Data Table 3.

530 **Assembling reference sets**

531 To identify additional reliably documented interactions between bacterial effectors and human
532 proteins for the positive control set (bhLit_BM-v1), we queried the IMEx consortium protein
533 interaction databases⁸⁷ through the PSICQUIC webservice⁸⁸ (May 10th, 2021) using the T3

534 effectors UniprotKB accession numbers and fetched all the PubMed identifiers of the articles
535 describing additional interactions. In total, we gathered 67 interactions between 29 T3 effectors
536 and 64 human proteins, described in 13 distinct publications that underwent the manual
537 curation step for inclusion in the PRS (Extended Data Table 3).

538 **Y2H assay sensitivity**

539 Effector ORFs from bhLit_BM-v1 and bhRRS-v1 (Extended Data Table 3) were transferred
540 into pDEST-DB (DB-X) and transformed into *Saccharomyces cerevisiae* Y8930 (MAT α). Yeast
541 strains containing the corresponding AD-Y human ORF were picked from hORFeome9.1⁸⁶ and
542 ORF identity verified by end-read Sanger sequencing of PCR products. Yeast strains harboring
543 plasmids containing ORFs from hsPRS-v2/hsRRS-v2⁸⁹ were provided by the Center for
544 Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA. DB-X and AD-Y were
545 mated fourfold with each other, as well as against yeast strains containing the corresponding
546 DB-empty or AD-empty plasmid. Growth scoring was performed as described above for the
547 fourfold verification. Pairs scoring positive at least three out of the four repeats qualified as
548 bona fide Y2H interactors.

549 **Interactome validation by yN2H**

550 200 interactions were randomly picked from HuMMI and all ORFs from the indicated datasets
551 (Extended Data Table 3) were transferred by Gateway LR reactions into pDEST-N2H-N1 and
552 pDEST-N2H-N2 plasmids containing a *LEU2* or *TRP1* auxotrophy marker, respectively⁸⁹.
553 Successful cloning was monitored by PCR-mediated evaluation of insert size, and positive
554 clones transformed into haploid *Saccharomyces cerevisiae* Y8930 (MAT α) and Y8800 (MAT α)
555 strains, respectively. Protein pairs from all datasets were randomly distributed across matching
556 96-well plates.

557 5 μ L of each haploid culture of opposite mating type grown to saturation was mated in 160 μ L
558 YEPD medium and incubated overnight. Additionally, each position was mated with yeast
559 stains containing empty N1 or N2 plasmids, to measure background. 10 μ L mated culture was
560 inoculated in 160 μ L SC-Leu-Trp and grown overnight. 50 μ L of this overnight culture was
561 reinoculated in 1.2 ml SC-Leu-Trp and incubated for 24 h at 1000 rpm. Cells were harvested
562 15 min at 3000 rpm, the supernatant discarded, and each cell pellet was fully resuspended in
563 100 μ l NanoLuc Assay solution (Promega corp. Madison, WI, USA, cat# 1120). Homogenized
564 solutions were transferred to white flat-bottom 96-well plates (Greiner Bio-One, Frickenhausen,
565 Germany, cat# 655904) and incubated in the dark for 1 h at room temperature. Luminescence
566 for each sample was measured on a SpectraMax ID3 (Molecular Devices, San Jose, CA, USA)
567 with 2 s integration time. The normalized luminescence ratio (NLR) was calculated by dividing
568 the raw luminescence of each pair (N1-X N2-Y) by the maximum luminescence value of one

569 of the two background measurements. All obtained NLR values were \log_2 transformed and the
570 positive fraction for each dataset was determined at \log_2 NLR thresholds between -2 and 2 , in
571 0.01 increments. Statistical results were robust across a wide range of stringency thresholds.
572 Extended Data Table 3 reports the results at \log_2 NLR = 0 . Reported P values were calculated
573 by Fisher's exact test.

574 **Interactome framework parameter calculation**

575 *Assay sensitivity* (S_a), i.e., the fraction of detectable interactions was assessed employing the
576 effector bhLit_BM-v1 (54 pairs) and bhRRS-v1 (73 pairs) as well as the human hsPRS-v2 (60
577 pairs) and hsRRS-v2 (78 pairs) for benchmarking. All reference sets were tested 4 times using
578 the Y2H screening pipeline. To assess sampling sensitivity (S_s) a repeat screen was
579 conducted. 288 bacterial effectors were screened 4 times against 5 pools comprising 1,475
580 human proteins. A saturation curve was calculated as described⁸⁵. Briefly, all combinations of
581 the number of interactions of the 4 repeats were assembled and the reciprocal values
582 calculated. From these a linear regression was determined to obtain the slope and the
583 intercept. Reciprocal parameters were calculated to find V_{max} and K_m and using the Michaelis-
584 Menten-formula a saturation curve was predicted. *Overall sensitivity* emerges from both
585 sampling and assay limitations and is calculated as $S_o = S_a * S_s$.

586 **Sequence similarity and interaction profile**

587 To investigate the relationship between the similarity of effector sequences and the similarity
588 of their interaction profiles we calculated the pairwise Jaccard index, which measures the
589 overlap between two effectors' interaction profiles. We calculated the Jaccard index of all
590 possible effector pairs within a homology cluster. This index represents the ratio of number of
591 human proteins targeted by both effectors to the total number of human proteins targeted by
592 either of them. For our analysis, we only considered effector pairs where the total number of
593 human proteins that are targeted by either effector was at least 3. We implemented the
594 calculations described here as commands in R version 4.2.1.

595 **Interface predictions**

596 We used as input a representative set of effectors identified in isolated strains (2300
597 sequences clustered at 90% sequence identity) and all effectors identified in MAGs (186). We
598 ran *mimicINT* as described in³² and available at [[https://github.com/TAGC-
599 NetworkBiology/mimicINT](https://github.com/TAGC-NetworkBiology/mimicINT)]. Briefly, *mimicINT* performs domain searches in effector sequences
600 with InterProScan⁹⁰ using the domain signatures from the InterPro database⁹¹ retaining
601 matches with an E-value below 10^{-5} . For host-like motif detection, *mimicINT* uses the SLiMProb
602 tool from the SLiMSuite software package⁹² by exploiting the motif definitions available in the
603 ELM database⁹³. Motifs are detected in disordered regions as defined by the IUPred

604 algorithm⁹⁴ using both short and long models (motif disorder propensity = 0.2, minimum size
605 of the disordered region = 5). The interface inference step relies on the 3did database⁹⁵ (ii) the
606 ELM database⁹³. The workflow checks whether any of the effector proteins contains at least
607 one domain or motif for which an interaction template is available. In this case, it infers the
608 interaction between the given protein and all the host proteins containing the cognate domain
609 (i.e., the interacting domain in the template). To control for false positive inference using motif-
610 domain templates, mimicINT provides two scoring strategies. First, considering binding
611 specificity of domains belonging to the same group (as PDZ or SH3)⁹⁶ an HMM-based domain
612 score⁹⁷ is computed used to rank or filter the inferred interactions. Second, given the
613 degenerate nature of motifs⁹⁸, mimicINT, using Monte-Carlo simulations, assesses the
614 probability of a given SLiM to occur by chance in query sequences and, thus, can be used to
615 filter false positives⁹⁹. This statistical approach randomly shuffles the disordered regions of the
616 input sequences to generate a large set of N randomized proteins.

617 Here, we first grouped effectors sequences by strain and effectors from MAGs were assigned
618 to the closest strain. In the first experiment, disordered regions were shuffled 100,000 times
619 using as background the effector sequences from the same strain (within-strain shuffling). In
620 the second, regions were shuffled 100,000 times using as disorder background the full set of
621 effector sequences (inter-strain shuffling). Subsequently, the occurrences of each detected
622 motif in each effector sequence were compared to the occurrences observed in the
623 corresponding set of shuffled sequences. We considered as significant all the motif
624 occurrences having an empirical *P* value lower than 0.1. To evaluate whether the number of
625 interface-resolved interactions inferred by mimicINT is significantly different from chance, we
626 generated 10,000 random networks by sampling human proteins from the interaction search
627 space in a degree-controlled manner. We then counted how many randomly generated
628 networks mimicINT inferred a higher number of interfaces than for the one observed in the
629 main screen network. Results and statistical details are in Extended Data Table 3.

630 **Holdup assay**

631 Domain production: 54 human PDZ domains and the 11 tandem constructs were
632 recombinantly expressed as His₆-MBP-PDZ constructs in *E. coli* BL21(DE3) pLysS in NZY
633 auto-induction LB medium (nzytech, MB17901)¹⁰⁰. PDZ domains were purified by Ni²⁺-affinity
634 with a 96-tip automated liquid-handling system (Tecan Freedom Evoware) using 800 µl of Ni²⁺
635 Beads (Chelating Sepharose Fast Flow immobilized metal affinity chromatography, Cytiva) for
636 each target. The domains were eluted in 2.5 ml of elution buffer: 250 mM imidazole, 300 mM
637 NaCl, 50 mM Tris, pH 8.0 buffer, and then desalted using PD10 columns (GE healthcare,
638 17085101) into 3.5 ml of 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM Imidazole buffer.
639 Concentration of desalted His₆-MBP-PDZ was determined using absorption at 280 nm on a

640 PHERAstar FSX plate reader (BMG LABTECH). Stock solutions were diluted to 4 μ M and
641 frozen at -20°C. To assess purity and confirm the concentrations, proteins were further
642 analyzed by SDS-PAGE (LabChip™ GXII, Perkin Elmer). Peptides: 10-mers corresponding to
643 the C-terminal sequences of effectors were ordered as synthetic biotinylated peptides from
644 GenicBio Limited (Shanghai, China); the N-terminal biotin was attached via a 6-aminohexanoic
645 acid linker, which we showed does not alter the peptide's binding or structural properties³⁴.
646 Purity was assessed by HPLC and mass spectrometry; all peptides were >95% pure.
647 Depending on the amino acid composition and charge peptides were solubilized in dH₂O, 1.4%
648 ammonia or 5% acetic acid, aliquoted at 10 mM concentration and stored at -20°C.

649 For the hold-up assay we followed published procedures^{34,35}. Briefly, 2.5 μ l of Streptavidin resin
650 (Cytiva, 17511301) were incubated for 15 min with 20 μ l of a 42 μ M biotinylated peptide
651 solution, in each well of a 384-well MultiScreenHTS™ filter plate (Millipore, MZHVN0W10).
652 The resin was washed with 10 resin volumes (resvol) of hold-up buffer (50 mM Tris HCl, 300
653 mM NaCl, 10 mM imidazole, 5 mM DTT), and depleted by incubation for 15 min with 5 resvol
654 of a 1 mM biotin solution, and three washes with 10 resvol of hold-up buffer. A single PDZ
655 domain was then added to each well, incubated for 15 min with the peptide bound to the resin
656 and the unbound PDZ was recovered by centrifugation into 384-well black assay plates for
657 fluorescence readout. The concentration is quantified by intrinsic Trp fluorescence,
658 fluorescein/mCherry was used for peak normalization. Binding affinities and equilibrium
659 dissociation constants (k_D) were calculated as in³⁴, using the mean PBM concentration for k_D
660 calculations. Raw values and statistical analysis are in Extended Data Table 3.

661 **Fluorescent polarization**

662 All FITC labelled peptides were synthesized as 10-mers by Biomatik, Canada, as acetate salts
663 of >98% purity. The FP experiments were performed with the His₆-MBP-PDZ proteins in 50
664 mM Tris, 300 mM NaCl, 1 mM DTT, pH 7.5 buffer in 384-well plates (Corning 3544). For direct
665 binding the His₆-MBP fused PDZ domains were two-fold serially diluted with 12 dilutions, and
666 a final volume of 10 μ l. These were then incubated with 50 nM of the FITC labelled viral
667 peptides and the plates were then read out after 1 h in FlexStation 3 (Molecular Devices) at
668 23°C, using 485 nm excitation and 520 nm emission. For competition experiments, the PDZ
669 domain and FITC peptide were kept constant at 6 μ M and 50 nM, respectively. The bacterial
670 effectors peptides in 1% ammonia buffer were added to the PDZ in a four-fold dilution, (5
671 concentrations: 0 to 31.25 μ M) and incubated at room temperature for 2 h. The FITC peptides
672 were then added and further incubated for 1 h at RT. The plates were then read as above.
673 Statistical analysis was performed using the Kruskal-Wallis test with Dunn's test followed by
674 an FDR-correction. Raw values and statistical analysis are in Extended Data Table 3.

675 **Effector convergence**

676 To estimate the significance of effector convergence, we performed a permutation test by
677 randomly sampling ‘target’ nodes ($n = 979$) from Y2H identifiable proteins from the human
678 reference interactome map, HuRI⁸⁶, as the sampling space ($n = 8,274$). We used sampling
679 with replacement to allow repeatedly picking a protein. In each iteration, the number of
680 distinctly targeted proteins was counted. The resulting distribution from 10,000 random
681 permutations was used to calculate the z-score of the experimentally observed targets ($n =$
682 349). The P value is the area under the curve for the standard normal distribution up to a given
683 z-score. We calculated the P value as implemented in the “pnorm()” R function using the z-
684 score as input. To account for the two-tailed test, the P value was multiplied by 2. To avoid
685 artifacts due to differential sampling we only considered interactions in the HuMMI_{MAIN},
686 excluding those human proteins targeted by effectors of the unknown strains and targets
687 outside HuRI. The rationale for the latter is that a substantial proportion of proteins that are not
688 in HuRI may not be suitable for Y2H analysis. Thus, restricting the analysis to the HuRI subset
689 increases the stringency.

690 To estimate the significance of the convergence of effectors from different strains (interspecies
691 convergence), we used a conditional permutation test that preserves the strain contribution.
692 For each iteration, we generated 18 samples, where for each sample, we randomly picked the
693 number of proteins equivalent to the observed targets of each strain (Extended Data Table 3).
694 From the full list of random picks that are assigned to all strains, the frequency of selecting a
695 protein was recorded. This frequency is the convergence value which indicates the number of
696 targeting strains. Using the convergence value distribution obtained from 10,000 iterations, we
697 identified the statistically significant number of strains sharing a target. The observed
698 convergence value ranges from 2 to 15 strains. We calculated the z-scores using the
699 convergence value distribution obtained from the conditional permutation test and the
700 associated P values as implemented in the “pnorm()” R function. The significant convergence
701 value (P value < 0.004) starts at 4 strains. We considered any target that is in common between
702 at least 4 strains to be subject to interspecies convergence.

703 **Function enrichment analysis**

704 We used the “gost()” function from the gprofiler2 version 0.2.1 R package¹⁰¹ to identify enriched
705 functions in effector targets. This function implements a hypergeometric test to estimate the
706 significance of the abundance of genes considering the frequency of the genes in the function
707 annotation databases. The main input argument for this function is the gene list (“query”). The
708 function allows the user to optionally set input arguments, including the background
709 (“custom_bg”), evidence codes (“evcodes”), annotation databases (“sources”), methods for

710 correcting the hypergeometric test P values (“*correction_method*”), and other arguments that
711 were set to their default options. We used the target official symbol identifiers as the “query”
712 argument. The list of HuRI proteins was the “*custom_bg*” argument. The annotations inferred
713 from electronic annotations were excluded by setting the “*exclude_iea*” argument to “*TRUE*”.
714 The hypergeometric test P values were corrected using Benjamin-Hochberg method by setting
715 the “*correction_method*” argument to “*fd*”. The argument (“*sources*”) was set to a vector
716 (“*GO:BP*”, “*KEGG*”, “*REAC*”), which encodes the search space across three function annotation
717 databases: gene ontology biological process terms (“*GO:BP*”) ¹⁰², Kyoto encyclopedia of genes
718 and genomes (“*KEGG*”) pathways ¹⁰³, and Reactome pathway database (“*REAC*”) ¹⁰⁴. After
719 plugging in these inputs into the “*gost()*” function, the output is a named list where “result” is a
720 data frame that tabulates the enrichment analysis results. We calculated the odds ratio and
721 the fold enrichment to estimate the effect size of each tested function. The odds ratio was
722 calculated for each function as the odds in the target set divided by the odds in the HuRI set.
723 The odds in the target set are the number of function-annotated target proteins divided by that
724 of the function-unannotated target proteins. Similarly, the odds in the HuRI set are the number
725 of function-annotated HuRI proteins divided by that of function-unannotated HuRI proteins.
726 The fold enrichment was calculated for each function by comparing the number of function-
727 annotated target proteins to that of the expected. The expected value represents the number
728 of function-annotated target proteins that is expected randomly based on the HuRI
729 background. It is the product of the total number of targets ($n = 349$) by the rarity. The rarity is
730 the number of function annotated HuRI proteins divided by the sum of annotated HuRI proteins.
731 The total HuRI proteins annotated for GO:BP, KEGG, and REAC, are 6988, 3250, and 4592,
732 respectively. Statistical details are in Extended Data Table 5.

733 **Metabolic subsystem analysis**

734 Several metabolism-related functions were significantly enriched in target proteins; therefore,
735 we tested the abundance of targeted enzymes in metabolic subsystems using the human
736 genome-scale metabolic model Recon3D ⁴⁶. To focus on metabolic enzymes as opposed to
737 signaling enzymes, we excluded ligases and kinases from Recon3D analyses. We performed
738 the hypergeometric test using the R function “*phyper()*” for each subsystem annotated in
739 Recon3D ($n = 95$). The inputs to this function are: the number of subsystem-annotated targeted
740 enzymes, the number of subsystem-annotated Recon3D enzymes, the number of subsystem-
741 unannotated Recon3D enzymes, and the number of targeted enzymes ($n = 16$). The nominal
742 P values were corrected using Benjamin-Hochberg. We calculated the odds ratio and the fold
743 enrichment using the same calculations described above for functional enrichments.

744 **Random walk-based determination of commensal effector network neighborhoods**

745 We have implemented a network propagation protocol based on a Random Walk with Restart
746 (RWR) algorithm RWR-MH¹⁰⁵ to explore the network vicinity of the commensal effectors in
747 HuRI⁵⁴, which contains 338 target proteins (HuMMI_{MAIN} screen) of 243 commensal effectors.
748 We used the human effector targets as seeds for the random walk and set the restart
749 probability to the default value of 0.7. In this way, we obtained a ranked list of proteins in the
750 network: the ones with the higher scores are more proximal to the seeds than those with lower
751 scores. To assign statistical significance to the computed RWR scores, we implemented a
752 normalization strategy based on degree-preserving network randomizations¹⁰⁶. We thus
753 generated 1,000 random networks from HuRI and ran the RWR algorithm to compute 1,000
754 scores for each network protein. We then computed an empirical *P* value for each protein in
755 the network keeping as neighbor proteins only those with an empirical *P* value < 0.01.

756 **Disease enrichment analysis**

757 We tested the association of all target proteins, or those subject to convergence, with human
758 diseases by performing a two-sided Fisher's exact test. We used the disease-causal genes
759 identified by the Open Targets genetic portal, which prioritizes genes at GWAS loci based on
760 variant-to-gene distance, molecular QTL colocalization, chromatin interaction, and variant
761 pathogenicity¹⁰⁷. This machine-learning approach assigns a locus to gene (l2g) score to
762 identify the most likely causal gene for the genetic variation signal of any marker SNP. We
763 considered a score of 0.5 or more as a threshold, as recommended by the authors¹⁰⁸. The
764 Fisher's exact test was performed using the function "*fisher.test()*" from "*stats*" R package
765 version 4.2.2 with its default inputs whenever applicable. The input to this function is a 2 x 2
766 contingency table, where columns represent the query set and the background set, and rows
767 denote the absence or presence of causal genes in the respective set. HuRI proteins were
768 used as the background set, and the query set was either the target proteins or those subject
769 to convergence. The calculated nominal *P* values from this function were then corrected using
770 the Benjamin-Hochberg method as implemented in the "*p.adjust()*" function. The odds ratio
771 and fold enrichment values were calculated as described in the functional enrichment section.
772 Statistical details are in Extended Data Table 5.

773 **Association with human traits and phenotype in network neighborhoods**

774 For each set of significant neighborhood-proteins we tested for enrichment of Open Targets
775 causal genes for human traits that had been investigated by 3 or more studies and for which
776 the Open Targets initiative identified 3 or more causal genes (l2g ≥ 0.5). We used a two-sided
777 Fisher's exact test to assess whether a given strain neighborhood is enriched in protein
778 associated with a human trait or phenotype followed by Benjamini-Hochberg multiple testing
779 correction. This yielded no significant association (FDR < 0.05). We therefore focused on 400

780 associations with a nominal P value < 0.01 and an OR > 3 . Some disease categorizations were
781 adjusted to better reflect etiology. Thus, Sjogren syndrome, eczema and psoriasis were
782 considered an ‘immunological’ rather than eye or skin traits, and osteoarthritis was labeled as
783 a disease of “musculoskeletal or connective tissue” rather than metabolic. For Fig. 4d some
784 closely related traits were merged, i.e., three asthma terms and three psoriasis terms.
785 Statistical details are in Extended Data Table 5.

786 **NF- κ B activation assay**

787 HEK 293 (RRID: CVCL_0045, DSMZ) were maintained in DMEM with 10% FBS and 100 U/mL
788 penicillin and 100 U/mL streptomycin at 37°C and 5% CO₂. IKK β (in pRK5 with a Flag-tag)
789 served as positive control whereas A20 (in pEF4 with a Flag-tag) as the negative control. In a
790 60 mm cell culture dish 1×10^6 cells were seeded in 3 ml Medium. After 24 h cells were
791 transfected using 10 ng NF- κ B reporter plasmid ($6 \times$ NF- κ B firefly luciferase pGL2), 50 ng pTK
792 reporter (renilla luciferase) and 2 μ g bacterial ORF in pMH-FLAG-HA. The DNA was added to
793 200 μ l 250 mM CaCl₂ solution (Carl Roth cat. no. 5239.1), vortexed and added dropwise to
794 200 μ l $2 \times$ HBS (50 mM HEPES (pH 7.0) (Carl Roth cat. no. 9105.4), 280 mM NaCl (Carl Roth
795 cat. no. 3957.2), 1.5 mM Na₂HPO₄ \times 2 H₂O (Carl Roth cat. no. 4984.1, pH 6.93) which was
796 vortexed. After 15 min incubation, the mixture was added dropwise to the cells. Medium was
797 changed after 6 h incubation. To assess NF- κ B inhibition, cells were treated for 4 h with 20
798 ng/ml TNF (Sigma-Aldrich cat. no. SRP3177) 24 h after transfection. Samples were washed,
799 lysed, centrifuged and the supernatant was measured using the dual luciferase reporter kit
800 (Promega, E1980) with a luminometer (Berthold Centro LB960 microplate reader, Software:
801 MikroWin 2010). NF- κ B induction was determined as Firefly luminescence to Renilla
802 luminescence. P values were calculated using the Kruskal-Wallis test with Dunn’s correction
803 followed by an FDR-correction. Raw values and statistical analysis are in Extended Data Table
804 6.

805 Protein expression levels were checked by Western Blots. Proteins were separated by SDS-
806 PAGE and transferred on polyvinylidene fluoride membranes, and after transfer blocked with
807 5% milk in $1 \times$ PBS + 0.1% Tween-20 (PBST) for 1 h at room temperature. Primary antibodies
808 were added in 2.5% BSA in PBS-T buffer at 4°C overnight. After 3 x 15min washes with PBS-
809 T anti-mouse secondary antibody was added at a 1:10,000 dilution for 1 h at RT (Jackson
810 ImmunoResearch Labs cat. no. 715-035-150, RRID:AB_2340770). Primary antibodies: anti-
811 Actin beta (SCBT cat. no. sc-47778, RRID:AB_626632) at a 1:10,000 dilution, anti-FLAG M2
812 (Sigma Aldrich cat. no. F3165, RRID:AB_259529) at a 1:500 dilution and anti-HA (Sigma-
813 Aldrich cat. no. 11583816001, RRID:AB_514505) at a 1:1,000 dilution. For detection the
814 LumiGlo reagent (CST cat. no. 7003S) and a chemiluminescence film (Sigma-Aldrich cat. no.
815 GE28-9068-36) were used.

816 **ICAM1 assay**

817 Caco-2 cells were maintained in DMEM Glutamax medium (Gibco) supplemented with 10%
818 FBS, 1% Pen/Strep at 37°C in a humidified 5% CO₂ incubator. Medium was refreshed twice
819 a week. Caco-2 cells were plated in both 24- and 96-well plates 24 h before transfection. Six
820 hours prior to transfection, culture medium was replaced with supplement-free DMEM. Co-
821 transfusions were performed using 40,000 MW linear polyethylenimine (PEI MAX®)
822 (Polysciences, Warrington, USA) at a ratio of 1:5 pDNA:PEI. Equimolar ratios of the eGFP-
823 plasmid and effector-plasmid were used to ensure equimolar representation of relevant ORFs.
824 In total, 250 ng and 1 µg pDNA was added per well of the 96- and 24-well plates, respectively.
825 pDNA-PEI complexes were formed by incubating pDNA and PEI at RT for 15 minutes, followed
826 by the addition of supplement-free DMEM and another incubation of 15 minutes at RT. Cells
827 were then exposed to the transfection mixture for 16 h, washed, and rested for 6 h in complete
828 DMEM. Subsequently, cells were stimulated using an activation mix containing 200 ng/ml PMA
829 (P8139-1MG, Sigma-Aldrich), 100 ng/ml LPS (L6529-1MG, Sigma-Aldrich), and 100 ng/ml
830 TNF (130-094-014, Miltenyi Biotec). In 24-well plates, cells were stimulated for 24 h and
831 detached from the plate using ice-cold PBS. In the 96-well plate, cells were stimulated for 48
832 h, treated with BD GolgiStop™ (554724, BD Biosciences) in the final 6 h of stimulation, and
833 detached using trypsin/EDTA. Cells were washed twice and ICAM1 was stained using an anti-
834 ICAM1 PE (#MHCD5404-4, Invitrogen) antibody. The mean fluorescent intensity of the GFP+
835 cell population was measured on a FACSFortessa™ flow cytometer (BD) and the data was
836 analyzed using FlowJo V10.8.1 (BD). After positive tests for normal data distribution,
837 significance was assessed using a one-way ANOVA with Dunnett's multiple comparisons test.
838 Raw values and statistical analysis are in Extended Data Table 6.

839 **Cytokine assays**

840 Caco-2 cells were plated in 100 mm cell culture dishes three days prior to transfection. The
841 transfection protocol was identical to that described above, however, a total of 20 µg pDNA
842 was used per dish. Upon overnight transfection, cells were detached using Trypsin/EDTA and
843 resuspended in cell sorting buffer (PBS + 2% FBS + 2mM EDTA). GFP+ cells were sorted into
844 ice-cold FBS using a BD FACSAria III cell sorter (BD) and transferred to a 96-well plate at
845 30,000 cells per well. Upon a 24 h rest-period, cells were activated for 48 h using the activation
846 mix described above. During cell stimulation, cell proliferation was monitored through
847 longitudinal imaging of cell confluency in the Incucyte S3 Live cell analysis system (Essen
848 BioScience). Cytokine levels were determined using the human inflammation panel 1
849 LEGENDplex™ kit (Biolegend) following the manufacturer's instructions. Cell culture
850 supernatant of the above samples was used to analyze IL1beta. To this end, IL1beta ELISAs
851 were performed using the ELISA MAX™ Deluxe Set Human IL1beta kit (437015, Biolegend)

852 following the protocol provided by the manufacturer. Statistical significance was evaluated
853 using Kruskal-Wallis test with uncorrected Dunn's test. Raw values and statistical analysis are
854 in Extended Data Table 6.

855 **Protein ecology**

856 Metagenomic assemblies from the Inflammatory Bowel Disease Multi'omics DataBases
857 (IBDMBD)⁶⁴ and from the skin metagenome¹⁰⁹ were downloaded, and each samples protein
858 repertoire predicted using Prodigal (options; -p meta)¹¹⁰. Effector proteins were compared to
859 the metagenomic protein repertoires using DIAMOND (options; >90% query length, >80%
860 identity). For analyses in Fig. 5, samples were grouped into patients with UC (n = 304), CD (n
861 = 508), and controls without IBD (n = 334). The annotations were then converted into binarised
862 vectors of presence and absence of each effector across the sample and the Fischer exact
863 test, implemented within scipy python module, was used to determine if the prevalence of each
864 effector occurring within CD or UC patient metagenomes compared to controls. Significance
865 was then corrected using the Benjamini-Hochberg method. The significance of differences in
866 prevalence distributions between healthy and either patient cohort were estimated by Wilcoxon
867 rank-sum test, implemented in the "*wilcox.test()*" R function. Statistical details in Extended Data
868 Table 6.

869 **Statistics and reproducibility**

870 Data were subjected to statistical analysis and plotted to Microsoft Excel 2010 or python or R
871 scripts. For comparison of normally distributed values we used one-way ANOVA, for
872 assessment of overlap for comparison of values not passing the normality tests we used
873 Kruskal-Wallis test with Dunn's corrected as appropriate and indicated in the figure legends
874 and methods. Enrichments were calculated using Fisher's exact test with Bonferroni FDR
875 correction. All statistical evaluations were done as two-sided tests. Generally, a corrected *P*
876 value < 0.05 was considered significant. GO, KEGG, and Reactome functional enrichments
877 were calculated using profiler with the respectively indicated background gene sets. For the
878 disease target enrichments and neighborhood associations no associations were significant
879 after multiple hypothesis correction, which is why nominally significant associations calculated
880 by Fisher's exact tests were used for Fig. 4c,d. All raw values, n, and statistical details are
881 presented in supplementary tables as indicated in the Figure legends and methods sections.

AUTHOR CONTRIBUTIONS

Project conception: PFB

T3SS and effector analyses: PH, TH, SA, CB, AZ, TR, PFB

ORF cloning: VY, MR, MA, AS, PFB

Interactome mapping and validation: VY, SR, BW, AS, PFB

Interaction curation: VY, MA, MB, AZ, CF, PFB

Data analyses: BD, VY, DS, CWL, MB, SAC, PS, CB, AZ, PFB

Interface identification and validation: SAC, AZ, JFM, SBM, JCT, RV

Effector ecology: TH, TC

Cell-based assays: VY, NvdH, FO, PFB, DK, MB

Visualization: VY, BD, JFM, AZ, PFB

Funding acquisition: PFB, CF, AZ, CB, TR, DK

Manuscript writing and editing: PFB, VY, BD, BW, TH, CF, AZ

ACKNOWLEDGEMENTS

Plasmids and strains for hsPRS/RRS_v2 were kindly provided by Marc Vidal, David E. Hill, and Mike Calderwood, CCSB, Dana-Farber Cancer Institute, Boston, MA. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC). Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources.

REPORTING SUMMARY

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

All sequence, interaction, and functional data generated in this study are available as supplementary information. The effectors identified and cloned for interactome mapping are presented in Extended Data Table 1. All protein-protein interaction data acquired in this study can be found in Extended Data Table 2 and Extended Data Table 3. The data for functional validation assays can be found in Extended Data Table 6. The protein interactions from this publication have been submitted to the IMEx (<http://www.imexconsortium.org>) consortium through IntAct¹¹¹ and assigned the identifier IM-29849. New effector sequences have been submitted to GenBank: BankIt2727690: OR372873 - OR373035 and OR509516 - OR509528.

CODE AVAILABILITY

All source code related to this paper is available as a zip file.

COMPETING INTERESTS

The authors declare no competing interests.

EXTENDED DATA TABLES:

Extended Data Table 1: T3SS in strains of the commensal human microbiome

Extended Data Table 2: Effector identification and cloning

Extended Data Table 3: Effector host interaction map

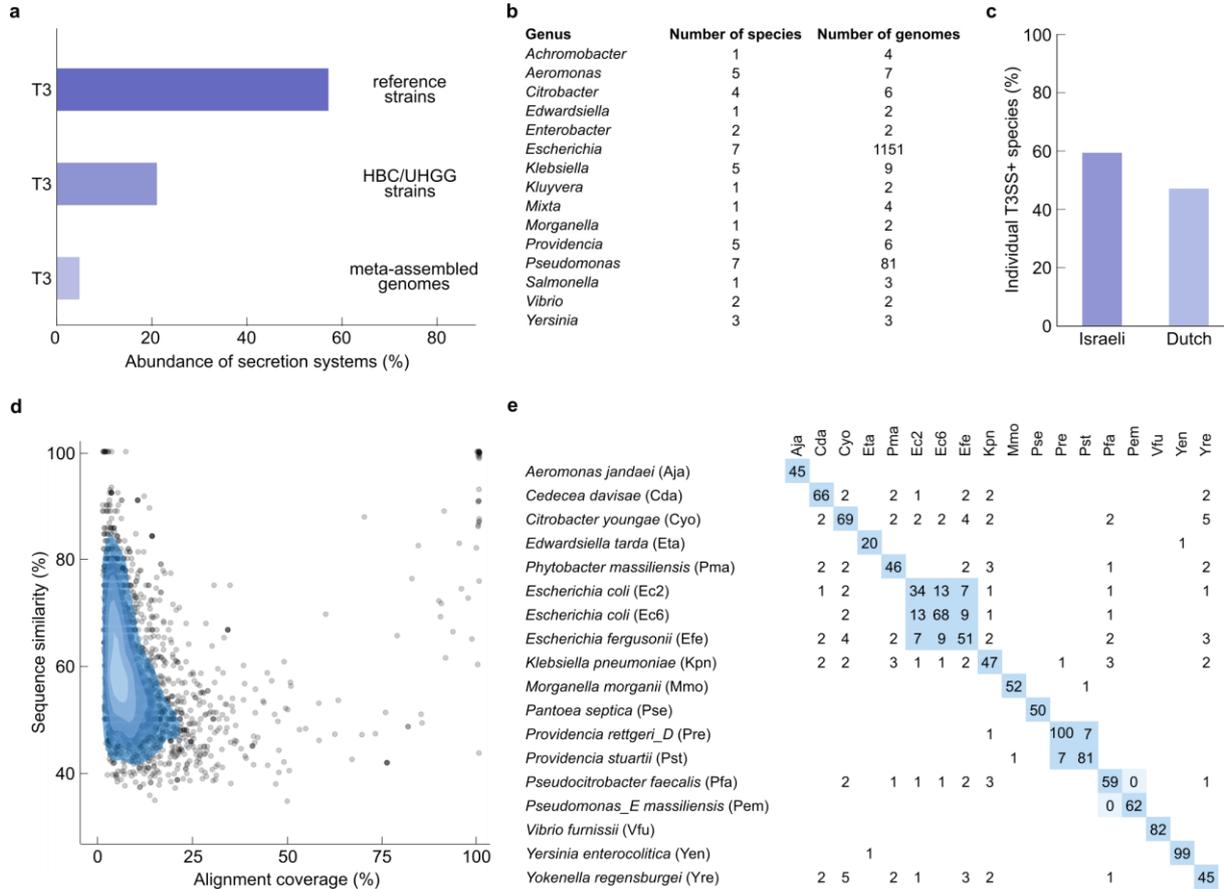
Extended Data Table 4: Interface identification and validation

Extended Data Table 5: Functional and disease enrichment

Extended Data Table 6: Functional assay data and IBD prevalence

882 **FIGURES**

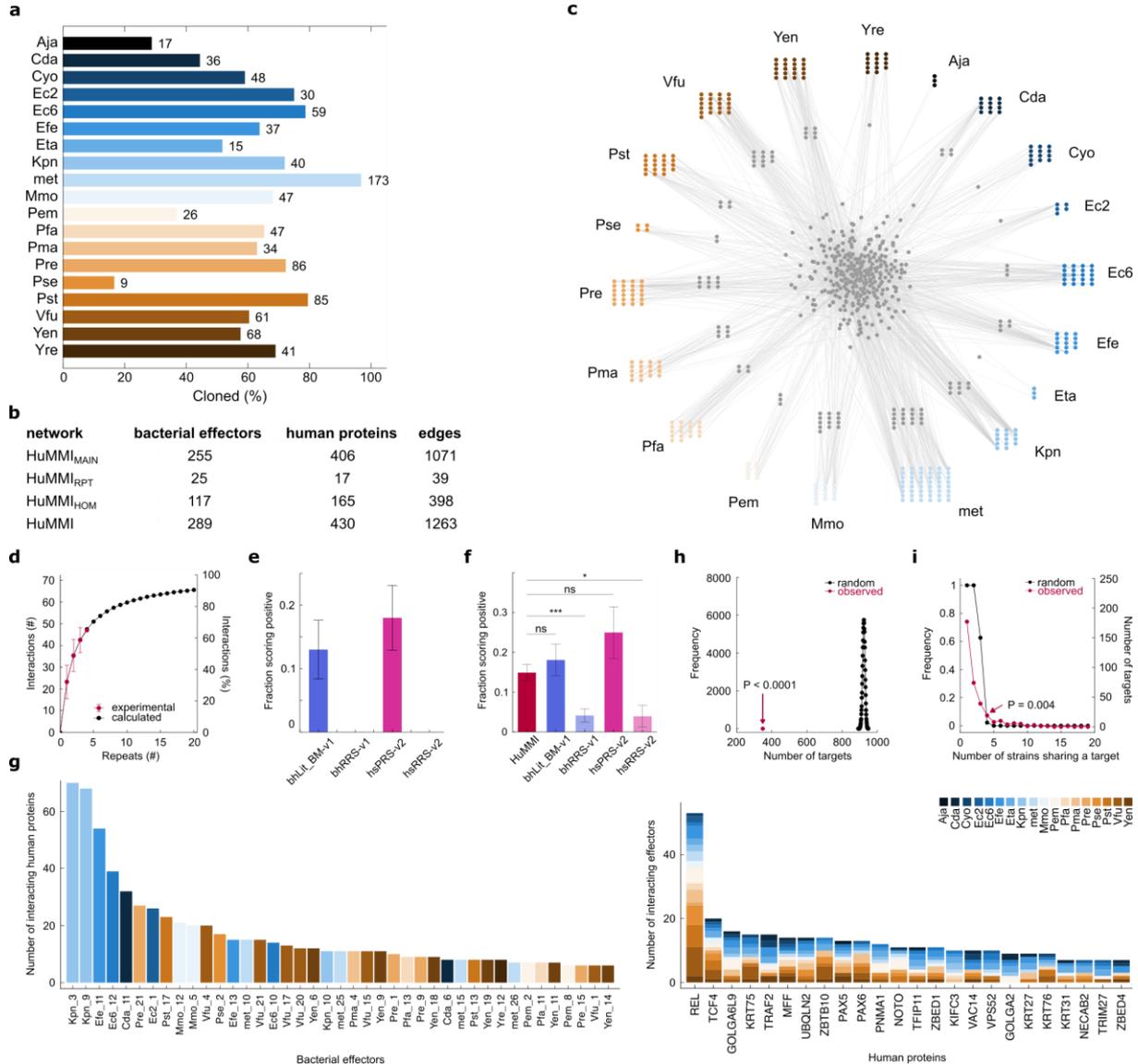
Figure 1



883

884 **Fig. 1 | T3SS in commensal bacterial species in the gut microbiome.** **a**, Proportion of
 885 *Pseudomonadota* genomes encoding complete T3SS among 77 reference strains of human
 886 intestinal and stool samples, in a collection of 4,475 strains isolated from normal human guts,
 887 and in meta-assembled genomes (MAG) of normal human guts. **b**, Most abundant genera and
 888 identified number of species and genomes encoding complete T3SS from the samples in **a**. **c**,
 889 Proportion of individuals in two human cohorts containing T3SS encoding microbial species.
 890 **d**, Similarity of 3,002 candidate effector-substrates for T3SS identified from commensal
 891 reference strains with 1,195 effectors from pathogenic microbes across the range of alignment
 892 coverages. **e**, Selection of 18 commensal *Pseudomonadota* strains with dissimilar effector
 893 complements used for subsequent functional analyses. Numbers indicate the count of shared
 894 effectors at >90% mutual sequence similarity across 90% common sequence length among
 895 the indicated strains. Full data for all panels in Extended Data Table 1.

Figure 2



896

897 **Fig. 2 | Meta-interactome network map of bacterial effectors with human proteins. a,**

898 Success rates of effector ORF cloning for each strain, and number of sequence verified ORFs

899 (right). **b,** Number of interactions and involved proteins in the HuMMI subsets. **c,** Verified

900 human microbiome meta-interactome (HuMMI) map. Grey nodes: human proteins; outer layer

901 human proteins targeted only by the nearest strain; central human proteins by effectors from

902 multiple strains. **d,** Sampling sensitivity: saturation curve calculated from the repeat

903 experiment: red dots represent average of verifiable interactions found in any combination of

904 indicated number of repeat screens; black dots and line: modeled saturation curve. **e,** Assay

905 sensitivity: percentage of identified interactions from bhLit_BM-v1 (n = 54 pairs), bhRRS-v1 (n

906 = 73 pairs), hsPRS-v2 (n = 60 pairs), hrRRS-v2 (n = 78 pairs) in our Y2H. Error bars present

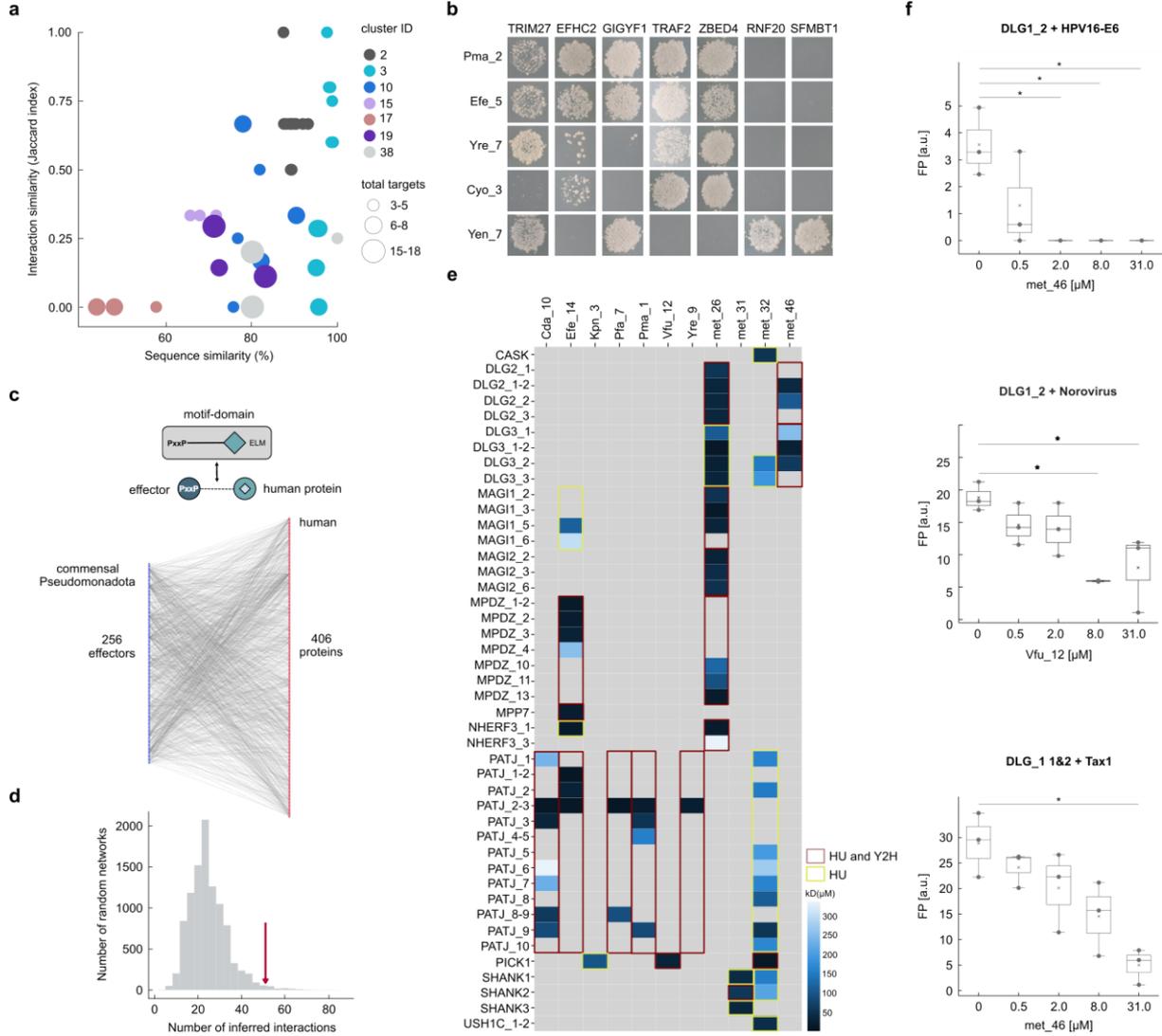
907 the standard error (SE) of proportion. **f,** Validation rate of a random sample of HuMMI

908 interactions (n = 295 pair configurations) compared to four reference sets in the yN2H

909 validation assay: bhLit_BM-v1 (n = 94 pair configurations), bhRRS-v1 (n = 145 pair

910 configurations), hsPRS-v2 (n = 44 pair configurations), hrRRS-v2 (n = 51 pair configurations).
911 * $P = 0.04$; *** $P = 0.0006$; ns “no significant difference” (Fisher exact test; Extended Data
912 Table 3). Error bars present SE of proportion. **g**, Left: degree distribution for the most
913 connected effectors; right: effector-degree distribution for most targeted human proteins.
914 Colors represent strains according to legend. **h**, Observed number of total effector targets in
915 the human reference interactome (HuRI), compared to random expectation (exp. $P < 0.0001$;
916 n = 10,000 randomizations). **(I)** Frequency distribution of human proteins targeted by effectors
917 from the indicated number of different strains (red), compared to random expectation (black; n
918 = 10,000). Targeting by effectors from four strains or more occurs significantly more often than
919 expected by chance (exp. $P = 0.004$; n = 10,000).

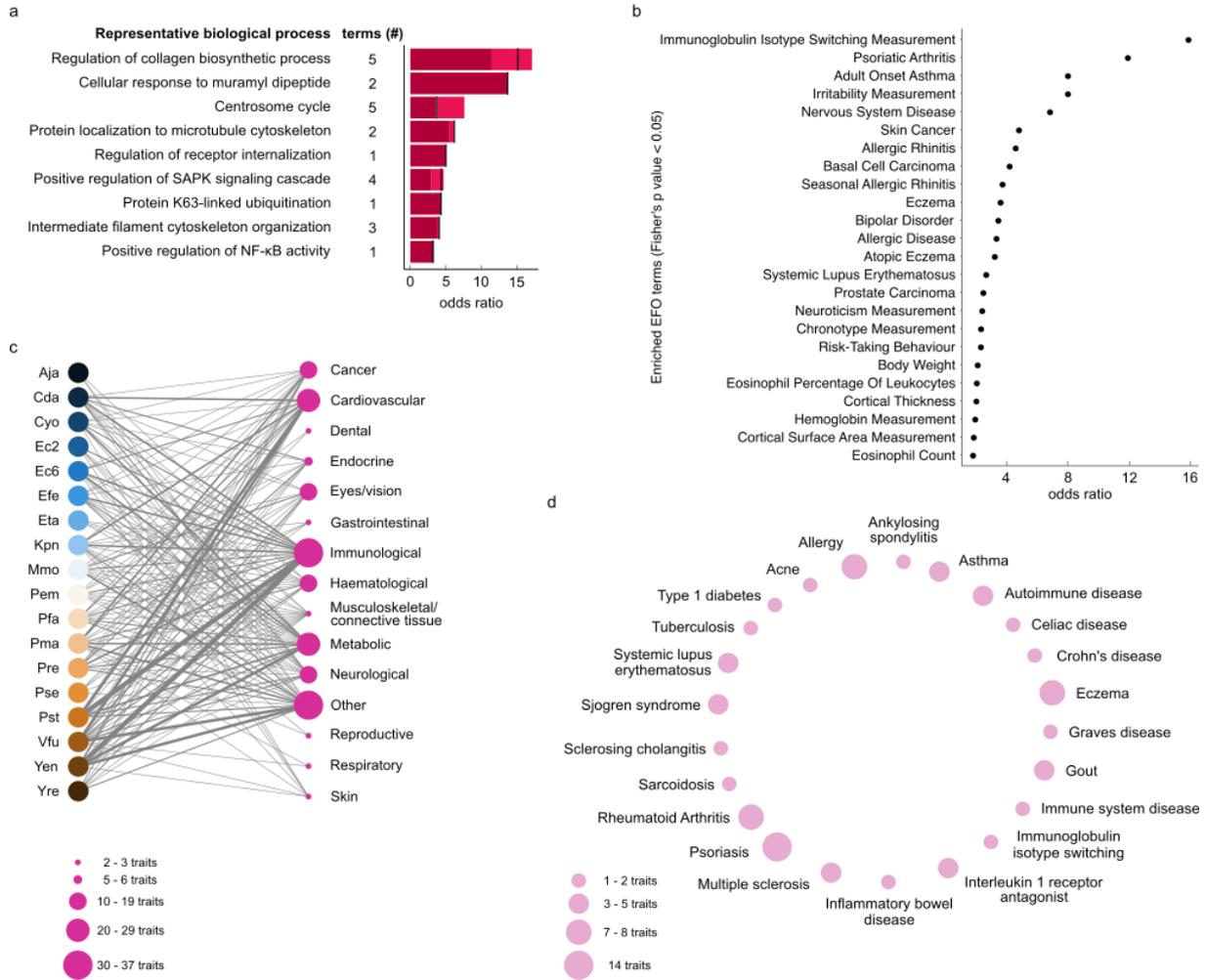
Figure 3



920

921 **Fig. 3 | Interaction specificity and interaction motifs.** **a**, Scatter plot of sequence- and
 922 Jaccard-interaction similarity for all effector pairs within indicated homology groups of
 923 HuMMI_{HOM} with ≥ 3 interactors and effectors. Node size indicates union of human proteins
 924 targeted by effector-pair according to legend. **b**, Y2H data for one of four repeats for homology
 925 cluster 3. **c**, Schematic of interaction motif-domain interface identification in the effector-host
 926 interaction. **d**, Count of motif-domain pairs matching at least one stringency criteria identified
 927 in HuMMI_{MAIN} (arrow) compared to random expectation (experimental P value, $n = 10,000$). **e**,
 928 Interaction strength of PDZ domains of human proteins with C-terminal 10 amino acid peptides
 929 of the effectors indicated on top. Calculated K_D according to legend. Overlap between HU and
 930 Y2H is indicated by colored frames. **f**, Competition of the interaction between human PDZ
 931 domains and viral PBM peptides by the indicated effector peptides. * $P < 0.05$ (Kruskal Wallis
 932 with Dunn's correction, $n = 3$). Boxes represent interquartile range (IQR), with the bold black
 933 line representing mean; whiskers indicate highest and lowest data point within 1.5 IQR.

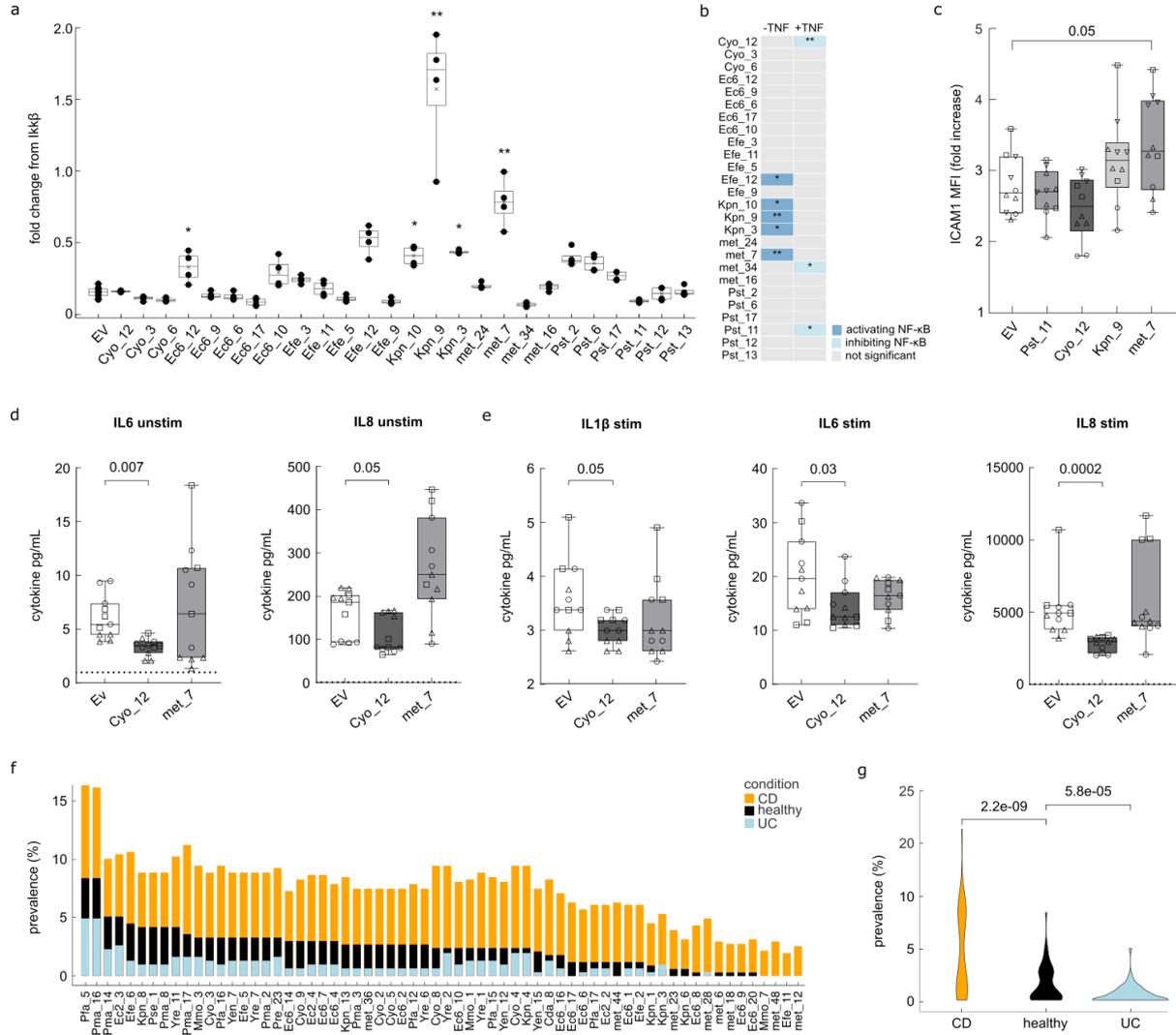
Figure 4



934

935 **Fig. 4 | Function and disease association of microbially targeted human proteins.** **a**,
 936 Odds ratios (OR) of representative functional annotations enriched among effector targeted
 937 human proteins (FDR < 0,05, Fisher's exact test with Bonferroni FDR correction). The number
 938 of represented terms is shown by terms (#). The lowest and highest OR observed for the
 939 represented group are indicated by light shaded area in each bar. Black line indicates OR for
 940 representative term. Full data and precise FDR and OR values in Extended Data Table 5. **b**,
 941 Genetic predisposition for traits and diseases enriched among human genes encoding effector
 942 targets in HuRI (cutoff FDR = 0.05, Fisher's exact test, n = 349). **c**, Disease groups for which
 943 genetic predisposition is enriched in network neighborhoods of effectors from the indicated
 944 strains. Trait node size corresponds to number of significantly targeted traits in that group
 945 according to legend. Stroke of strain-group edge reflects number of underlying significant
 946 effector-trait links ($\alpha < 0.01$ and OR > 3, Fisher's exact test). **d**, Specific diseases underlying
 947 the 'immunological' group in c. Node size reflects the number of underlying effector-trait
 948 associations according to legend.

Figure 5



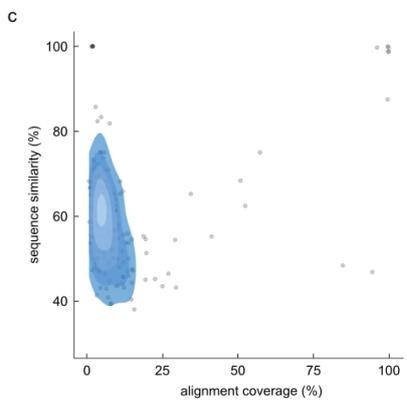
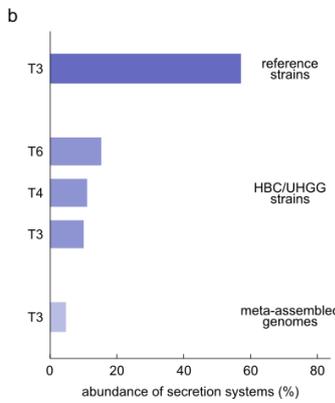
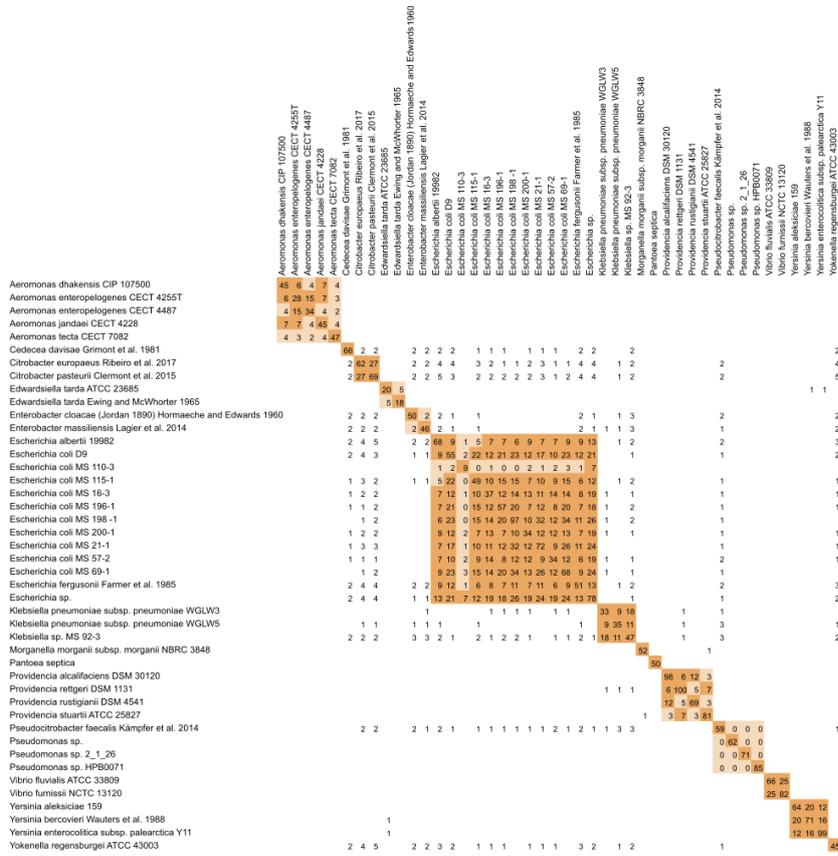
949

950 **Fig. 5 | Effector impact on human cell function and clinical prevalence in IBDs.** **a**, Relative
 951 NF-κB transcriptional reporter activity of HEK293 cells expressing the indicated effectors or
 952 empty vector (EV) in unstimulated conditions (Kruskal-Wallis test with Dunn's correction, * $P <$
 953 0.05, ** $P <$ 0.01, $n = 4$). Boxes represent IQR, black line indicates the mean, whiskers indicate
 954 highest and lowest data point within 1.5 IQR. **b**, Summary of significant impact of effectors on
 955 normalized NF-κB transcriptional reporter activity in baseline conditions and after TNF
 956 stimulation (Kruskal-Wallis test with Dunn's correction, * $P <$ 0.05, ** $P <$ 0.01, $n = 4$). **c**, Fold-
 957 induction of ICAM1 expression following pro-inflammatory stimulation of Caco-2 cells
 958 transfected with the indicated effectors (one-way ANOVA with Dunnett's multiple comparison
 959 test, $n = 10$). **d**, Concentration of cytokines secreted by Caco-2 cells in basal conditions
 960 transfected with the indicated effectors. EV indicates empty vector mock control. P values
 961 calculated by Kruskal-Wallis test ($n = 11$). Dashed line indicates detection limit of assay. **e**,
 962 Concentration of cytokines secreted by Caco-2 cells stimulated by a pro-inflammatory cocktail
 963 transfected with the indicated effectors. EV indicates empty vector mock control. Indicated P

964 values calculated by Kruskal-Wallis test ($n = 11$). Dashed line: detection limit of assay. C – E
965 Boxes represent IQR, black line indicates the mean, whiskers indicate highest and lowest data
966 point. **f**, Effector prevalence in metagenomes of CD ($n = 504$), and UC patients ($n = 302$)
967 compared to healthy controls. Effectors are significantly more prevalent in CD patient
968 metagenomes (FDR < 0.01; Fisher exact test, Benjamini-Hochberg correction). **g**, Effector
969 prevalence distribution among the indicated cohorts. P values calculated by Wilcoxon rank-
970 sum test, n as in **f**.

971 **EXTENDED DATA FIGURES**

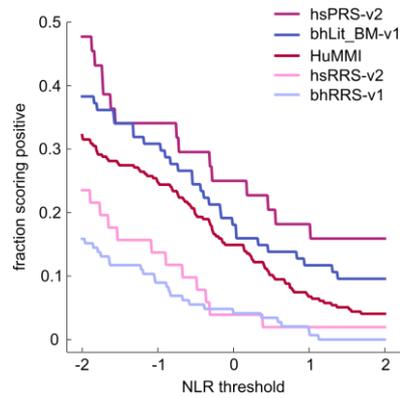
Extended Figure 1
a



972

973 **Extended Data Fig. 1 | T3SS in strains of the commensal gut microbiome. a**, Effector-
 974 complement comparison of the 44 T3SS+ Pseudomonadota reference strains. Numbers
 975 indicate the count of shared effectors at >90% mutual sequence similarity across 90% common
 976 sequence length among the indicated strains. **b**, Abundance of secretion systems in
 977 Pseudomonadota genomes among the 77 reference strains of human intestinal and stool
 978 samples, in a collection of 4,475 strains isolated from normal human guts (HBC/UHGG strains)
 979 and in meta-assembled genomes (MAG) of normal human guts. **c**, Similarity of identified 186
 980 candidate effectors from the 770 T3SS+ MAGs with 1,195 effectors from pathogenic microbes
 981 across the range of alignment coverages. Full data for all panels in Extended Data Table 1.

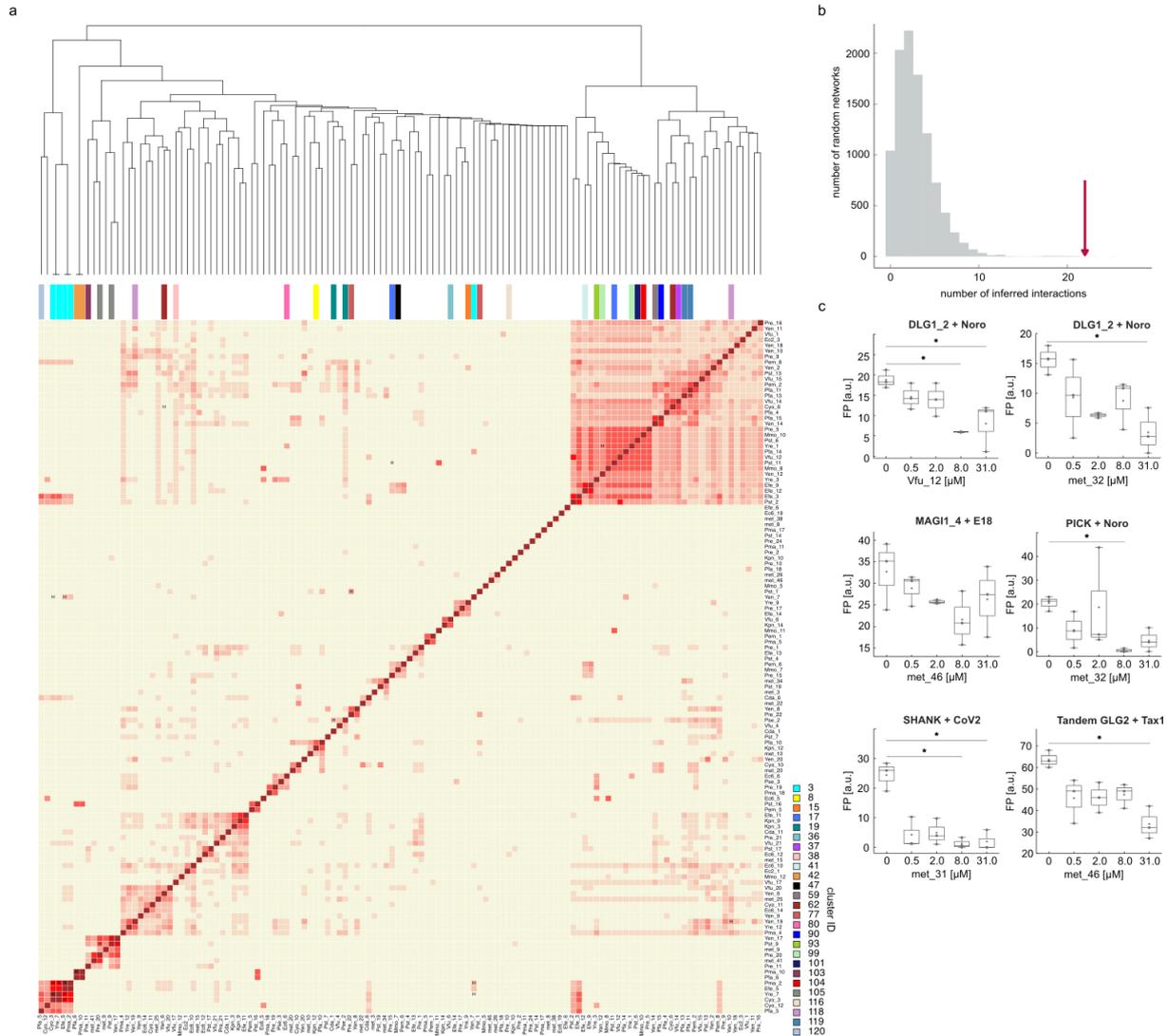
Extended Figure 2



982

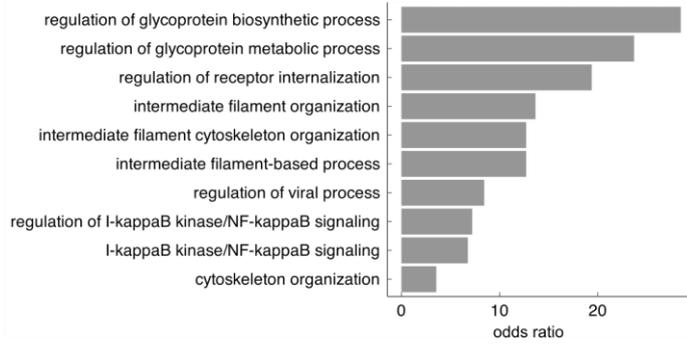
983 **Extended Data Fig. 2 | Detection rates of protein pairs in different sets across varying**
984 **thresholds in yN2H.** Fractions scoring positive of the HuMMI dataset and benchmarking
985 datasets (hsPRS-v2, bhLit_BM-v1, hsRRS-v2, bhRRS-v1) depending on the threshold of the
986 normalized luminescence ratio (NLR). Full data in Extended Data Table 3.

Extended Figure 3



Extended Data Fig. 3 | Interaction specificity and interaction motifs. **a**, Jaccard-interaction similarity of all interacting effector-pairs with at least 3 shared human interactors. Color-intensity correlates with Jaccard-index. Effector pairs marked with “H” share the same homology cluster. Clusters are color-coded according to legend. **b**, Count of motif-domain pairs matching at least two stringency criteria identified in HuMMI_{MAIN} (arrow) compared to $n = 10,000$ randomized control networks (empirical $P = 0.0003$). **c**, Competition of the interaction between human PDZ domain and viral PBM peptide by indicated C-terminal effector peptides. * $P < 0.05$ (Kruskal Wallis with Dunn’s correction, $n = 3$). Boxes indicate IQR, black line represents mean, whiskers indicate highest and lowest data point within 1.5 IQR. Precise P values and n for each test are shown in Extended Data Table 4.

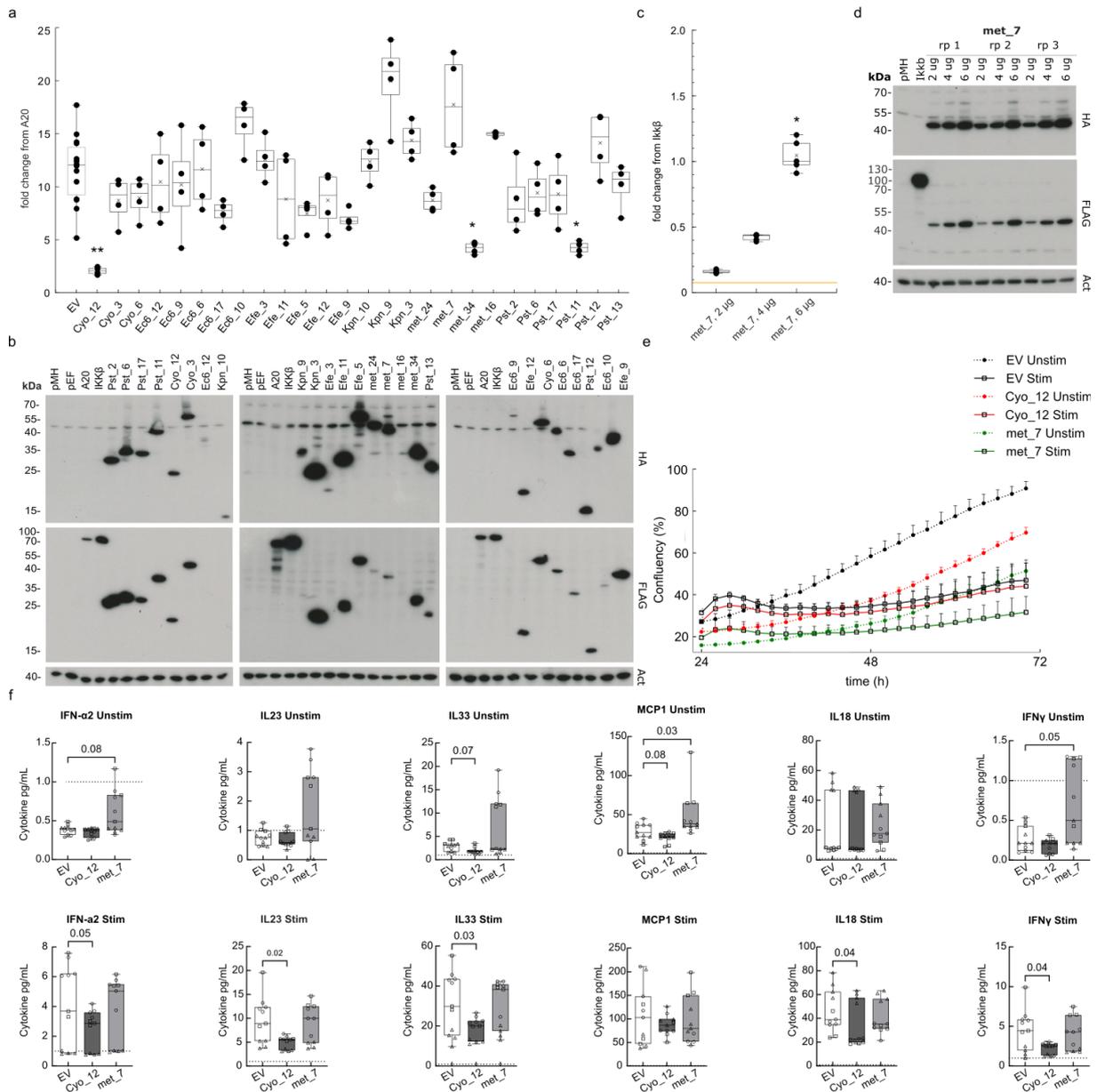
Extended Figure 4



998

999 **Extended Data Fig. 4 | GO enrichment for convergence proteins.** OR for functional
1000 annotations enriched among effector-targeted human proteins that are subject of convergence
1001 (FDR < 0.05, Fisher's exact test with Bonferroni FDR correction). Full data and precise FDR
1002 and OR values in Extended Data Table 5.

Extended Figure 5



1003

1004 **Extended Figure 5 | Effector impact on human cell function.** **a.** Relative NF-κB

1005 transcriptional reporter activity of HEK293 cells expressing the indicated effectors under TNF-

1006 stimulated conditions (Kruskal-Wallis test with Dunn's correction, * $P < 0.05$, ** $P = 0.01$, $n =$

1007 4). Boxes represent IQR, with the bold black line representing the mean; whiskers indicate

1008 highest and lowest data point within 1.5 IQR. **b.** Representative anti-Hemagglutinin (HA) and

1009 anti-Flag (FLAG) western blots showing expression of transfected effector proteins relative to

1010 actin control (ACT). Empty pMH-Flag-HA (pMH), empty pEF4 (pEF). **c.** Titration of met_7

1011 shows a concentration dependent specific increase of NF-κB reporter activity. Yellow line

1012 represents the empty vector value. (Kruskal-Wallis test with Dunn's correction, * $P < 0.05$, error

1013 bars: standard deviation of the mean, $n = 5$). Boxes represent IQR, with the bold black line

1014 representing the mean; whiskers indicate highest and lowest data point within 1.5 IQR. **d.**

1015 Representative anti-Hemagglutinin (HA) and anti-Flag (FLAG) western blots for experiment in
1016 c showing expression of transfected effector proteins relative to actin control (ACT). **e**,
1017 Representative proliferation curves of Caco-2 cells transfected with empty vector (EV), Cyo_12
1018 or met_7 in basal conditions (unstim) or following pro-inflammatory stimulation (stim) over 72
1019 h after sorting. **f**, Concentration of cytokines secreted by Caco-2 cells transfected with the
1020 indicated effectors in basal conditions (Unstim) or following pro-inflammatory stimulation
1021 (Stim). EV indicates empty vector mock control. Indicated *P* values calculated by Kruskal-
1022 Wallis test with Dunn's multiple hypothesis correction ($n = 11$). Boxes represent IQR, with the
1023 bold black line representing the mean; whiskers indicate highest and lowest data point. Raw
1024 measurements, *n*, and precise *P* values for all panels in Extended Data Table 6.
1025
1026

1027 **REFERENCES**

- 1028 1 Plichta, D. R., Graham, D. B., Subramanian, S. & Xavier, R. J. Therapeutic
1029 Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-
1030 Microbiome Relationships. *Cell* **178**, 1041-1056, doi:10.1016/j.cell.2019.07.045 (2019).
- 1031 2 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat. Med.* **24**, 392-
1032 400, doi:10.1038/nm.4517 (2018).
- 1033 3 Depner, M. *et al.* Maturation of the gut microbiome during the first year of life contributes
1034 to the protective farm effect on childhood asthma. *Nat Med* **26**, 1766-1775,
1035 doi:10.1038/s41591-020-1095-x (2020).
- 1036 4 Paun, A., Yau, C. & Danska, J. S. Immune recognition and response to the intestinal
1037 microbiome in type 1 diabetes. *J Autoimmun* **71**, 10-18, doi:10.1016/j.jaut.2016.02.004
1038 (2016).
- 1039 5 Keshavarzian, A. *et al.* Colonic Bacterial Composition in Parkinson's Disease. *Mov.*
1040 *Disord.* **30**, 1351-1360, doi:10.1002/mds.26307 (2015).
- 1041 6 Oren, A. & Garrity, G. M. Valid publication of the names of forty-two phyla of
1042 prokaryotes. *Int J Syst Evol Microbiol* **71**, doi:10.1099/ijsem.0.005056 (2021).
- 1043 7 Shin, N. R., Whon, T. W. & Bae, J. W. Proteobacteria: microbial signature of dysbiosis
1044 in gut microbiota. *Trends Biotechnol* **33**, 496-503, doi:10.1016/j.tibtech.2015.06.011
1045 (2015).
- 1046 8 Deng, W. Y. *et al.* Assembly, structure, function and regulation of type III secretion
1047 systems. *Nature Reviews Microbiology* **15**, 323-337, doi:10.1038/nrmicro.2017.20
1048 (2017).
- 1049 9 Miwa, H. & Okazaki, S. How effectors promote beneficial interactions. *Current Opinion*
1050 *in Plant Biology* **38**, 148-154, doi:10.1016/j.pbi.2017.05.011 (2017).
- 1051 10 Eichinger, V. *et al.* EffectiveDB--updates and novel features for a better annotation of
1052 bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res*
1053 **44**, D669-674, doi:10.1093/nar/gkv1269 (2016).
- 1054 11 Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved
1055 metagenomic analyses. *Nat Biotechnol* **37**, 186-192, doi:10.1038/s41587-018-0009-7
1056 (2019).
- 1057 12 Poyet, M. *et al.* A library of human gut bacterial isolates paired with longitudinal
1058 multiomics data enables mechanistic microbiome research. *Nat Med* **25**, 1442-1452,
1059 doi:10.1038/s41591-019-0559-3 (2019).
- 1060 13 Groussin, M. *et al.* Elevated rates of horizontal gene transfer in the industrialized human
1061 microbiome. *Cell* **184**, 2053-2067 e2018, doi:10.1016/j.cell.2021.02.052 (2021).
- 1062 14 Yang, X. B., Pan, J. F., Wang, Y. & Shen, X. H. Type VI Secretion Systems Present
1063 New Insights on Pathogenic *Yersinia*. *Frontiers in Cellular and Infection Microbiology*
1064 **8**, doi:10.3389/fcimb.2018.00260 (2018).
- 1065 15 Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**,
1066 499-504, doi:10.1038/s41586-019-0965-1 (2019).
- 1067 16 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over
1068 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*
1069 **176**, 649-662 e620, doi:10.1016/j.cell.2019.01.001 (2019).
- 1070 17 Shalon, D. *et al.* Profiling the human intestinal environment under physiological
1071 conditions. *Nature*, doi:10.1038/s41586-023-05989-7 (2023).

- 1072 18 Leviatan, S., Shoer, S., Rothschild, D., Gorodetski, M. & Segal, E. An expanded
1073 reference map of the human gut microbiome reveals hundreds of previously unknown
1074 species. *Nat Commun* **13**, 3863, doi:10.1038/s41467-022-31502-1 (2022).
- 1075 19 Jing, R. *et al.* DeepT3 2.0: improving type III secreted effector predictions by an
1076 integrative deep learning framework. *NAR Genom Bioinform* **3**, lqab086,
1077 doi:10.1093/nargab/lqab086 (2021).
- 1078 20 Arnold, R. *et al.* Sequence-based prediction of type III secreted proteins. *PLoS Pathog*
1079 **5**, e1000376, doi:10.1371/journal.ppat.1000376 (2009).
- 1080 21 Goldberg, T., Rost, B. & Bromberg, Y. Computational prediction shines light on type III
1081 secretion origins. *Scientific reports* **6**, 34516, doi:10.1038/srep34516 (2016).
- 1082 22 Wang, J. W. *et al.* BastionHub: a universal platform for integrating and analyzing
1083 substrates secreted by Gram-negative bacteria. *Nucleic Acids Research* **49**, D651-
1084 D659, doi:10.1093/nar/gkaa899 (2021).
- 1085 23 Ma, W. B., Dong, F. F. T., Stavrinos, J. & Guttman, D. S. Type III effector
1086 diversification via both pathoadaptation and horizontal transfer in response to a
1087 coevolutionary arms race. *Plos Genetics* **2**, 2131-2142,
1088 doi:10.1371/journal.pgen.0020209 (2006).
- 1089 24 Rohmer, L., Guttman, D. S. & Dangl, J. L. Diverse evolutionary mechanisms shape the
1090 type III effector virulence factor repertoire in the plant pathogen *Pseudomonas*
1091 *syringae*. *Genetics* **167**, 1341-1360, doi:10.1534/genetics.103.019638 (2004).
- 1092 25 Kim, D. K. *et al.* A proteome-scale map of the SARS-CoV-2-human contactome. *Nat*
1093 *Biotechnol*, doi:10.1038/s41587-022-01475-z (2022).
- 1094 26 Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat*
1095 *Methods* **6**, 83-90, doi:10.1038/nmeth.1280 (2009).
- 1096 27 Braun, P. Interactome mapping for analysis of complex phenotypes: insights from
1097 benchmarking binary interaction assays. *Proteomics* **12**, 1499-1518,
1098 doi:10.1002/pmic.201100598 (2012).
- 1099 28 Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein
1100 interactions. *Nat Methods* **6**, 91-97 (2009).
- 1101 29 Wessling, R. *et al.* Convergent targeting of a common host protein-network by
1102 pathogen effectors from three kingdoms of life. *Cell host & microbe* **16**, 364-375,
1103 doi:10.1016/j.chom.2014.08.004 (2014).
- 1104 30 Mukhtar, M. S. *et al.* Independently evolved virulence effectors converge onto hubs in
1105 a plant immune system network. *Science* **333**, 596-601, doi:10.1126/science.1203659
1106 (2011).
- 1107 31 Osborne, R. *et al.* Symbiont-host interactome mapping reveals effector-targeted
1108 modulation of hormone networks and activation of growth promotion. *Nat Commun* **14**,
1109 4065, doi:10.1038/s41467-023-39885-5 (2023).
- 1110 32 Choteau, S. A. *et al.* mimicINT: a workflow for microbe-host protein interaction
1111 inference. *bioRxiv*, 2022.2011.2004.515250, doi:10.1101/2022.11.04.515250 (2022).
- 1112 33 Gutierrez-Gonzalez, L. H. *et al.* Peptide Targeting of PDZ-Dependent Interactions as
1113 Pharmacological Intervention in Immune-Related Diseases. *Molecules* **26**,
1114 doi:10.3390/molecules26216367 (2021).
- 1115 34 Gogl, G. *et al.* Quantitative fragmentomics allow affinity mapping of interactomes. *Nat*
1116 *Commun* **13**, 5472, doi:10.1038/s41467-022-33018-0 (2022).

- 1117 35 Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-
1118 throughput holdup assay. *Nat Methods* **12**, 787-793, doi:10.1038/nmeth.3438 (2015).
- 1119 36 Javier, R. T. & Rice, A. P. Emerging Theme: Cellular PDZ Proteins as Common Targets
1120 of Pathogenic Viruses. *J Virol* **85**, 11544-11556, doi:10.1128/jvi.05410-11 (2011).
- 1121 37 Zhang, H. *et al.* Pathogenesis and Mechanism of Gastrointestinal Infection With
1122 COVID-19. *Front Immunol* **12**, 674074, doi:10.3389/fimmu.2021.674074 (2021).
- 1123 38 Damin, D. C., Ziegelmann, P. K. & Damin, A. P. Human papillomavirus infection and
1124 colorectal cancer risk: a meta-analysis. *Colorectal Dis* **15**, e420-428,
1125 doi:10.1111/codi.12257 (2013).
- 1126 39 Karst, S. M. & Tibbetts, S. A. Recent Advances in Understanding Norovirus
1127 Pathogenesis. *Journal of Medical Virology* **88**, 1837-1843, doi:10.1002/jmv.24559
1128 (2016).
- 1129 40 Maseko, S. B. *et al.* Identification of small molecule antivirals against HTLV-1 by
1130 targeting the hDLG1-Tax-1 protein-protein interaction. *Antiviral Res*, 105675,
1131 doi:10.1016/j.antiviral.2023.105675 (2023).
- 1132 41 Gonzalez, R. & Elena, S. F. The Interplay between the Host Microbiome and
1133 Pathogenic Viral Infections. *Mbio* **12**, doi:10.1128/mBio.02496-21 (2021).
- 1134 42 Girardin, S. E. *et al.* Nod2 is a general sensor of peptidoglycan through muramyl
1135 dipeptide (MDP) detection. *Journal of Biological Chemistry* **278**, 8869-8872,
1136 doi:10.1074/jbc.C200651200 (2003).
- 1137 43 Knights, D. *et al.* Complex host genetics influence the microbiome in inflammatory
1138 bowel disease. *Genome Med* **6**, 107, doi:10.1186/s13073-014-0107-1 (2014).
- 1139 44 Brennan, J. J. & Gilmore, T. D. Evolutionary Origins of Toll-like Receptor Signaling.
1140 *Mol. Biol. Evol.* **35**, 1576-1587, doi:10.1093/molbev/msy050 (2018).
- 1141 45 D'Alessio, S. *et al.* Revisiting fibrosis in inflammatory bowel disease: the gut thickens.
1142 *Nature Reviews Gastroenterology & Hepatology* **19**, 169-184, doi:10.1038/s41575-
1143 021-00543-0 (2022).
- 1144 46 Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human
1145 metabolism. *Nat Biotechnol* **36**, 272-281, doi:10.1038/nbt.4072 (2018).
- 1146 47 Vidal, M., Cusick, M. E. & Barabasi, A.-L. Interactome Networks and Human Disease.
1147 *Cell* **144**, 986-998, doi:10.1016/j.cell.2011.02.016 (2011).
- 1148 48 Gulbahce, N. *et al.* Viral Perturbations of Host Networks Reflect Disease Etiology. *Plos*
1149 *Comp. Biol.* **8**, doi:e1002531
10.1371/journal.pcbi.1002531 (2012).
- 1151 49 Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification
1152 and prioritisation. *Nucleic Acids Research* **49**, D1302-D1310,
1153 doi:10.1093/nar/gkaa1027 (2021).
- 1154 50 Bunker, J. J. *et al.* Natural polyreactive IgA antibodies coat the intestinal microbiota.
1155 *Science* **358**, eaan6619, doi:doi:10.1126/science.aan6619 (2017).
- 1156 51 Pabst, O. & Slack, E. IgA and the intestinal microbiota: the importance of being specific.
1157 *Mucosal Immunology* **13**, 12-21, doi:10.1038/s41385-019-0227-4 (2020).
- 1158 52 Alcazar, C. G. *et al.* The association between early-life gut microbiota and childhood
1159 respiratory diseases: a systematic review. *Lancet Microbe* **3**, e867-e880,
1160 doi:10.1016/S2666-5247(22)00184-7 (2022).

- 1161 53 Mahmud, M. R. *et al.* Impact of gut microbiome on skin health: gut-skin axis observed
1162 through the lenses of therapeutics and skin diseases. *Gut Microbes* **14**, 2096995,
1163 doi:10.1080/19490976.2022.2096995 (2022).
- 1164 54 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**,
1165 402-408, doi:10.1038/s41586-020-2188-x (2020).
- 1166 55 Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease.
1167 *New England Journal of Medicine* **375**, 2369-2379, doi:10.1056/NEJMra1600266
1168 (2016).
- 1169 56 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of
1170 multiple integrin genes in inflammatory bowel disease. *Nature Genetics* **49**, 256-261,
1171 doi:10.1038/ng.3760 (2017).
- 1172 57 Vainer, B., Nielsen, O. H. & Horn, T. Comparative studies of the colonic in situ
1173 expression of intercellular adhesion molecules (ICAM-1, -2, and -3), beta2 integrins
1174 (LFA-1, Mac-1, and p150,95), and PECAM-1 in ulcerative colitis and Crohn's disease.
1175 *Am J Surg Pathol* **24**, 1115-1124, doi:10.1097/00000478-200008000-00009 (2000).
- 1176 58 Vainer, B. Intercellular adhesion molecule-1 (ICAM-1) in ulcerative colitis: presence,
1177 visualization, and significance. *Inflamm Res* **54**, 313-327, doi:10.1007/s00011-005-
1178 1363-8 (2005).
- 1179 59 Biswas, S., Bryant, R. V. & Travis, S. Interfering with leukocyte trafficking in Crohn's
1180 disease. *Best Practice & Research Clinical Gastroenterology* **38-39**, 101617,
1181 doi:<https://doi.org/10.1016/j.bpg.2019.05.004> (2019).
- 1182 60 Brand, S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new
1183 immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's
1184 disease. *Gut* **58**, 1152-1167, doi:10.1136/gut.2008.163667 (2009).
- 1185 61 Zhao, J. *et al.* Th17 Cells in Inflammatory Bowel Disease: Cytokines, Plasticity, and
1186 Therapies. *J Immunol Res* **2021**, 8816041, doi:10.1155/2021/8816041 (2021).
- 1187 62 Mitsuyama, K. *et al.* IL-8 as an important chemoattractant for neutrophils in ulcerative
1188 colitis and Crohn's disease. *Clin Exp Immunol* **96**, 432-436, doi:10.1111/j.1365-
1189 2249.1994.tb06047.x (1994).
- 1190 63 Herrero-Cervera, A., Soehnlein, O. & Kenne, E. Neutrophils in chronic inflammatory
1191 diseases. *Cell Mol Immunol* **19**, 177-191, doi:10.1038/s41423-021-00832-3 (2022).
- 1192 64 Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel
1193 diseases. *Nature* **569**, 655-+, doi:10.1038/s41586-019-1237-9 (2019).
- 1194 65 Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory
1195 bowel disease. *Nature Microbiology* **4**, 293-305, doi:10.1038/s41564-018-0306-4
1196 (2019).
- 1197 66 Kelly, D., Conway, S. & Aminov, R. Commensal gut bacteria: mechanisms of immune
1198 modulation. *Trends Immunol* **26**, 326-333, doi:<https://doi.org/10.1016/j.it.2005.04.008>
1199 (2005).
- 1200 67 Büttner, D. Protein export according to schedule: architecture, assembly, and
1201 regulation of type III secretion systems from plant- and animal-pathogenic bacteria.
1202 *Microbiol Mol Biol Rev* **76**, 262-310, doi:10.1128/mmbr.05017-11 (2012).
- 1203 68 Rodriguez, P. A. *et al.* Systems Biology of Plant-Microbiome Interactions. *Mol. Plant.*
1204 **12**, 804-821, doi:10.1016/j.molp.2019.05.006 (2019).
- 1205 69 Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility
1206 to Crohn's disease. *Nature* **411**, 599-603, doi:10.1038/35079107 (2001).

- 1207 70 Lichtenstein, G. R. *et al.* ACG Clinical Guideline: Management of Crohn's Disease in
1208 Adults. *American Journal of Gastroenterology* **113**, 481-517, doi:10.1038/ajg.2018.27
1209 (2018).
- 1210 71 Wera, O., Lancellotti, P. & Oury, C. The Dual Role of Neutrophils in Inflammatory Bowel
1211 Diseases. *Journal of Clinical Medicine* **5**, doi:10.3390/jcm5120118 (2016).
- 1212 72 Basnet, S., Palmenberg, A. C. & Gern, J. E. Rhinoviruses and Their Receptors. *Chest*
1213 **155**, 1018-1025, doi:<https://doi.org/10.1016/j.chest.2018.12.012> (2019).
- 1214 73 Hayashi, Y. *et al.* Rhinovirus Infection and Virus-Induced Asthma. *Viruses* **14**,
1215 doi:10.3390/v14122616 (2022).
- 1216 74 Zhang, Y. *et al.* The ORMDL3 Asthma Gene Regulates ICAM1 and Has Multiple Effects
1217 on Cellular Inflammation. *Am J Respir Crit Care Med* **199**, 478-488,
1218 doi:10.1164/rccm.201803-0438OC (2019).
- 1219 75 Reimer, L. C. *et al.* BacDive in 2019: bacterial phenotypic data for High-throughput
1220 biodiversity analysis. *Nucleic Acids Res* **47**, D631-D636, doi:10.1093/nar/gky879
1221 (2019).
- 1222 76 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
1223 assessing the quality of microbial genomes recovered from isolates, single cells, and
1224 metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
- 1225 77 Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved
1226 functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids*
1227 *Res* **44**, D286-293, doi:10.1093/nar/gkv1248 (2016).
- 1228 78 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
1229 generation sequencing data. *Bioinformatics* **28**, 3150-3152,
1230 doi:10.1093/bioinformatics/bts565 (2012).
- 1231 79 Wang, J. *et al.* BastionHub: a universal platform for integrating and analyzing
1232 substrates secreted by Gram-negative bacteria. *Nucleic Acids Res* **49**, D651-D659,
1233 doi:10.1093/nar/gkaa899 (2021).
- 1234 80 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**,
1235 421, doi:10.1186/1471-2105-10-421 (2009).
- 1236 81 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to
1237 classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-
1238 1927, doi:10.1093/bioinformatics/btz848 (2019).
- 1239 82 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new
1240 developments. *Nucleic Acids Res* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).
- 1241 83 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High
1242 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
1243 *Nat Commun* **9**, 5114, doi:10.1038/s41467-018-07641-9 (2018).
- 1244 84 Altmann, M., Altmann, S., Falter, C. & Falter-Braun, P. High-Quality Yeast-2-Hybrid
1245 Interaction Network Mapping. *Curr Protoc Plant Biol* **3**, e20067,
1246 doi:10.1002/cppb.20067 (2018).
- 1247 85 Altmann, M. *et al.* Extensive signal integration by the phytohormone protein network.
1248 *Nature* **583**, 271-276, doi:10.1038/s41586-020-2460-0 (2020).
- 1249 86 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**,
1250 402-408, doi:10.1038/s41586-020-2188-x (2020).

- 1251 87 Orchard, S. *et al.* Protein interaction data curation: the International Molecular
1252 Exchange (IMEx) consortium. *Nature Methods* **9**, 345-350, doi:10.1038/nmeth.1931
1253 (2012).
- 1254 88 del-Toro, N. *et al.* A new reference implementation of the PSICQUIC web service.
1255 *Nucleic Acids Res* **41**, W601-606, doi:10.1093/nar/gkt392 (2013).
- 1256 89 Choi, S. G. *et al.* Maximizing binary interactome mapping with a minimal number of
1257 assays. *Nat. Commun.* **10**, 3907, doi:10.1038/s41467-019-11809-2 (2019).
- 1258 90 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
1259 *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).
- 1260 91 Blum, M. *et al.* The InterPro protein families and domains database: 20 years on.
1261 *Nucleic Acids Res* **49**, D344-D354, doi:10.1093/nar/gkaa977 (2021).
- 1262 92 Edwards, R. J., Paulsen, K., Aguilar Gomez, C. M. & Perez-Bercoff, A. Computational
1263 Prediction of Disordered Protein Motifs Using SLiMSuite. *Methods Mol Biol* **2141**, 37-
1264 72, doi:10.1007/978-1-0716-0524-0_3 (2020).
- 1265 93 Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res*
1266 **48**, D296-D306, doi:10.1093/nar/gkz1030 (2020).
- 1267 94 Dosztanyi, Z. Prediction of protein disorder based on IUPred. *Protein Sci* **27**, 331-340,
1268 doi:10.1002/pro.3334 (2018).
- 1269 95 Mosca, R., Ceol, A., Stein, A., Olivella, R. & Aloy, P. 3did: a catalog of domain-based
1270 interactions of known three-dimensional structure. *Nucleic Acids Res* **42**, D374-379,
1271 doi:10.1093/nar/gkt887 (2014).
- 1272 96 Gfeller, D. *et al.* The multiple-specificity landscape of modular peptide recognition
1273 domains. *Mol Syst Biol* **7**, 484, doi:10.1038/msb.2011.18 (2011).
- 1274 97 Davey, N. E. *et al.* Attributes of short linear motifs. *Mol Biosyst* **8**, 268-281,
1275 doi:10.1039/c1mb05231d (2012).
- 1276 98 Davey, N. E. *et al.* SLiMPrints: conservation-based discovery of functional motif
1277 fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res* **40**, 10628-
1278 10641, doi:10.1093/nar/gks854 (2012).
- 1279 99 Hagai, T., Azia, A., Babu, M. M. & Andino, R. Use of host-like peptide motifs in viral
1280 proteins is a prevalent strategy in host-virus interactions. *Cell reports* **7**, 1729-1739,
1281 doi:10.1016/j.celrep.2014.04.052 (2014).
- 1282 100 Duhoo, Y. *et al.* High-Throughput Production of a New Library of Human Single and
1283 Tandem PDZ Domains Allows Quantitative PDZ-Peptide Interaction Screening
1284 Through High-Throughput Holdup Assay. *Methods Mol Biol* **2025**, 439-476,
1285 doi:10.1007/978-1-4939-9624-7_21 (2019).
- 1286 101 Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package
1287 for gene list functional enrichment analysis and namespace conversion toolset
1288 g:Profiler. *F1000Res* **9**, doi:10.12688/f1000research.24956.2 (2020).
- 1289 102 Gene Ontology, C. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**,
1290 doi:10.1093/genetics/iyad031 (2023).
- 1291 103 Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG
1292 for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* **51**, D587-
1293 D592, doi:10.1093/nar/gkac963 (2023).
- 1294 104 Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**,
1295 D687-D692, doi:10.1093/nar/gkab1028 (2022).

- 1296 105 Valdeolivas, A. *et al.* Random walk with restart on multiplex and heterogeneous
1297 biological networks. *Bioinformatics* **35**, 497-505, doi:10.1093/bioinformatics/bty637
1298 (2019).
- 1299 106 Biran, H., Kupiec, M. & Sharan, R. Comparative Analysis of Normalization Methods for
1300 Network Propagation. *Front Genet* **10**, 4, doi:10.3389/fgene.2019.00004 (2019).
- 1301 107 Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and
1302 genes at all published human GWAS trait-associated loci. *Nat Genet* **53**, 1527-1533,
1303 doi:10.1038/s41588-021-00945-5 (2021).
- 1304 108 Barrio-Hernandez, I. *et al.* Network expansion of genetic associations defines a
1305 pleiotropy map of human cell biology. *Nat Genet* **55**, 389-398, doi:10.1038/s41588-
1306 023-01327-9 (2023).
- 1307 109 Oh, J. *et al.* Biogeography and individuality shape function in the human skin
1308 metagenome. *Nature* **514**, 59-64, doi:10.1038/nature13786 (2014).
- 1309 110 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1310 identification. *BMC Bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).
- 1311 111 Del Toro, N. *et al.* The IntAct database: efficient access to fine-grained molecular
1312 interaction data. *Nucleic Acids Res* **50**, D648-D653, doi:10.1093/nar/gkab1006 (2022).
- 1313