

1 Interspecies variation in hominid gut microbiota controls host gene regulation

2 Amanda L. Muehlbauer^{1,2}, Allison L. Richards³, Adnan Alazizi³, Michael Burns⁴, Andres
3 Gomez⁵, Jonathan B. Clayton^{6,7}, Klara Petrzekova^{8,9,10}, Camilla Cascardo³, Justyna Resztak³,
4 Xiaoquan Wen¹¹, Roger Pique-Regi^{3,12}, Francesca Luca^{3,12,*}, and Ran Blekhman^{1,2,*}

5 *Corresponding authors (blekhman@umn.edu, fluca@wayne.edu)

- 6 1. Department of Genetics, Cell Biology and Development, University of Minnesota,
7 Minneapolis, Minnesota, USA
- 8 2. Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis,
9 Minnesota, USA
- 10 3. Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan
11 48201, USA
- 12 4. Department of Biology, Loyola University, Chicago, Illinois, 60660, USA
- 13 5. Department of Animal Science, University of Minnesota, Saint Paul, Minnesota, USA
- 14 6. Department of Biology, University of Nebraska at Omaha, Omaha, Nebraska, USA
- 15 7. Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln,
16 Nebraska, USA
- 17 8. The Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic
- 18 9. Liberec Zoo, Liberec, Czech Republic
- 19 10. The Czech Academy of Sciences, Institute of Parasitology, Ceske Budejovice, Czech
20 Republic
- 21 11. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
- 22 12. Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan
23 48201, USA

24 Abstract

25 The gut microbiome exhibits extreme compositional variation between hominid hosts.
26 However, it is unclear how this variation impacts host physiology, and whether this effect can be
27 mediated through microbial regulation of host gene expression in interacting epithelial cells.
28 Here, we characterized the transcriptional response of colonic epithelial cells *in vitro* to live
29 microbial communities extracted from humans, chimpanzees, gorillas, and orangutans. We found
30 most host genes exhibit a conserved response, whereby they respond similarly to the four
31 hominid microbiomes, while some genes respond only to microbiomes from specific host
32 species. Genes that exhibit such a divergent response are associated with relevant intestinal
33 diseases in humans, such as inflammatory bowel disease and Crohn's disease. Lastly, we found
34 that inflammation-associated microbial species regulate the expression of host genes previously

35 associated with inflammatory bowel disease, suggesting health-related consequences for
36 species-specific host-microbiome interactions across hominids.

37 **Introduction**

38 The microbiome of the primate gastrointestinal tract plays an important role in host
39 physiology and health. Extreme variation in the gut microbiome has been observed between
40 healthy human individuals; this variation is even more pronounced between different species of
41 great apes (Human Microbiome Project Consortium, 2012; Nishida & Ochman, 2019).
42 Microbiome composition is strongly correlated with the species of the host, a pattern known as
43 co-diversification. Within hominids and other nonhuman primates, co-diversification between
44 host and microbial symbionts has led to overall microbiome composition clustering along the
45 expected phylogenetic relationships of the host species, including bacterial, archeal and
46 eukaryotic groups within the gut microbiome (Amato, et al., 2019; Mann et al., 2019; Moeller et
47 al., 2012; Ochman et al., 2010; Raymann et al., 2017). However, reports show that these
48 phylogenetic constraints are flexible, depending on diet and subsistence strategy (Gomez et al.,
49 2019). For example, compared with industrialized human groups, small scale rural or agricultural
50 human populations share a greater number of gut microbiome traits with wild nonhuman
51 primates (Amato, et al., 2019; Gomez et al., 2019).

52 Different hominid species harbor many of the same bacterial phyla in the gastrointestinal
53 tract, but in varying abundances. For example, both the human and chimpanzee guts are
54 primarily colonized by Bacteroidetes and Firmicutes, but the chimpanzee gut also harbors higher
55 abundances of microbial phyla that are relatively rare in humans, including Actinobacteria,
56 Euryarchaeota, Tenericutes, and Verrucomicrobia (Nishida & Ochman, 2019; Ochman et al.,
57 2010). Gorillas, besides also displaying presence of these rare taxa, harbor greater abundances of
58 Chloroflexi, Tenericutes, and Fibrobacteres (Gomez et al., 2015, 2016; Hicks et al., 2018).
59 Although the orangutan microbiome has not been characterized as thoroughly, a previous report
60 has shown that orangutan guts harbor higher diversity in archaeal lineages compared to other
61 great apes, in addition to similar microbial phyla as gorillas and chimpanzees (Delsuc et al.,
62 2014; Raymann et al., 2017). At lower microbial taxonomic levels, very different microbial
63 species are present in human and chimpanzee microbiomes, resulting in greater divergence
64 (Nishida & Ochman, 2019).

65 Overall gut microbiome composition is shaped by a combination of host genetics, host
66 physiology, and environmental factors. Studies have shown that host genetic variation influences
67 microbiome composition within humans, but has yet to be studied in other hominids (Blekhman
68 et al., 2015; Goodrich et al., 2014). Among environmental influences, diet has a large impact on
69 the primate gut microbiome (Gomez, et al., 2016; Hicks et al., 2018; Nagpal et al., 2018). Most
70 non-human great ape species in the wild and in captivity subsist on a primarily plant-based diet
71 of fruit and vegetation that is occasionally supplemented by animal protein, such as meat or
72 insects (Tutin & Fernandez, 1993; Vogel et al., 2015; Watts et al., 2012). In contrast, human diets
73 are usually omnivorous and highly variable depending on cultural influences, agricultural

74 practices, geographic location, and individual dietary preferences (Lang et al., 2018; Wu et al.,
75 2011). Other environmental factors that can influence microbiome composition between primates
76 include variation in geography, seasonality, and other social behaviors such as grooming
77 (Grieneisen et al., 2019; Tung et al., 2015). In addition, physiological differences between
78 primate species, such as differences in gut morphology and digestive processes, also contribute
79 to differences in microbiome composition (Amato, et al., 2019). Although a large effort has been
80 made to characterize the factors that influence variation in the microbiome, it is unclear how
81 variation in microbiome composition between great ape species can impact relevant host
82 phenotypes.

83 A likely mechanism by which the microbiome can affect host physiology is through
84 regulating the expression of host genes in interacting intestinal epithelial cells (Luca et al., 2018;
85 Richards et al., 2016, 2019). Studies in animal models have demonstrated that gut microbiota can
86 drive changes in host gene expression by altering epigenetic programming, such as histone
87 modification, transcription factor binding, and methylation (Camp et al., 2014; Krautkramer et
88 al., 2016; Pan et al., 2018; Qin et al., 2018). For example, Camp et al. found that the microbiome
89 drives the differential expression of transcription factors enriched in accessible binding sites
90 (Camp et al., 2014). In addition, Pan et al. found that the microbiome can alter DNA methylation
91 in the gut epithelial cells of mice (Pan et al., 2018). Moreover, in cell culture, inter-individual
92 variation in microbiome composition can drive differential responses in host gene expression at
93 the intestinal level (Richards et al., 2019). However, we do not know how interspecies variation
94 in the microbiome affects gene regulation in host cells. When considering the microbiota
95 variation amongst great ape species and their influences on host gene expression, *in vivo* studies
96 in experimental animal models are limited. Furthermore, *in vivo* experiments can be confounded
97 by a multitude of factors, such as differences in diet between the animal model species and the
98 primate species of interest, microbiota colonization history of the animal model, and inherent
99 differences in the genetic backgrounds between the animal model and the primate species (Luca
100 et al., 2018).

101 Here, we use an *in vitro* experimental system (Richards et al., 2016, 2019) to assess host
102 gene expression changes in response to diverse gut microbiota from four great ape species:
103 humans (*Homo sapiens*), and captive chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*
104 *gorilla*), and orangutans (*Pongo abelii*). This experimental design allows us to determine causal
105 relationships between gut microbiome composition and gene expression changes in colonic
106 epithelial cells that are induced by the microbiome while controlling for potentially confounding
107 environmental and technical effects (Richards et al., 2016, 2019). We have leveraged this design
108 to ascertain how host genes respond to between-species variation in microbiome composition
109 across hominids, characterize the function of host genes that respond to microbiota from each

110 great ape species, and identify microbial taxa and pathways that likely drive expression of
111 specific host genes.

112 **Results**

113 To assess how host genes respond to variation in the microbiome, we extracted live
114 microbiota from 19 fecal samples from four hominid species (4 humans, 3 chimpanzees, 6
115 gorillas, and 3 orangutans), and treated human colonic epithelial cells (colonocytes) with the
116 extracted microbiota using an experimental technique from a previously published method (see
117 SI Table 1) (Richards et al., 2016, 2019). We quantified changes in gene expression in the
118 colonocytes as a response to the primate microbiota using RNA-seq (**Fig. 1A**). Additionally, we
119 used 16S rRNA sequencing and shotgun metagenomics to characterize the composition of the
120 microbiome in these samples. A principal coordinate analysis of Bray-Curtis dissimilarities
121 confirmed that the microbiome samples cluster by primate host species of origin (**Fig 1B**; SI Fig.
122 1). This observation is consistent with previous findings showing that the phylogenetic
123 relationship between primate host species is reflected in their microbiomes (Ochman et al.,
124 2010), and that interspecies microbiome distinctions between wild apes is maintained in the
125 captive individuals included in our study.

126 The bacterial composition of the samples confirmed clear distinctions between hominid
127 species at the phylum level (**Fig. 1C**, SI Fig. 2), with nine of the most abundant microbial phyla
128 showing significantly different levels between hominid species (Kruskal-Wallis test,
129 Benjamini-Hochberg FDR <0.1, SI Table 2). The human microbiome samples have a high
130 relative abundance of Bacteroidetes and Firmicutes, which have both been previously identified
131 as dominant phyla in the human gut (Human Microbiome Project Consortium, 2012; Turnbaugh
132 et al., 2007). In addition, Actinobacteria abundance is significantly different between hominid
133 species (Kruskal-Wallis Test, Benjamini-Hochberg q-value = 0.00567; ANOVA,
134 Benjamini-Hochberg q-value = 3.82×10^{-9}), with chimpanzees showing the greatest abundance
135 (see **Fig. 1C**). Furthermore, we identified 21 microbial species that are differentially abundant
136 between hominid host species (SI Table 3, Kruskal Wallis, Benjamini-Hochberg FDR <0.1).
137 Examples of several microbes that have variable abundance across species, including
138 *Bacteroides ovatus*, which shows higher abundance in humans compared to other hominids;
139 *Phascolarctobacterium succinatutens*, which shows lower abundance in humans compared to
140 other hominids; and *Prevotella copri*, which has higher abundance in gorilla and orangutan, are
141 shown in **Fig. 1D**.

142 To characterize the host response to the microbiome, we used likelihood ratio tests
143 combined with a negative binomial model (DESeq2) to identify host genes that change their

144 expression after inoculation with microbiomes from the four hominid host species (see
145 Methods). We identified 4,329 host genes that respond to the microbiome of at least one hominid
146 species (**Fig. 2A**, Benjamini-Hochberg FDR<0.1). The majority of differentially expressed genes
147 (2,261 genes, 52%) respond to the microbiomes of all four hominids (see **Fig. 2A, 2B**; full
148 dataset available in SI Table 4; see Methods). Despite this overall consistent response, we find
149 164 host genes that respond in a species-specific manner; namely, respond to the microbiome of
150 one hominid species but not the other three. For example, *SHROOM3* responds to the human
151 microbiome, but shows no response to the chimpanzee, gorilla, and orangutan microbiomes (**Fig.**
152 **2C**). Similarly, *B3GAT2*, *DUSP11*, and *DARS2* respond in a species-specific manner to the
153 chimpanzee, gorilla, and orangutan microbiomes, respectively (**Fig. 2C**). We also find 394 host
154 genes that respond to microbiomes from two hominid species; e.g., *CBRI* responds to orangutan
155 and gorilla microbiomes (**Fig. 2C**). Likewise, 1,313 host genes respond to microbiomes from
156 three hominid species, and 13,531 genes show no response to any of the hominid microbiomes
157 (e.g., *INVS*; see **Fig. 2C**).

158 To understand how genes with a host species-specific response may interact with each
159 other, we visualized interaction networks for differentially expressed host genes that respond to
160 microbiomes from each hominid species (Krämer et al., 2014)(Ingenuity Pathway Analysis,
161 <http://www.ingenuity.com>; **Fig. 3A, 3B**, SI Fig. 3, SI Fig. 4; see Methods). The most significant
162 interaction network of host genes that respond only to human microbiomes is enriched with
163 functional categories related to cancer, cell death and survival, and organismal injuries and
164 abnormalities (**Fig. 3A**, SI Table 5). This is consistent with previous studies showing that the
165 microbiome may influence host disease through changes in host gene regulation, but also
166 suggests that this effect may be specific to human microbiomes (Camp et al., 2014; Krautkramer
167 et al., 2016; Pan et al., 2018; Qin et al., 2018). By comparison, the most significant interaction
168 network of genes that respond specifically to orangutan microbiomes is enriched for functional
169 categories related to carbohydrate metabolism, lipid metabolism, and small molecule
170 biochemistry (**Fig. 3B**, SI Table 6). This is consistent with the observation that orangutan diets,
171 compared to that of gorillas or chimpanzees, could incorporate a greater proportion of ripe fruits
172 and highly digestible/simple sugars in peak seasons (up to 100% dependence on fruit) (Remis,
173 1997; Taylor, 2006). In addition, previous reports point to a highly diverse archeal community in
174 orangutans compared to other apes, which could be associated with an increased capacity to
175 metabolize highly fermentable plant materials (Raymann et al., 2017). For functions enriched in
176 the most significant networks for genes that respond only to gorilla microbiomes and only to
177 chimpanzee microbiomes, see SI Table 7 and SI Table 8.

178 To further characterize the biological functions represented by host genes that respond to
179 variation in hominid microbiomes, we categorized differentially expressed genes into two
180 groups: low-divergence genes, which show a similar magnitude and direction of response to the

181 four hominid microbiomes, and high-divergence genes, which show a highly variable response to
182 the four hominid microbiomes (following the approach of Hagai and colleagues (Hagai et al.,
183 2018); see Methods and SI Table 9). We find that low-divergence genes, namely, differentially
184 expressed genes that show a similar response to microbiomes from all four primate species, tend
185 to be enriched for functions related to basic cell processes, such as RNA processing, cell cycle,
186 and RNA metabolic processing (**Fig. 3C**, Benjamini-Hochberg FDR<0.1, SI Tables 10-15). This
187 suggests that these genes are likely involved in basic host responses to bacterial cells, rather than
188 response to specific microbial features. Interestingly, high-divergence genes, namely, genes that
189 respond differently to the microbiomes from the four primate host species, tend to be enriched
190 for categories related to disease, inflammation, and cancer (**Fig. 3C**, SI Fig. 5, SI Fig. 6). Of
191 note, colorectal cancer, rheumatoid arthritis, and *Salmonella* infection functional categories are
192 enriched among high-divergence genes, and have all been associated with gut microbiome
193 composition in previous studies (Dahmus et al., 2018; Ferreira et al., 2011; Scher & Abramson,
194 2011). Moreover, when considering host genes that have been previously associated with
195 complex human traits through genome wide association studies (GWAS) using data in the
196 GWAS Catalog (Buniello et al., 2019), we find that high-divergence genes are enriched with
197 traits and diseases that have also been linked to the microbiome, such as Crohn's Disease,
198 Inflammatory Bowel Disease, and body mass index (**Fig. 3D**; see Methods). This might indicate
199 that these complex disease phenotypes may be modulated by differences in composition of the
200 gut microbial community through the regulation of these key host genes.

201 Next, we sought to identify genes whose response is directly correlated with the
202 abundance of specific microbial taxa. To do so, we used mixed linear models that integrated host
203 response transcriptomic data (via RNA-seq) and microbial species abundance information data
204 (via shotgun metagenomics; see Methods). We identified 25 microbial species that drive the
205 expression of 80 differentially expressed host genes across the four hominids (**Fig. 4A** and SI
206 Table 16, 162 host gene-microbial taxon pairs in total, Benjamini-Hochberg FDR<0.05). A
207 heatmap of the interactions reveals two roughly defined major clusters, one of which includes a
208 subcluster of host genes that are downregulated by microbial taxa that are rare or absent in
209 humans but present in the other hominids, such as *Prevotella copri*, *Methanobrevibacter*
210 (unclassified), and *Phascolarctobacterium succinatutens* (highlighted in **Fig. 4A**; also see **Fig.**
211 **1D** for *P. copri* and *P. succinatutens* abundances across hominids). Genes that are downregulated
212 in the presence of these microbial species are significantly enriched for several immune-related
213 pathways, such as cytokine activity, IL-7 signaling, malaria, Legionellosis, and TNF signaling
214 (SI Tables 17-20). Using a similar method, we identified 89 microbial pathways that drive the
215 expression of 310 unique host genes for a total of 2061 significant microbial pathway-host gene
216 interactions (Benjamini-Hochberg FDR<0.05). For simplicity, we focused on the top 48
217 microbial pathways that drive the expression of the top 44 unique host genes (**Fig. 4B** and SI
218 Table 21), with a total of 216 microbial pathway-host genes pairs (examples of specific

219 interactions can be found in SI Fig. 7). Clustering of this interaction data revealed three main
220 clusters (I, II, III), with genes in cluster II associated with pathways that are more abundant in
221 humans compared to other hominid microbiomes. These host genes are enriched in functional
222 categories related to inflammation and infectious disease, including Legionellosis, malaria, and
223 pertussis, and overlap with genes found in the cluster described above in the species-level
224 analysis (SI Table 22).

225 To investigate specific host gene-microbe interactions, we considered the network of all
226 high-divergence host genes for which expression is driven by microbial species (28 host genes
227 and 14 microbial species; Benjamini-Hochberg FDR<0.01). We find that certain microbial taxa
228 are represented in highly connected nodes and likely control the regulation of several
229 high-divergence host genes (**Fig. 4C**). For example, two *Bacteroides* species, *B. ovatus* and *B.*
230 *uniformis*, drive the expression of several host genes, including *LIF* and *DUSP5* respectively,
231 both of which have been previously associated with inflammation (Habibian et al., 2017; Yue et
232 al., 2015). *Bacteroides* is a highly abundant microbial genus in the human gut and is known to
233 have mixed effects on human health (Wexler, 2007). Notably, *B. ovatus* is highly abundant in the
234 human microbiome samples, but is at low abundances in the orangutan gut microbiomes and
235 entirely absent in the chimpanzee and gorilla microbiomes (**Fig. 1D**, FDR<0.1).

236 To explore the possible phenotypic consequences of host genes for which expression is
237 driven by certain microbial species, we integrated gene-trait associations identified through
238 transcription-wide association study (TWAS). TWAS identifies associations between gene
239 expression and complex traits by considering genetically predicted gene expression from
240 expression quantitative trait locus (eQTL) studies and SNP-trait associations from GWAS. We
241 considered genes implicated in 114 complex traits through Probabilistic TWAS (Zhang et al.,
242 2019), and found that expression of 44 out of 57 high-divergence host genes is associated with
243 43 complex phenotypes (**Fig 4D**). These include diseases and phenotypes previously linked to
244 the gut microbiome, including Crohn's disease, inflammatory bowel disease, ulcerative colitis,
245 body mass index, body fat percentage, and schizophrenia (**Fig. 4D**, SI Table 23). We found
246 several microbial taxa that have higher abundance in the non-human microbiomes, including *P.*
247 *copri*, and *P. succinatutens*, which have previously been hypothesized to have protective effects,
248 downregulate the expression of host gene *LIF*, which has been linked to ulcerative colitis,
249 inflammatory bowel disease, and Crohn's disease in our TWAS analysis (Fig. 4C, Fig.4D) (De
250 Vadder et al., 2016; Morgan et al., 2012). These results are consistent with findings from the
251 enrichment analysis reported in **Fig. 3C** and **Fig. 3D**, where we found phenotypes related to
252 inflammation were driven by high-divergence genes. Furthermore, we find that *Eubacterium*
253 *rectale* and *B. ovatus*, microbes that have higher abundance in humans and that have been
254 previously associated with inflammatory bowel disease (Noor et al., 2010; Zhang et al., 2017),

255 upregulate the expression of *CSF3*, which has been reported as upregulated in ulcerative colitis
256 patients (de Lange & Barrett, 2015; Hotte et al., 2012)

257 To investigate specific host gene-pathway interactions, we constructed a network of the
258 most significant interactions between microbial pathways and high-divergence host genes as
259 described above (**Fig 4E**; see Methods). We find that nine of these 17 host genes, including
260 *DUSP5*, *CYR61*, *NFKBIZ*, *PTGS2*, *IL6*, *CXCL8*, *IL36G*, *IL1B*, and *IL36RN* (all displayed at the
261 top layer in **Fig. 4E**) have been implicated in immune function or inflammation (Cox et al.,
262 2004; Emre & Imhof, 2014; Gales et al., 2013; Habibian et al., 2017; Hörber et al., 2016; Müller
263 et al., 2018; Onoufriadis et al., 2011; Ren & Torres, 2009; Rincon, 2012; Wang et al., 2017). We
264 found that these genes are associated with several microbial pathways, including
265 Phosphopanthothenate biosynthesis I, chorismate biosynthesis, UDP-N-acetylmuramoyl
266 pentapeptide biosynthesis II (lysine-containing), and UDP-N-acetylmuramoyl pentapeptide
267 biosynthesis I (meso-diaminopimelate containing).

268 Discussion

269 Interactions between hominid hosts and their microbiomes have been an underexplored
270 area of research, and the complexity of the host-microbiome relationship makes identifying the
271 specific microbial features that causally impact the host phenotype inherently challenging. Here,
272 we use an *in vitro* model to assess how gut microbiomes from different host species impact gene
273 regulation, which is a likely mechanism for microbes to drive changes in host phenotype and
274 health. Inoculating host colonic epithelial cells with live gut microbiome communities from four
275 great ape species, we find that most host genes are regulated similarly by microbiomes from all
276 four hominid microbiomes. However, some host genes are regulated only by microbiomes from a
277 single hominid; these genes are enriched with immunity functions and are involved in the
278 development of inflammatory bowel disease.

279 Chimpanzees, gorillas, and orangutans are our closest extant relatives, making these
280 species an important study system for understanding human evolution as well as the genetic and
281 environmental etiology of human-specific diseases. Distinct physiological, cognitive, and
282 behavioral differences between primate species are hypothesized to be the result of changes in
283 host gene regulation (Britten & Davidson, 1971; Enard et al., 2002; Gilad et al., 2006; King &
284 Wilson, 1975). Indeed, studies have identified genes showing a species-specific expression
285 pattern, and genes for which regulation likely evolves under natural selection (Blekhman et al.,
286 2008; Brawand et al., 2011). Here, we show that microbiomes of different hominid species elicit
287 different gene expression responses in the same type of intestinal epithelial cells (human
288 colonocytes). Although we show that most host genes respond to microbiomes from different
289 hominids in a similar manner, we also identified genes that exhibit a species-specific response.
290 Thus, it may be tempting to hypothesize that some of the species-specific differences in gene

291 expression observed previously are driven by interactions with the gut microbiome. These
292 species-specific microbiome-regulated host genes might facilitate host-specific adaptations to
293 physiological or dietary constraints; for example, our analysis indicates that genes with a
294 response to only orangutan microbiomes are enriched for carbohydrate metabolism, lipid
295 metabolism, and small molecule biochemistry, which suggests that the interaction of the
296 orangutan microbiome and colonic epithelial cells may aid in digestion of specific
297 macronutrients, especially those associated with diets rich in high-energy, highly digestible plant
298 sources (e.g. ripe fruit).

299 In addition to environmental adaptations, species-specific responses to the microbiota
300 may indicate tightly controlled symbiotic relationships that may result in disease phenotypes
301 when altered. We find that high-divergence genes – namely, genes that respond discordantly to
302 microbiomes from different hominid species – are enriched for traits associated with disease,
303 such as inflammation and aberrant apoptosis. This suggests that genes with a response highly
304 sensitive to the variation across hominid microbiomes may possibly play a role in host disease
305 traits. These genes are also significantly associated with relevant disease traits in the GWAS
306 catalog and in our TWAS analysis, including Crohn’s Disease (CD) and Inflammatory Bowel
307 Disease (IBD). Significant distinctions exist in gut microbiome composition and diversity across
308 apes with marked differences in subsistence strategies: for instance, industrialized human
309 societies and primates in captivity have lower gut microbiome diversity and show higher
310 incidences of noncommunicable diseases than small-scale human populations and wild
311 non-human primates, respectively (Clayton et al., 2016; Gomez, et al., 2016). Thus, one
312 hypothesis is that these unique features of the microbiome are causal for the development of
313 diseases common in humans living in industrialized areas, but not in non-industrialized human
314 populations or in non-human wild primates, such as IBD. Our results are consistent with this
315 hypothesis, and further suggest that a mechanism by which the microbiome can affect disease
316 risk is through regulating the expression of host genes in interacting colonic epithelial cells. For
317 example, we found that several microbes that have lower abundance in humans compared to the
318 other hominids, including *P. copri* and *P. succinatutens*, downregulate the expression of the gene
319 *LIF*, which has been associated with IBD (SI Fig. 8). This suggests that these microbes may
320 confer a protective effect through regulation of host genes, and their absence in humans is
321 possibly detrimental. Conversely, we found that microbes that have higher abundance in humans
322 compared to the other hominids, including *B. ovatus* and *E. rectale*, upregulate the expression of
323 *CSF3*, which has been associated with inflammatory bowel disease (SI Fig. 8). This suggests that
324 these microbes may have a human-specific pathogenic effect. Moreover, some of the genes we
325 found to be regulated by the microbiome in a species-specific manner, such as *IL1B*, *IL6*, *IL36G*,
326 *IL36RN*, and *CXCL8*, have been previously implicated in IBD (Gijsbers et al., 2004; Khor et al.,
327 2011; Müller et al., 2018; Parisinos et al., 2018; Russell et al., 2016; Schulze et al., 2008), while

328 others, such as *DUSP5*, *CYR61*, *NFKBIZ*, and *PTGS2*, have roles in immune response (Cox et
329 al., 2004; Emre & Imhof, 2014; Habibian et al., 2017; Hörber et al., 2016).

330 Our ability to interpret these results in a comparative evolutionary context is limited by
331 the unavailability of colonocytes from the non-human hominids in the study. The non-human
332 hominids in the study are all captive, thus their microbiomes might not be representative of wild
333 animals (Clayton et al., 2016); however, we find that the microbiomes used in this study still
334 cluster by host species identity, and between-species variation in microbiome composition is
335 preserved. Another limitation of our analysis is that the taxonomic profiling of metagenomic
336 shotgun sequencing data relies on databases that are biased towards microbes residing in human
337 microbiomes, and might impact our ability to detect and accurately quantify certain microbes in
338 the non-human samples. Moreover, the *in vitro* approach used here represents a simplified
339 version of the complex interactions occurring at the organismal level. Nevertheless, our approach
340 allows for tightly controlled experimental conditions that can be tailored to the specific question
341 of interest, by focusing on the relevant host cell type and microbiomes, and massively reducing
342 confounding effects of cellular composition and the environment. Indeed, our approach allows
343 controlling for various factors that may affect both the microbiome and host gene regulation,
344 such as organismal-level variables (e.g., infection and hormones), host genetic variation,
345 environmental factors (e.g., host diet), and oscillations and circadian dynamics in the
346 microbiome and host gene expression.

347 In conclusion, we find that gut microbial communities from different hominids mostly
348 elicit a conserved regulatory response in host cells, whereby most host genes respond similarly to
349 hominid microbiomes. However, we also find that some host genes show a divergent response,
350 and a number of host genes respond only to microbiomes from one hominid species and not the
351 others. These genes are enriched in functional categories related to immunity and inflammation,
352 and are over-represented in pathways involved in autoinflammatory diseases, such as IBD and
353 Crohn's disease. These results represent an important step towards understanding the causal
354 relationships between variation in the gut microbiome across hominids and the regulation of
355 intestinal epithelial cells. We hope that future studies will expand on this work using organoid
356 culture and animal models to characterize the contribution of specific microbes to the
357 development of disease through regulation of host genes.

358 **Acknowledgements**

359 We would like to thank the primate zookeepers at the Ostrava Zoo, and especially Jana
360 Pluhackova, for their help with chimpanzee fecal sample collection. We would also like to thank
361 the primate zookeepers at the Como Park Zoo and Conservatory for their help with gorilla and
362 orangutan fecal sample collection. Finally, we would like to thank Blekman Lab members, and
363 especially Rich Abdill, Beth Adamovicz, Laura Grieneisen, and Sambhawa Priya, for their

364 comments and advice on the manuscript. This work is supported by NIH award R35-GM128716
365 (to R. B.) and NIH award R01-GM109215 (to F. L. and R. P.-R.). Partial funding was provided
366 by the Czech-American Scientific cooperation (LH15175) supported by the Ministry of
367 Education, Youth and Sports of the Czech Republic (to K. P.). This work was carried out, in part,
368 by resources provided by the Minnesota Supercomputing Institute.

369 **Methods**

370 **Sample acquisition and live microbiota extraction**

371 See SI Table 1 for full details about the human and non-human primate fecal samples
372 used in this analysis. Non-human fecal samples from gorillas and orangutans were collected from
373 captive animals immediately after defecation. One orangutan who donated two samples was on a
374 low dose of antibiotics for chronic colitis. Samples were collected as soon as possible (within an
375 hour of defecation) into a 50mL conical tube containing 20mL of cryoprotectant solution
376 consisting of a 50:50 mixture of glycerol and saline solution. The cryoprotectant was filter
377 sterilized through a 0.22 μ m filter. Samples were shaken vigorously to distribute the
378 cryoprotectant. Gorilla and orangutan samples were stored at -80°C within 1 hour after collection
379 and shipped to the lab on dry ice. Chimpanzee samples were stored at -20°C within 1 hour of
380 collection and then shipped to the U.S. lab on dry ice within one day. Human fecal samples were
381 purchased from OpenBiome and arrived frozen on dry ice. The following briefly describes the
382 protocol by which OpenBiome processes stool samples. The sample is collected by OpenBiome
383 and given to a technician within 1 hour of defecation. The mass of the sample is measured and
384 transferred to a sterile biosafety cabinet. The stool sample is put into a sterile filter bag, and a
385 sterile filtered dilutant of 12.5% glycerol is added with a normal saline buffer (0.90% [wt/vol]
386 NaCl in water). The sample solution is then introduced to a homogenizer blender for 60 s and
387 aliquoted into sterile bottles. The bottles are then immediately frozen at -80°C. Any sample not
388 fully processed within 2 hours of passage is destroyed.

389 To extract fecal microbiota from the non-human primate samples, inside a sterile
390 low-oxygen cabinet we placed fecal material into a sterilized disposable standard commercial
391 blender cup, added 20ml glycerol to reach approximately 30mL glycerol and 200mL normal
392 saline buffer (0.90% [wt/vol] NaCl in water). Fecal material was blended until fully
393 homogenized (about 1-2 min). Blended material was transferred to the same side of the
394 membrane in a 330-micron filter bag and the liquid suspension of the bacterial community was
395 collected on the other side of the filter. The resulting microbiota suspension was then mixed and
396 aliquoted into small tubes and stored at -80°C.

397 The research and sample collection in this study complied with protocols approved
398 through the University of Minnesota Institutional Animal Care and Use Committee.

399 **Colonocyte with hominid-derived microbiota treatment experiment**

400 The experimental protocol used for the treatment of colonocytes with microbiota has
401 previously been described in Richards et al., 2016 (Richards et al., 2016). Experiments were
402 conducted using primary human colonic epithelial cells (HCoEpiC, lot: 9763), hereby called
403 colonocytes (ScienCell Research Laboratories, Carlsbad, California, USA, 2950). The cells were
404 cultured on plates or flasks coated with poly-l-lysine (PLL), according to the supplier's
405 specifications (ScienCell 0413). Colonocytes were cultured in colonic epithelial cell medium

406 supplemented with colonic epithelial cell growth supplement and penicillin-streptomycin
407 according to the manufacturer's protocol (ScienCell 2951) at 37 °C with 5% CO₂. At 24 hours
408 before treatment, cells were changed to antibiotic-free medium and moved to an incubator at 37
409 °C, 5% CO₂, and a reduced 5% O₂.

410 Fecal microbiota were not thawed until the day of the experiment. Prior to treatment, the
411 microbiota was thawed at 30 °C, and the microbial density (OD₆₀₀) was assessed via a
412 spectrophotometer (Bio-Rad SmartSpec 3000). Medium was removed from the colonocytes and
413 fresh antibiotic-free medium was added to the cells, with a final microbial ratio of 10:1
414 microbe:colonocyte in each well. Additional wells containing only colonocytes were also
415 cultured in the same 24-well plate for use as controls.

416 After 2 hours, the wells were scraped on ice, pelleted, and washed with cold
417 phosphate-buffered saline (PBS) and then resuspended in lysis buffer (Dynabeads mRNA Direct
418 kit, ThermoFisher Scientific, Waltham, Massachusetts, USA) and stored at -80 °C until
419 extraction of colonocyte RNA for RNA-seq. We conducted both metagenomic shotgun
420 sequencing and 16s rRNA sequencing on the microbiomes at four points: before preparation
421 (raw), after preparation (prepared), cultured with colonocytes (colonocytes) and cultured without
422 colonocytes (control). Human fecal microbiome samples were purchased as "prepared" from
423 Openbiome and therefore were not sequenced raw.

424 **RNA-seq experiment and data processing**

425 Poly-adenylated mRNA was isolated from thawed cell lysates using the Dynabeads
426 mRNA Direct Kit (Ambion) following the manufacturer's instructions. RNA-seq libraries were
427 prepared using a protocol modified from the NEBNext Ultradirectional (NEB) library
428 preparation protocol to use Barcodes from BIOO Scientific added by ligation, as described in
429 Richards et al. (Richards et al., 2019). The libraries were then pooled and sequenced on two
430 lanes of the Illumina Next-seq 500 in the Luca/Pique-Regi laboratory using the high output kits
431 for 75 cycles to obtain paired-end reads. Reads were 80 bp in length. Read counts ranged
432 between 12,632,223 and 36,747,968 reads per sample, with a mean of 18,726,038 and median of
433 16,993,999 reads per sample.

434 FastQC was used to determine quality of reads from raw data (FastQC, version 0.11.5).
435 Trimmomatic was used to trim adapters. FastQC was again used to determine quality of reads
436 after trimming of adapters (Trimmomatic version 0.33). Transcripts were aligned to database
437 GRCh38 and was performed using HISAT2 (HISAT2 version 2.0.2)(Kim et al., 2019). After
438 alignment, read counts ranged between 10,817,737 and 33,592,529 aligned reads per sample,
439 with a mean of 17,142,585.72 and a median of 15,542,693.5 aligned reads per sample. Overall,
440 the average alignment rate was ~70% across samples (SI Fig. 9). The R 'Subread' package with
441 the 'featureCounts' program was used to make the transcript abundance file (R version 3.3.3,
442 Subread version 1.4.6).

443 **16s rRNA sequencing**

444 Sequencing on the 16s rRNA V4 region was performed at the University of Minnesota
445 Genomics Center using the protocol described in Gohl et. al, 2016 (Gohl et al., 2016). DNA
446 isolated from fecal samples was quantified with qPCR and the V4 region of the 16s rRNA gene
447 was amplified using PCR with barcodes for multiplexing.
448 The forward indexing primer sequence is -
449 **AATGATACGGCGACCACCGAGATCTACAC**[i5]**TCGTTCGGCAGCGTC** and the reverse
450 indexing primer sequence is -
451 **CAAGCAGAAGACGGCATAACGAGAT**[i7]**GTCTCGTGGGCTCGG** (where the bolded
452 regions are the p5 and p7 flow cell adapters and [i5] and [i7] refer to the index sequence codes
453 used by Illumina). The qPCR step starts with an initial denaturing step at 95 °C for 5 min
454 followed by 35 cycles of denaturation (20s at 98 °C), annealing (15s at 66 °C) and elongation (1
455 min at 72 °C). After qPCR, samples are normalized to 167,000 molecules/ μ l. This is based on
456 the volume of sample used for PCR1 (3ul), so 500,000 molecules is roughly 10x the target
457 sequencing coverage. The next PCR (PCR1) step is similar to the qPCR step, except with only
458 25 cycles of denaturation, annealing and elongation. After the first round of amplification, PCR1
459 products are diluted 1:100 and 5 μ l of 1:100 PCR1 is used in the second PCR reaction. The next
460 step (PCR2) is similar to the previous two PCR protocols, except with only 10 cycles of
461 denaturation, annealing and elongation. Next, Pooled samples were denatured with NaOH,
462 diluted to 8 pM in Illumina's HT1 buffer, spiked with 15% PhiX, and heat denatured at 96 °C for
463 2 minutes immediately prior to loading. A MiSeq 600 cycle v3 kit was used to sequence the
464 sample. The following Nextera adapter sequences for post-run trimming are also used. For read 1
465 -
466 CTGTCTCTTATACACATCTCCGAGCCCACGAGACNNNNNNNNATCTCGTATGCCGTCT
467 TCTGCTTG and for read 2 -
468 CTGTCTCTTATACACATCTGACGCTGCCGACGANNNNNNNNGTGTAGATCTCGGTGGT
469 CGCCGTATCATT

470 **Metagenomic shotgun sequencing**

471 Metagenomic shotgun sequencing on prepared microbiota samples was performed at the
472 University of Minnesota Genomics Center (UMGC). DNA samples were quantified using a
473 fluorimetric PicoGreen assay. For a sample to pass QC, it needs to quantify greater than 0.2
474 ng/ μ l. If the samples pass QC they enter the TruSeq NexteraXT DNA library preparation queue.
475 gDNA samples were converted to Illumina sequencing libraries using Illumina's NexteraXT
476 DNA Sample Preparation Kit (Cat. # FC-130-1005). 1 ng of gDNA is simultaneously
477 fragmented and tagged with a unique adapter sequence. This "tagmentation" step is mediated by
478 a transposase. The tagmented DNA is simultaneously indexed and amplified 12 PCR cycles.
479 Final library size distribution is validated using capillary electrophoresis and quantified using
480 fluorimetry (PicoGreen). Truseq libraries were hybridized to a NextSeq (either Single Read or

481 Paired End). Clustering occurs on-board where the bound library molecules are clonally
482 amplified and sequenced using Illumina's SBS chemistry. NextSeq uses 2-color chemistry to
483 image the clusters. Upon completion of read 1, a 7 base paired index read is performed in the
484 case of single indexed libraries. If dual indexing was used during library preparation, 2 separate
485 8 or 10 base pair index reads are performed. Finally, clustered library fragments were
486 re-synthesized in the reverse direction thus producing the template for paired end read 2. Base
487 call (.bcl) files for each cycle of sequencing are generated by Illumina Real Time Analysis (RTA)
488 software. The base call files are demultiplexed and then converted to index specific fastq files
489 using the MiSeq Reporter software on-instrument.

490 **Characterizing microbiota**

491 To identify microbial features from the metagenomic shotgun sequencing data, including
492 taxa and pathway abundances, we used the HUMAnN2 pipeline with Metaphlan2 (HUMAnN2
493 v0.11.1, Metaphlan2 v0.2.6.0)(Franzosa et al., 2018; Truong et al., 2015). FastQC v0.11.7 was
494 used to determine quality of sequencing reads before trimming. Sequencing adapters were
495 trimmed from the raw reads using Trimmomatic (Trimmomatic v0.33) (Bolger et al., 2014).
496 FastQC v0.11.7 was again used to determine quality of sequencing reads after trimming the
497 sequencing adapters from the reads (SI Fig. 10). Metaphlan2 was used to assign taxonomy at all
498 taxonomic levels to the sequencing reads in each sequencing file, and in particular to get relative
499 abundances of microbial taxa for each sample. The HUMAnN2 pipeline utilizes bowtie v0.2.2
500 for read alignment (Langmead & Salzberg, 2012), DIAMOND v0.8.22 for high throughput
501 protein alignment (Buchfink et al., 2015), MinPath (Ye & Doak, 2009) for pathway
502 reconstruction from protein family predictions. The UniRef90 database was used for determining
503 gene family abundances (Suzek et al., 2015). We found a total of 166 named microbial species
504 detected in at least one sample (SI Fig. 11).

505 **Principal Coordinate Analysis of Samples**

506 Using the 16s rRNA data from the fecal microbiota samples, we used the R package
507 'DADA2' (DADA2, version 1.2.2) to identify amplicon sequence variants (ASVs) from the
508 reads(Callahan et al., 2016). DADA2 was used to filter and trim sequences from raw reads.
509 Forward reads were trimmed to position 240 and reverse reads were trimmed to position 160.
510 Reads were truncated at the first quality score less than or equal to 2. Reads with more than two
511 errors were discarded after truncation. Amplicon sequences were dereplicated using the function
512 'derepFastq.' Sample composition was inferred using the 'dada' function. Chimeras were
513 removed using 'removeBimeraDenovo.' We assigned taxonomy to the resulting ASVs using
514 'assignTaxonomy.' Using the R package 'vegan' (version 2.5-3), we calculated Bray-Curtis
515 dissimilarities and plotted these as a principal coordinate analysis plot (Fig. 1B).

516 **Species-specific differential expression analysis**

517 We filtered the RNA-seq counts table so that we only consider protein coding genes,
518 reducing the number of considered genes from 60,674 to 19,715. Host genes were filtered for
519 only protein coding genes using the R package ‘biomaRt’ with ensembl build 37. Within
520 DESeq2 (DESeq2 version 1.14.1), RNA-seq counts were further filtered such that each gene had
521 to be present at least once over all the samples, leaving 17860 tested genes (Love et al., 2014).
522 DESeq2 uses a negative binomial model to model the count data while it also estimates an
523 appropriate size factor to normalize each sample by its sequencing depth. Additionally, the
524 overdispersion parameter governing the negative binomial distribution is estimated per each gene
525 and using a regularization approach that can monitor outliers and adjust for the mean-variance
526 dependency. The parameter governing the mean gene expression after adjusting to its sequencing
527 depth is modeled as a linear combination that incorporates known batch effects (i.e., plate) and
528 the effect of the biological variable of interest (i.e., each microbiome):

529 Host gene expression \sim ExperimentPlate + Microbiome effects.

530 or, in mathematical terms:

531
$$Y_{nj} = \sum_s \beta_{js}^S M_{ns} + \sum_p \beta_{jp}^P P_{np}$$

532 Where Y_{nj} represents the internal DEseq parameter for mean gene expression for gene j and
533 experiment n , M_{ns} is the treatment indicator (control or microbiome for species s), and the β_{js}^M
534 parameter is the microbiome effect for each species. To model plate as a known batch effect we
535 use P_{np} and β_{jp}^P for the plate indicator variable and its effect on gene expression.

536 For four hominid microbiomes, $2^4=16$ effect configurations are possible (for each species
537 combination of which parameters $\beta_{js}^M = 0$), and we ran a likelihood test for each configuration
538 L_i : a gene can respond to a single primate microbiome (chimp, gorilla, human, or orangutan), a
539 gene can respond to two of the four primate microbiomes (chimp-gorilla, chimp-human,
540 chimp-orangutan, gorilla-human, gorilla-orangutan, human-orangutan), a gene can respond to
541 three of the four primate microbiomes (chimp-gorilla-human, chimp-human-orangutan,
542 chimp-gorilla-orangutan, or gorilla-human-orangutan), a gene can respond to all four primate
543 microbiomes, or a gene can show no response to any of the four primate microbiomes. The no
544 response case is considered the base case, or null model for all the likelihood ratio tests
545 performed.

546 To identify genes that respond to microbiomes from a specific primate species and to
547 detect the total number of differentially expressed genes that respond to each of the fifteen
548 possible non-null combinations of primate microbiomes we ran a likelihood ratio test against the

549 base model, which assumes that the host gene shows no response: Host gene expression ~
550 Experiment Plate and all the coefficients are zero. After determining across all genes and
551 configurations which were statistically significant FDR<10%. We used the likelihood statistics
552 L_{ji} for each gene j and configuration i to calculate the most probably configuration
553 $P(H_{ji}|D) = L_{ji}/\sum_i L_{ji}$.

554 **Enrichment analysis**

555 Enrichment analysis was performed using Ingenuity Pathway Analysis (IPA, QIAGEN
556 Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>,). We
557 analyzed genes that show a response to microbiomes from a specific primate species. Here, we
558 define those genes as genes that are upregulated or downregulated in response to a specific
559 primate host species, or that show no response to microbiomes from that primate species and
560 show a response to the other three primate host species. For example, genes that show a response
561 only to human microbiomes will be upregulated or downregulated in response to human
562 microbiomes, or show no response to human microbiomes and a response to chimpanzee, gorilla,
563 and orangutan microbiomes. Genes that show a response to three species but not the fourth are
564 also showing a species specific response to the fourth primate species.

565 We further validated these results using the R package ‘ClusterProfiler’ for enrichment
566 analysis using all detected genes present in at least one sample as the background set
567 (ClusterProfiler v3.2.14, Fig. 3C) (Yu et al., 2012). We used ENRICH for enrichment analysis
568 of the high and low-divergence genes and extracted the top ten response categories from the GO
569 Biological, GO Molecular, KEGG, and Reactome databases (SI Fig. 5, SI Fig. 6)(Chen et al.,
570 2013; Kuleshov et al., 2016).

571 To identify enrichment of high-divergence genes among genes that were previously found
572 to be associated with complex human disease and traits, we used data from the GWAS catalog
573 (Buniello et al., 2019). Since each GWAS has a different distribution of p-values and
574 significance cutoffs, we chose to use a set of $-\log_{10}(\text{p-value})$ cutoffs in the range of 8-50 (plotted
575 along the x axis in Fig. 3D). For a given trait, we identified the overlap between the genes
576 significantly associated with the disease at each cutoff and high-divergence genes³⁴, and
577 calculated a fold enrichment (plotted along the y axis in Fig. 3D), defined as the ratio of
578 observed/expected overlap between the two gene sets. We used a Fisher Exact Test to calculate a
579 p-value for each cutoff, and traits for which this value was significant after FDR corrections were
580 marked with a colored line in Fig. 3D.

581 K-means clustering was performed using the ‘kmeans’ function in base R (version 3.3.3)
582 on the cluster of microbes *P. copri*, *Methanobrevibacter* and *P. succinatutens* for the genes in
583 Fig. 4A. Enrichment analysis was performed using ENRICH on the two clusters of genes. A
584 k-means clustering analysis was also performed on the full set of microbial pathway-host gene
585 correlations in Fig. 4B to produce three clusters of genes.

586 **Log fold change of genes by primate species**

587 To calculate the fold changes for each gene for each of the four primate species, we used
588 a similar DESeq2 model to the one described above:

589
$$\text{Gene expression} \sim \text{ExperimentPlate} + \text{Species}$$

590 or, in mathematical terms:

591
$$Y_{nj} = \sum_s \beta_{js}^S M_{ns} + \sum_p \beta_{jp}^P P_{np}$$

592 Here, Species is a vector indicating which primate species the microbiome sample originated
593 from, and ExperimentPlate controls for the batch effect as before, but we just test the marginal
594 effect of each species-specific parameter β_{js}^S being not different than the untreated control β_j^C .
595 We use the contrast argument in DESeq2 to extract comparisons of each primate species against
596 the control. Thus, this resulted in log fold change calculations for each gene as it responds to
597 each of the four primate species' microbiomes. These values are available in SI Tables 24-27.

598 **Divergence scores for differentially expressed, conserved genes**

599 Using DESeq2, we identified genes that responded to microbiome treatment. The model
600 to determine whether a gene responds to treatment, we used the following model:

601
$$\text{Gene expression} \sim \text{ExperimentPlate} + \text{Treatment}$$

602 Where ExperimentPlate controls for the batch effect of the experiment, and Treatment is a binary
603 vector indicating whether the colonocytes are treated with a microbiome or act as a control for
604 the experiment. Mathematically:

605
$$Y_{nj} = \mu + \beta_j^T T_n + \sum_p \beta_{jp}^P P_{np}$$

606 Where Y_{nj} represents the internal DEseq mean gene expression parameter for gene j and
607 experiment n as before, T_n is the treatment indicator (control = 0 or microbiome = 1), and the
608 β_j^T parameter is the microbiome effect. Plate effects are modeled as before. To model plate as a
609 known batch effect we use P_{np} and β_{jp}^P for the plate indicator variable and its effect on gene
610 expression.

611 Log fold changes for each gene were calculated as described above, and then used to
612 calculate a divergence metric for each gene. We used a similar divergence calculation as
613 described in Hagai et al. (Hagai et al., 2018). Namely, for the genes identified as responding to

614 treatment with microbiomes, we used the log fold changes for each species in the following
615 equation:

$$616 \quad \text{Divergence} = \log_2 \left[\frac{1}{6} \sum_{i,j} (\log FC \text{ primate}_i - \log FC \text{ primate}_j)^2 \right]$$

617 Following Hagai et al., the top 25% of genes were assigned a “high-divergence” status,
618 and the lowest 25% of genes were assigned a “low-divergence” status. These genes were used in
619 the enrichment analyses described below.

620 The rest of the genes are considered “medium divergence” genes. These genes are used in
621 the enrichment analysis as a background set (Fig. 3A, Fig. 3C).

622 **Pairwise correlations between host genes and microbial species and pathways**

623 Using the microbial species abundances calculated from the metagenomic shotgun
624 sequencing, we ran correlation analysis between genes that are differentially expressed with
625 respect to treatment with microbiota and abundances of microbial species. Metaphlan2 reports
626 microbial species as a proportion of the total microbial community per sample. Microbial species
627 were filtered such that only microbial species present in at least half of the samples and that
628 reached a total summed relative abundance of 9% were included in the analysis, leaving 36
629 microbial species. We applied a center log-ratio transformation to the filtered microbial species
630 abundance data. Microbial pathways were filtered such that the total of each pathway had to be
631 greater than a summed threshold of 8000 reads per kilobase (RPK), leaving 95 microbial
632 pathways to be included in the analysis. Microbial pathways were normalized using the centered
633 log ratio transformation in a similar manner to the microbial species.

634 Using DESeq2, we identified which microbial species or pathways are associated with
635 differentially expressed genes using the following model:

$$636 \quad \text{Gene Expression} \sim \text{ExperimentPlate} + \text{Treatment} + \text{Microbial feature abundance}$$

637 Mathematically:

$$638 \quad Y_{nj} = \mu + \sum_p \beta_{jp}^P P_{np} + \beta_j^T T_n + \beta_j^{A(f)} A_{fn}$$

639 Where Y_{nj} represents the internal DEseq parameter for gene expression for gene j and
640 experiment n as before, T_n is the treatment indicator (control = 0 or microbiome = 1), and the
641 β_j^T parameter is the microbiome effect. Plate effects are modeled as before. The parameter
642 $\beta_j^{A(f)} A_{fn}$ is used to model the effect of the microbiome feature (i.e., microbial species or
643 pathway) f on gene expression. We statistically test effect $\beta_j^{A(f)} \neq 0$ in a separate DESeq model

644 run for each feature f . We used an FDR correction on the combined results from all models. The
645 microbial species abundance is a continuous numeric value that represents that center log ratio
646 transformed relative abundance of the microbial feature f A_{fn} for each sample n .

647 **TWAS analysis**

648 To directly investigate whether discovered effects on gene expression may contribute to
649 complex traits, we considered PTWAS gene-trait associations for 114 traits from Zhang et
650 al.(Zhang et al., 2019). PTWAS utilizes probabilistic eQTL annotations derived from
651 multi-variant Bayesian fine-mapping analysis of eQTL data across 49 tissues from GTEx v8 to
652 detect associations between gene expression levels and complex trait risk. Using the host genes
653 that were highly correlated with a microbial species and fell into the high-divergence category
654 (FDR<0.05), we overlapped the significant results with genes causally implicated in complex
655 traits across all tissues by Zhang et al. (PTWAS scan, 5% FDR). We repeated the same analysis
656 with the host genes that were highly correlated with a microbial pathway (FDR<0.01) and fell
657 into the high-divergence category.

658 **Data Availability**

659 Raw data for 16s rRNA sequencing, RNA-sequencing, and metagenomic shotgun
660 sequencing are available on the Sequence Read Archive (SRA) under submission ID
661 SUB7918466. For data tables used in this analysis, including tables for RNA-seq gene
662 expression counts, metagenomic shotgun sequencing species abundances, amplicon sequence
663 variant abundances, pathway abundances and metadata, see our Figshare project under the same
664 title as the manuscript at <https://figshare.com/account/home#/projects/87626>.

665 **Competing interests**

666 The authors have no competing interests to report.

667 References

- 668 Amato, K. R., G Sanders, J., Song, S. J., Nute, M., Metcalf, J. L., Thompson, L. R., Morton, J.
669 T., Amir, A., J McKenzie, V., Humphrey, G., Gogul, G., Gaffney, J., L Baden, A., A O
670 Britton, G., P Cuzzo, F., Di Fiore, A., J Dominy, N., L Goldberg, T., Gomez, A., ... R
671 Leigh, S. (2019). Evolutionary trends in host physiology outweigh dietary niche in
672 structuring primate gut microbiomes. *The ISME Journal*, *13*(3), 576–587.
673 <https://doi.org/10.1038/s41396-018-0175-0>
- 674 Amato, K. R., Mallott, E. K., McDonald, D., Dominy, N. J., Goldberg, T., Lambert, J. E.,
675 Swedell, L., Metcalf, J. L., Gomez, A., Britton, G. A. O., Stumpf, R. M., Leigh, S. R., &
676 Knight, R. (2019). Convergence of human and Old World monkey gut microbiomes
677 demonstrates the importance of human ecology over phylogeny. *Genome Biology*, *20*(1),
678 201. <https://doi.org/10.1186/s13059-019-1807-z>
- 679 Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., Spector, T. D.,
680 Keinan, A., Ley, R. E., Gevers, D., & Clark, A. G. (2015). Host genetic variation impacts
681 microbiome composition across human body sites. *Genome Biology*, *16*, 191.
682 <https://doi.org/10.1186/s13059-015-0759-1>
- 683 Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K., & Gilad, Y. (2008). Gene regulation in
684 primates evolves under tissue-specific selection pressures. *PLoS Genetics*, *4*(11), e1000271.
685 <https://doi.org/10.1371/journal.pgen.1000271>
- 686 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
687 sequence data. *Bioinformatics*, *30*(15), 2114–2120.
688 <https://doi.org/10.1093/bioinformatics/btu170>
- 689 Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti,
690 A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F.,
691 Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene
692 expression levels in mammalian organs. *Nature*, *478*(7369), 343–348.
693 <https://doi.org/10.1038/nature10532>
- 694 Britten, R. J., & Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a
695 speculation on the origins of evolutionary novelty. *The Quarterly Review of Biology*, *46*(2),
696 111–138. <https://doi.org/10.1086/406830>
- 697 Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using
698 DIAMOND. *Nature Methods*, *12*(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- 699 Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C.,
700 McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousitou, O., Whetzel, P.
701 L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., ...
702 Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide
703 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*,
704 *47*(D1), D1005–D1012. <https://doi.org/10.1093/nar/gky1120>

- 705 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P.
706 (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature*
707 *Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- 708 Camp, J. G., Frank, C. L., Lickwar, C. R., Guturu, H., Rube, T., Wenger, A. M., Chen, J.,
709 Bejerano, G., Crawford, G. E., & Rawls, J. F. (2014). Microbiota modulate transcription in
710 the intestinal epithelium without remodeling the accessible chromatin landscape. *Genome*
711 *Research*, 24(9), 1504–1516. <https://doi.org/10.1101/gr.165845.113>
- 712 Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan,
713 A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.
714 *BMC Bioinformatics*, 14, 128. <https://doi.org/10.1186/1471-2105-14-128>
- 715 Clayton, J. B., Vangay, P., Huang, H., Ward, T., Hillmann, B. M., Al-Ghalith, G. A., Travis, D.
716 A., Long, H. T., Tuan, B. V., Minh, V. V., Cabana, F., Nadler, T., Toddes, B., Murphy, T.,
717 Glander, K. E., Johnson, T. J., & Knights, D. (2016). Captivity humanizes the primate
718 microbiome. *Proceedings of the National Academy of Sciences of the United States of*
719 *America*, 113(37), 10376–10381. <https://doi.org/10.1073/pnas.1521835113>
- 720 Cox, D. G., Pontes, C., Guino, E., Navarro, M., Osorio, A., Canzian, F., Moreno, V., & Bellvitge
721 Colorectal Cancer Study Group. (2004). Polymorphisms in prostaglandin synthase
722 2/cyclooxygenase 2 (PTGS2/COX2) and risk of colorectal cancer. *British Journal of*
723 *Cancer*, 91(2), 339–343. <https://doi.org/10.1038/sj.bjc.6601906>
- 724 Dahmus, J. D., Kotler, D. L., Kastenber, D. M., & Kistler, C. A. (2018). The gut microbiome
725 and colorectal cancer: a review of bacterial pathogenesis. *Journal of Gastrointestinal*
726 *Oncology*, 9(4), 769–777. <https://doi.org/10.21037/jgo.2018.04.07>
- 727 de Lange, K. M., & Barrett, J. C. (2015). Understanding inflammatory bowel disease via
728 immunogenetics. *Journal of Autoimmunity*, 64, 91–100.
729 <https://doi.org/10.1016/j.jaut.2015.07.013>
- 730 Delsuc, F., Metcalf, J. L., Wegener Parfrey, L., Song, S. J., González, A., & Knight, R. (2014).
731 Convergence of gut microbiomes in myrmecophagous mammals. *Molecular Ecology*, 23(6),
732 1301–1317. <https://doi.org/10.1111/mec.12501>
- 733 De Vadder, F., Kovatcheva-Datchary, P., Zitoun, C., Duchamp, A., Bäckhed, F., & Mithieux, G.
734 (2016). Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal
735 Gluconeogenesis. *Cell Metabolism*, 24(1), 151–157.
736 <https://doi.org/10.1016/j.cmet.2016.06.013>
- 737 Emre, Y., & Imhof, B. A. (2014). Matricellular protein CCN1/CYR61: a new player in
738 inflammation and leukocyte trafficking. *Seminars in Immunopathology*, 36(2), 253–259.
739 <https://doi.org/10.1007/s00281-014-0420-1>
- 740 Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K.,
741 Muchmore, E., Varki, A., Ravid, R., Doxiadis, G. M., Bontrop, R. E., & Pääbo, S. (2002).
742 Intra- and interspecific variation in primate gene expression patterns. *Science*, 296(5566),
743 340–343. <https://doi.org/10.1126/science.1068996>

- 744 Ferreira, R. B. R., Gill, N., Willing, B. P., Antunes, L. C. M., Russell, S. L., Croxen, M. A., &
745 Finlay, B. B. (2011). The intestinal microbiota plays a role in Salmonella-induced colitis
746 independent of pathogen colonization. *PloS One*, 6(5), e20338.
747 <https://doi.org/10.1371/journal.pone.0020338>
- 748 Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G.,
749 Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018).
750 Species-level functional profiling of metagenomes and metatranscriptomes. *Nature*
751 *Methods*, 15(11), 962–968. <https://doi.org/10.1038/s41592-018-0176-y>
- 752 Gales, D., Clark, C., Manne, U., & Samuel, T. (2013). The Chemokine CXCL8 in
753 Carcinogenesis and Drug Response. *ISRN Oncology*, 2013, 859154.
754 <https://doi.org/10.1155/2013/859154>
- 755 Gijssbers, K., Van Assche, G., Joossens, S., Struyf, S., Proost, P., Rutgeerts, P., Geboes, K., & Van
756 Damme, J. (2004). CXCR1-binding chemokines in inflammatory bowel diseases:
757 down-regulated IL-8/CXCL8 production by leukocytes in Crohn’s disease and selective
758 GCP-2/CXCL6 expression in inflamed intestinal tissue. *European Journal of Immunology*,
759 34(7), 1992–2000. <https://doi.org/10.1002/eji.200324807>
- 760 Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P., & White, K. P. (2006). Expression profiling in
761 primates reveals a rapid evolution of human transcription factors. *Nature*, 440(7081),
762 242–245. <https://doi.org/10.1038/nature04559>
- 763 Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J.
764 B., Johnson, T. J., Hunter, R., Knights, D., & Beckman, K. B. (2016). Systematic
765 improvement of amplicon marker gene methods for increased accuracy in microbiome
766 studies. *Nature Biotechnology*, 34(9), 942–949. <https://doi.org/10.1038/nbt.3601>
- 767 Gomez, A., Petrzalkova, K. J., Burns, M. B., Yeoman, C. J., Amato, K. R., Vlckova, K., Modry,
768 D., Todd, A., Jost Robinson, C. A., Remis, M. J., Torralba, M. G., Morton, E., Umaña, J. D.,
769 Carbonero, F., Gaskins, H. R., Nelson, K. E., Wilson, B. A., Stumpf, R. M., White, B. A.,
770 ... Blekhman, R. (2016). Gut Microbiome of Coexisting BaAka Pygmies and Bantu
771 Reflects Gradients of Traditional Subsistence Patterns. *Cell Reports*, 14(9), 2142–2153.
772 <https://doi.org/10.1016/j.celrep.2016.02.013>
- 773 Gomez, A., Petrzalkova, K., Yeoman, C. J., Vlckova, K., Mrázek, J., Koppova, I., Carbonero, F.,
774 Ulanov, A., Modry, D., Todd, A., Torralba, M., Nelson, K. E., Gaskins, H. R., Wilson, B.,
775 Stumpf, R. M., White, B. A., & Leigh, S. R. (2015). Gut microbiome composition and
776 metabolomic profiles of wild western lowland gorillas (*Gorilla gorilla gorilla*) reflect host
777 ecology. *Molecular Ecology*, 24(10), 2551–2565. <https://doi.org/10.1111/mec.13181>
- 778 Gomez, A., Rothman, J. M., Petrzalkova, K., Yeoman, C. J., Vlckova, K., Umaña, J. D., Carr,
779 M., Modry, D., Todd, A., Torralba, M., Nelson, K. E., Stumpf, R. M., Wilson, B. A.,
780 Blekhman, R., White, B. A., & Leigh, S. R. (2016). Temporal variation selects for
781 diet-microbe co-metabolic traits in the gut of Gorilla spp. *The ISME Journal*, 10(2),
782 514–526. <https://doi.org/10.1038/ismej.2015.146>

- 783 Gomez, A., Sharma, A. K., Mallott, E. K., Petrzekova, K. J., Jost Robinson, C. A., Yeoman, C.
784 J., Carbonero, F., Pafco, B., Rothman, J. M., Ulanov, A., Vlckova, K., Amato, K. R.,
785 Schnorr, S. L., Dominy, N. J., Modry, D., Todd, A., Torralba, M., Nelson, K. E., Burns, M.
786 B., ... Leigh, S. R. (2019). Plasticity in the Human Gut Microbiome Defies Evolutionary
787 Constraints. *mSphere*, 4(4). <https://doi.org/10.1128/mSphere.00271-19>
- 788 Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont,
789 M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G., & Ley, R. E.
790 (2014). Human genetics shape the gut microbiome. *Cell*, 159(4), 789–799.
791 <https://doi.org/10.1016/j.cell.2014.09.053>
- 792 Grieneisen Laura E., Charpentier Marie J. E., Alberts Susan C., Blekhman Ran, Bradburd
793 Gideon, Tung Jenny, & Archie Elizabeth A. (2019). Genes, geology and germs: gut
794 microbiota across a primate hybrid zone are explained by site soil properties, not host
795 species. *Proceedings of the Royal Society B: Biological Sciences*, 286(1901), 20190431.
796 <https://doi.org/10.1098/rspb.2019.0431>
- 797 Habibian, J. S., Jelic, M., Bagchi, R. A., Lane, R. H., McKnight, R. A., McKinsey, T. A.,
798 Morrison, R. F., & Ferguson, B. S. (2017). DUSP5 functions as a feedback regulator of
799 TNF α -induced ERK1/2 dephosphorylation and inflammatory gene expression in adipocytes.
800 *Scientific Reports*, 7(1), 12879. <https://doi.org/10.1038/s41598-017-12861-y>
- 801 Hagai, T., Chen, X., Miragaia, R. J., Rostom, R., Gomes, T., Kunowska, N., Henriksson, J., Park,
802 J.-E., Proserpio, V., Donati, G., Bossini-Castillo, L., Vieira Braga, F. A., Naamati, G.,
803 Fletcher, J., Stephenson, E., Vegh, P., Trynka, G., Kondova, I., Dennis, M., ... Teichmann,
804 S. A. (2018). Gene expression variability across cells and species shapes innate immunity.
805 *Nature*, 563(7730), 197–202. <https://doi.org/10.1038/s41586-018-0657-2>
- 806 Hicks, A. L., Lee, K. J., Couto-Rodriguez, M., Patel, J., Sinha, R., Guo, C., Olson, S. H.,
807 Seimon, A., Seimon, T. A., Ondzie, A. U., Karesh, W. B., Reed, P., Cameron, K. N., Lipkin,
808 W. I., & Williams, B. L. (2018). Gut microbiomes of wild great apes fluctuate seasonally in
809 response to diet. *Nature Communications*, 9(1), 1786.
810 <https://doi.org/10.1038/s41467-018-04204-w>
- 811 Hörber, S., Hildebrand, D. G., Lieb, W. S., Lorscheid, S., Hailfinger, S., Schulze-Osthoff, K., &
812 Essmann, F. (2016). The Atypical Inhibitor of NF- κ B, I κ B ζ , Controls Macrophage
813 Interleukin-10 Expression. *The Journal of Biological Chemistry*, 291(24), 12851–12861.
814 <https://doi.org/10.1074/jbc.M116.718825>
- 815 Hotte, N. S. C., Salim, S. Y., Tso, R. H., Albert, E. J., Bach, P., Walker, J., Dieleman, L. A.,
816 Fedorak, R. N., & Madsen, K. L. (2012). Patients with inflammatory bowel disease exhibit
817 dysregulated responses to microbial DNA. *PLoS One*, 7(5), e37932.
818 <https://doi.org/10.1371/journal.pone.0037932>
- 819 Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy
820 human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
- 821 Khor, B., Gardet, A., & Xavier, R. J. (2011). Genetics and pathogenesis of inflammatory bowel

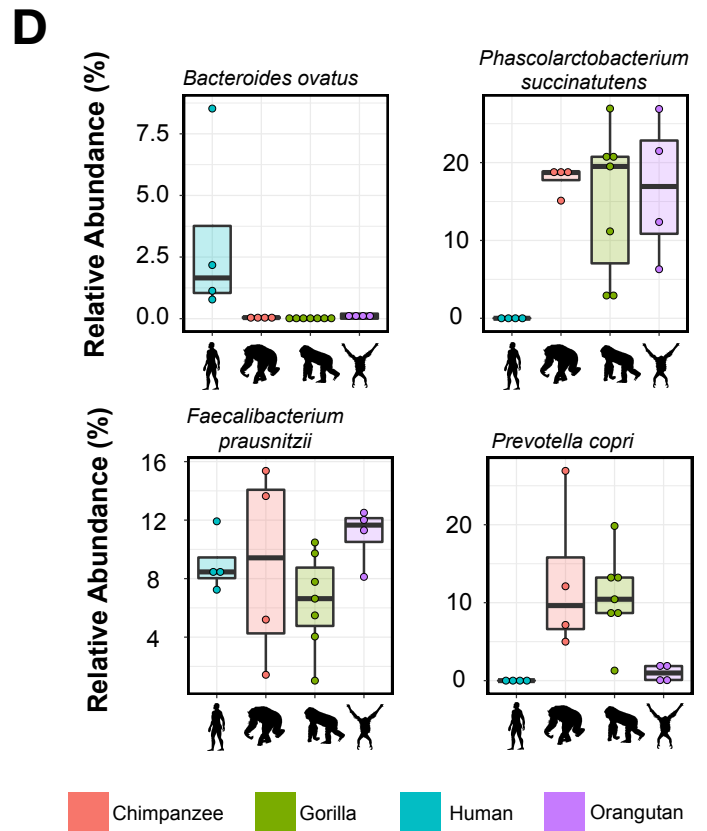
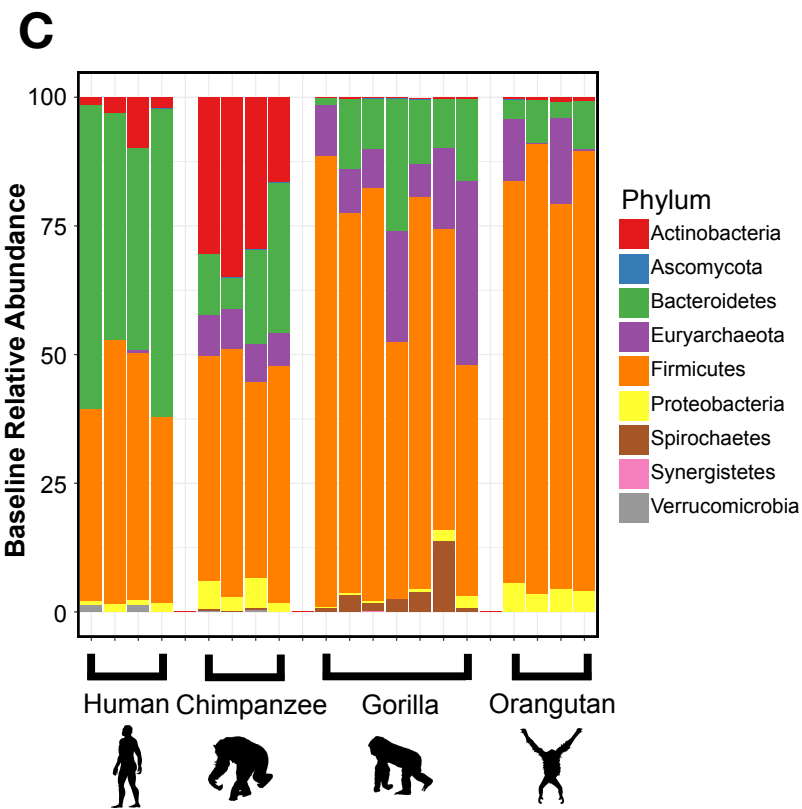
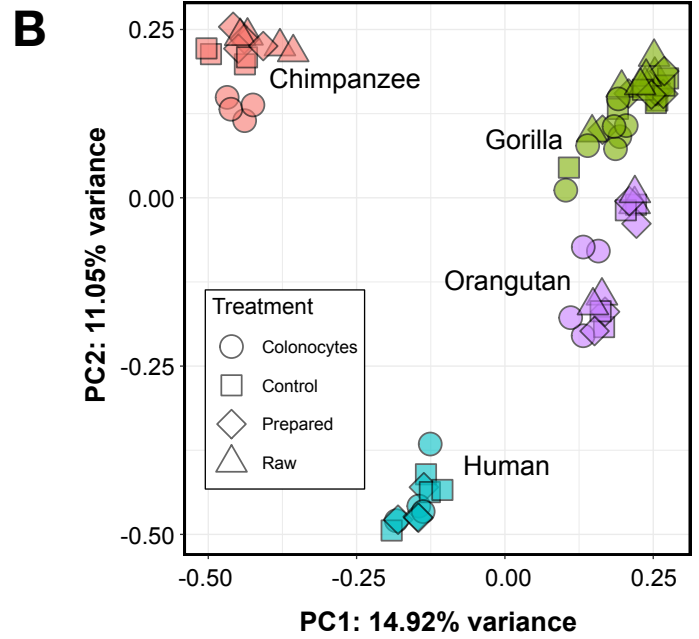
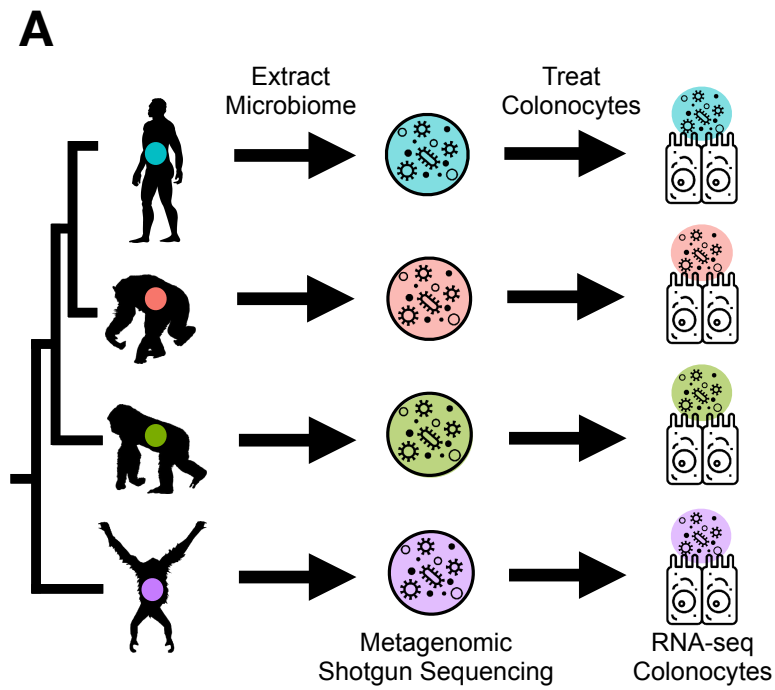
- 822 disease. *Nature*, 474(7351), 307–317. <https://doi.org/10.1038/nature10209>
- 823 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome
824 alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8),
825 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- 826 King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees.
827 *Science*, 188(4184), 107–116. <https://doi.org/10.1126/science.1090005>
- 828 Krämer, A., Green, J., Pollard, J., Jr, & Tugendreich, S. (2014). Causal analysis approaches in
829 Ingenuity Pathway Analysis. *Bioinformatics*, 30(4), 523–530.
830 <https://doi.org/10.1093/bioinformatics/btt703>
- 831 Krautkramer, K. A., Kreznar, J. H., Romano, K. A., Vivas, E. I., Barrett-Wilt, G. A., Rabaglia,
832 M. E., Keller, M. P., Attie, A. D., Rey, F. E., & Denu, J. M. (2016). Diet-Microbiota
833 Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. *Molecular*
834 *Cell*, 64(5), 982–992. <https://doi.org/10.1016/j.molcel.2016.10.025>
- 835 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev,
836 S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D.,
837 Gundersen, G. W., & Ma’ayan, A. (2016). Enrichr: a comprehensive gene set enrichment
838 analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97.
839 <https://doi.org/10.1093/nar/gkw377>
- 840 Lang, J. M., Pan, C., Cantor, R. M., Tang, W. H. W., Garcia-Garcia, J. C., Kurtz, I., Hazen, S. L.,
841 Bergeron, N., Krauss, R. M., & Lusi, A. J. (2018). Impact of Individual Traits, Saturated
842 Fat, and Protein Source on the Gut Microbiome. *mBio*, 9(6).
843 <https://doi.org/10.1128/mBio.01604-18>
- 844 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
845 *Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- 846 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
847 dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
848 <https://doi.org/10.1186/s13059-014-0550-8>
- 849 Luca, F., Kupfer, S. S., Knights, D., Khoruts, A., & Blekman, R. (2018). Functional Genomics
850 of Host-Microbiome Interactions in Humans. *Trends in Genetics: TIG*, 34(1), 30–40.
851 <https://doi.org/10.1016/j.tig.2017.10.001>
- 852 Mann, A. E., Mazel, F., Lemay, M. A., Morien, E., Billy, V., Kowalewski, M., Di Fiore, A., Link,
853 A., Goldberg, T. L., Tecot, S., Baden, A. L., Gomez, A., Sauter, M. L., Cuozzo, F. P., Rice,
854 G. A. O., Dominy, N. J., Stumpf, R., Lewis, R. J., Swedell, L., ... Wegener Parfrey, L.
855 (2019). Biodiversity of protists and nematodes in the wild nonhuman primate gut. *The ISME*
856 *Journal*. <https://doi.org/10.1038/s41396-019-0551-4>
- 857 Moeller, A. H., Degnan, P. H., Pusey, A. E., Wilson, M. L., Hahn, B. H., & Ochman, H. (2012).
858 Chimpanzees and humans harbour compositionally similar gut enterotypes. *Nature*
859 *Communications*, 3, 1179. <https://doi.org/10.1038/ncomms2159>
- 860 Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., Reyes, J. A.,

- 861 Shah, S. A., LeLeiko, N., Snapper, S. B., Bousvaros, A., Korzenik, J., Sands, B. E., Xavier,
862 R. J., & Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory
863 bowel disease and treatment. *Genome Biology*, *13*(9), R79.
864 <https://doi.org/10.1186/gb-2012-13-9-r79>
- 865 Müller, A., Hennig, A., Lorscheid, S., Grondona, P., Schulze-Osthoff, K., Hailfinger, S., &
866 Kramer, D. (2018). IκBζ is a key transcriptional regulator of IL-36-driven psoriasis-related
867 gene expression in keratinocytes. *Proceedings of the National Academy of Sciences of the*
868 *United States of America*, *115*(40), 10088–10093. <https://doi.org/10.1073/pnas.1801377115>
- 869 Nagpal, R., Shively, C. A., Appt, S. A., Register, T. C., Michalson, K. T., Vitolins, M. Z., &
870 Yadav, H. (2018). Gut Microbiome Composition in Non-human Primates Consuming a
871 Western or Mediterranean Diet. *Frontiers in Nutrition*, *5*, 28.
872 <https://doi.org/10.3389/fnut.2018.00028>
- 873 Nishida, A. H., & Ochman, H. (2019). A great-ape view of the gut microbiome. *Nature Reviews.*
874 *Genetics*. <https://doi.org/10.1038/s41576-018-0085-z>
- 875 Noor, S. O., Ridgway, K., Scovell, L., Kemsley, E. K., Lund, E. K., Jamieson, C., Johnson, I. T.,
876 & Narbad, A. (2010). Ulcerative colitis and irritable bowel patients exhibit distinct
877 abnormalities of the gut microbiota. *BMC Gastroenterology*, *10*, 134.
878 <https://doi.org/10.1186/1471-230X-10-134>
- 879 Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B. N., Peeters, M., Hahn, B. H., &
880 Hugenholtz, P. (2010). Evolutionary relationships of wild hominids recapitulated by gut
881 microbial communities. *PLoS Biology*, *8*(11), e1000546.
882 <https://doi.org/10.1371/journal.pbio.1000546>
- 883 Onoufriadis, A., Simpson, M. A., Pink, A. E., Di Meglio, P., Smith, C. H., Pullabhatla, V.,
884 Knight, J., Spain, S. L., Nestle, F. O., Burden, A. D., Capon, F., Trembath, R. C., & Barker,
885 J. N. (2011). Mutations in IL36RN/IL1F5 are associated with the severe episodic
886 inflammatory skin disease known as generalized pustular psoriasis. *American Journal of*
887 *Human Genetics*, *89*(3), 432–437. <https://doi.org/10.1016/j.ajhg.2011.07.022>
- 888 Pan, W.-H., Sommer, F., Falk-Paulsen, M., Ulas, T., Best, P., Fazio, A., Kachroo, P., Luzius, A.,
889 Jentzsch, M., Rehman, A., Müller, F., Lengauer, T., Walter, J., Künzel, S., Baines, J. F.,
890 Schreiber, S., Franke, A., Schultze, J. L., Bäckhed, F., & Rosenstiel, P. (2018). Exposure to
891 the gut microbiota drives distinct methylome and transcriptome changes in intestinal
892 epithelial cells during postnatal development. *Genome Medicine*, *10*(1), 27.
893 <https://doi.org/10.1186/s13073-018-0534-5>
- 894 Parisinos, C. A., Serghiou, S., Katsoulis, M., George, M. J., Patel, R. S., Hemingway, H., &
895 Hingorani, A. D. (2018). Variation in Interleukin 6 Receptor Gene Associates With Risk of
896 Crohn's Disease and Ulcerative Colitis. *Gastroenterology*, *155*(2), 303–306.e2.
897 <https://doi.org/10.1053/j.gastro.2018.05.022>
- 898 Qin, Y., Roberts, J. D., Grimm, S. A., Lih, F. B., Deterding, L. J., Li, R., Chrysovergis, K., &
899 Wade, P. A. (2018). An obesity-associated gut microbiome reprograms the intestinal

- 900 epigenome and leads to altered colonic gene expression. *Genome Biology*, 19(1), 7.
901 <https://doi.org/10.1186/s13059-018-1389-1>
- 902 Raymann, K., Moeller, A. H., Goodman, A. L., & Ochman, H. (2017). Unexplored Archaeal
903 Diversity in the Great Ape Gut Microbiome. *mSphere*, 2(1).
904 <https://doi.org/10.1128/mSphere.00026-17>
- 905 Remis, M. J. (1997). Western lowland gorillas (*Gorilla gorilla gorilla*) as seasonal frugivores: use
906 of variable resources. *American Journal of Primatology*, 43(2), 87–109.
907 <https://doi.org/3.0.CO;2-T>>10.1002/(SICI)1098-2345(1997)43:2<87::AID-AJP1>3.0.CO;2
908 -T
- 909 Ren, K., & Torres, R. (2009). Role of interleukin-1beta during pain and inflammation. *Brain*
910 *Research Reviews*, 60(1), 57–64. <https://doi.org/10.1016/j.brainresrev.2008.12.020>
- 911 Richards, A. L., Burns, M. B., Alazizi, A., Barreiro, L. B., Pique-Regi, R., Blekhman, R., &
912 Luca, F. (2016). Genetic and transcriptional analysis of human host response to healthy gut
913 microbiota. *mSystems*, 1(4). <https://doi.org/10.1128/mSystems.00067-16>
- 914 Richards, A. L., Muehlbauer, A. L., Alazizi, A., Burns, M. B., Findley, A., Messina, F., Gould, T.
915 J., Cascardo, C., Pique-Regi, R., Blekhman, R., & Luca, F. (2019). Gut Microbiota Has a
916 Widespread and Modifiable Effect on Host Gene Regulation. *mSystems*, 4(5).
917 <https://doi.org/10.1128/mSystems.00323-18>
- 918 Rincon, M. (2012). Interleukin-6: from an inflammatory marker to a target for inflammatory
919 diseases. *Trends in Immunology*, 33(11), 571–577. <https://doi.org/10.1016/j.it.2012.07.003>
- 920 Russell, S. E., Horan, R. M., Stefanska, A. M., Carey, A., Leon, G., Aguilera, M., Statovci, D.,
921 Moran, T., Fallon, P. G., Shanahan, F., Brint, E. K., Melgar, S., Hussey, S., & Walsh, P. T.
922 (2016). IL-36 α expression is elevated in ulcerative colitis and promotes colonic
923 inflammation. *Mucosal Immunology*, 9(5), 1193–1204. <https://doi.org/10.1038/mi.2015.134>
- 924 Scher, J. U., & Abramson, S. B. (2011). The microbiome and rheumatoid arthritis. *Nature*
925 *Reviews. Rheumatology*, 7(10), 569–578. <https://doi.org/10.1038/nrrheum.2011.121>
- 926 Schulze, H. A., Häsler, R., Mah, N., Lu, T., Nikolaus, S., Costello, C. M., & Schreiber, S. (2008).
927 From model cell line to in vivo gene expression: disease-related intestinal gene expression
928 in IBD. *Genes and Immunity*, 9(3), 240–248. <https://doi.org/10.1038/gene.2008.11>
- 929 Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., & UniProt Consortium. (2015).
930 UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity
931 searches. *Bioinformatics*, 31(6), 926–932. <https://doi.org/10.1093/bioinformatics/btu739>
- 932 Taylor, A. B. (2006). Feeding behavior, diet, and the functional consequences of jaw form in
933 orangutans, with implications for the evolution of Pongo. *Journal of Human Evolution*,
934 50(4), 377–393. <https://doi.org/10.1016/j.jhevol.2005.10.006>
- 935 Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A.,
936 Huttenhower, C., & Segata, N. (2015). MetaPhlan2 for enhanced metagenomic taxonomic
937 profiling. *Nature Methods*, 12(10), 902–903. <https://doi.org/10.1038/nmeth.3589>
- 938 Tung, J., Barreiro, L. B., Burns, M. B., Grenier, J.-C., Lynch, J., Grieneisen, L. E., Altmann, J.,

- 939 Alberts, S. C., Blekhman, R., & Archie, E. A. (2015). Social networks predict gut
940 microbiome composition in wild baboons. *eLife*, 4. <https://doi.org/10.7554/eLife.05224>
- 941 Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I.
942 (2007). The human microbiome project. *Nature*, 449(7164), 804–810.
943 <https://doi.org/10.1038/nature06244>
- 944 Tutin, C. E. G., & Fernandez, M. (1993). Composition of the diet of chimpanzees and
945 comparisons with that of sympatric lowland gorillas in the lopé reserve, gabon. *American*
946 *Journal of Primatology*, 30(3), 195–211. <https://doi.org/10.1002/ajp.1350300305>
- 947 Vogel, E. R., Harrison, M. E., Zulfa, A., Bransford, T. D., Alavi, S. E., Husson, S.,
948 Morrogh-Bernard, H., Santiano, Firtsman, T., Utami-Atmoko, S. S., van Noordwijk, M. A.,
949 & Farida, W. R. (2015). Nutritional Differences between Two Orangutan Habitats:
950 Implications for Population Density. *PloS One*, 10(10), e0138612.
951 <https://doi.org/10.1371/journal.pone.0138612>
- 952 Wang, W., Yu, X., Wu, C., & Jin, H. (2017). IL-36 γ inhibits differentiation and induces
953 inflammation of keratinocyte via Wnt signaling pathway in psoriasis. *International Journal*
954 *of Medical Sciences*, 14(10), 1002–1007. <https://doi.org/10.7150/ijms.20809>
- 955 Watts, D. P., Potts, K. B., Lwanga, J. S., & Mitani, J. C. (2012). Diet of chimpanzees (Pan
956 troglodytes schweinfurthii) at Ngogo, Kibale National Park, Uganda, 1. Diet composition
957 and diversity. *American Journal of Primatology*, 74(2), 114–129.
958 <https://doi.org/10.1002/ajp.21016>
- 959 Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology*
960 *Reviews*, 20(4), 593–621. <https://doi.org/10.1128/CMR.00008-07>
- 961 Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M.,
962 Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R.,
963 Nessel, L., Li, H., Bushman, F. D., & Lewis, J. D. (2011). Linking long-term dietary
964 patterns with gut microbial enterotypes. *Science*, 334(6052), 105–108.
965 <https://doi.org/10.1126/science.1208344>
- 966 Ye, Y., & Doak, T. G. (2009). A parsimony approach to biological pathway
967 reconstruction/inference for genomes and metagenomes. *PLoS Computational Biology*,
968 5(8), e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>
- 969 Yue, X., Wu, L., & Hu, W. (2015). The regulation of leukemia inhibitory factor. *Cancer Cell &*
970 *Microenvironment*, 2(3). <https://doi.org/10.14800/ccm.877>
- 971 Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing
972 Biological Themes Among Gene Clusters. *Omics: A Journal of Integrative Biology*, 16(5),
973 284–287. <https://doi.org/10.1089/omi.2011.0118>
- 974 Zhang, S.-L., Wang, S.-N., & Miao, C.-Y. (2017). Influence of Microbiota on Intestinal Immune
975 System in Ulcerative Colitis and Its Intervention. *Frontiers in Immunology*, 8, 1674.
976 <https://doi.org/10.3389/fimmu.2017.01674>
- 977 Zhang, Y., Quick, C., Yu, K., Barbeira, A., The GTEx Consortium, Luca, F., Pique-Regi, R., Im,

978 H. K., & Wen, X. (2019). Investigating tissue-relevant causal molecular mechanisms of
979 complex traits using probabilistic TWAS analysis. In *bioRxiv* (p. 808295).
980 <https://doi.org/10.1101/808295>



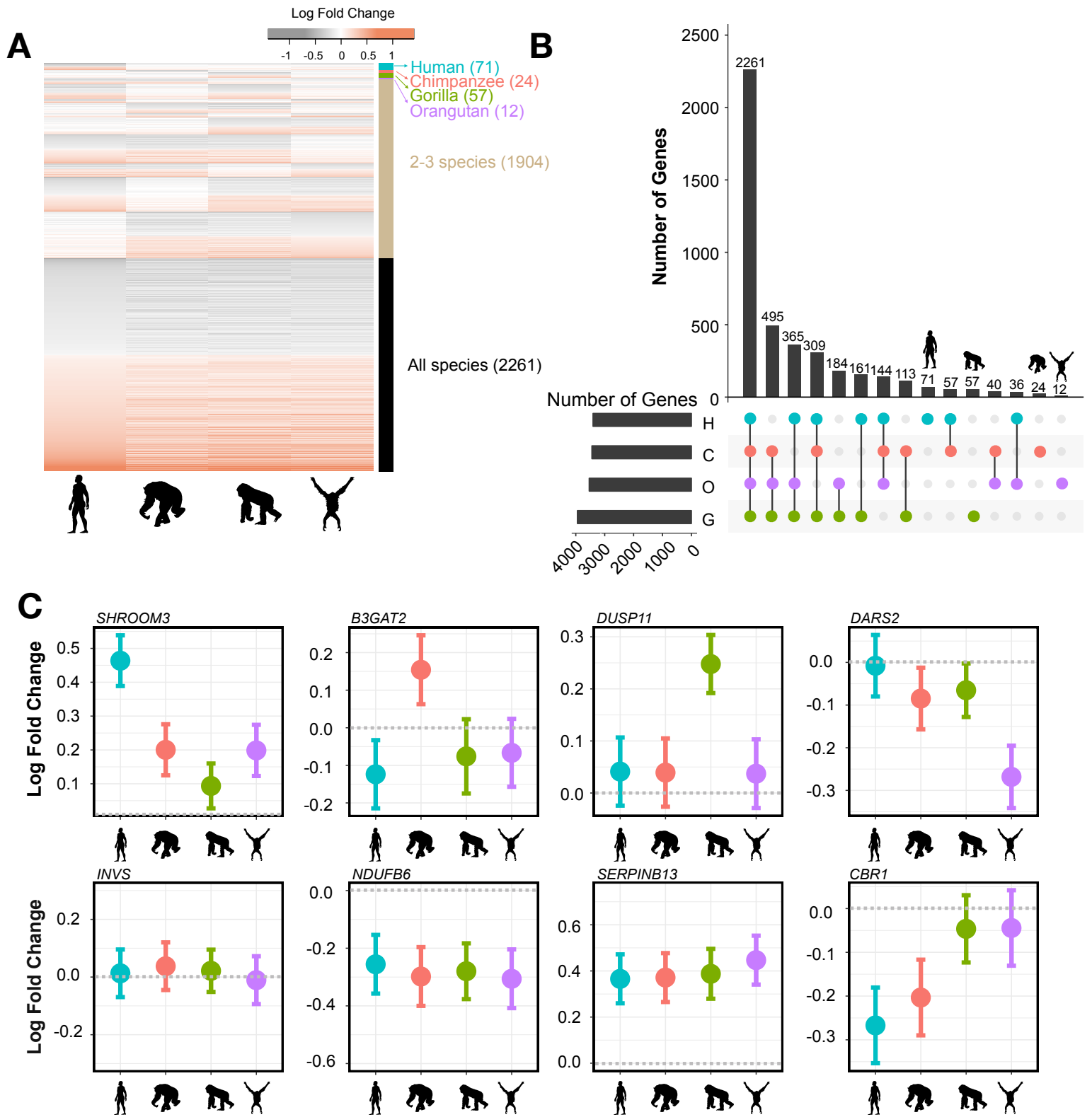
981 **Figure 1. Experimental design and gut microbiome composition**

982 **(A)** Experimental design. Live microbiomes were extracted from fecal samples from
983 chimpanzees (n=4), gorillas (n=7), humans (n=4), and orangutans (n=4). Microbes were
984 incubated with human colonic epithelial cells for 2 hours, after which RNA-seq was performed
985 on the epithelial cells. Metagenomic shotgun sequencing and 16s rRNA sequencing were
986 performed on the microbiome samples to determine microbiome composition.

987 **(B)** Bray-Curtis dissimilarity of the all microbiome samples from all four primate species at
988 different stages of the experiment: in raw fecal samples (raw), after extraction from fecal samples
989 and before the experiment (prepared), after incubation with colonocytes (colonocytes), and after
990 incubation without colonocytes (control).

991 **(C)** Stacked barplot showing the relative abundances of microbial phyla for each hominid fecal
992 sample prior to culturing but after extracting the microbiota for treatment.

993 **(D)** Examples of microbial species (from shotgun metagenomics) that show various patterns of
994 abundance across hominid species. *Bacteroides ovatus* shows a high abundance in humans
995 relative to the other hominid species. *Phascolarctobacterium succinatutens* is highly abundant in
996 the non-human hominid but not present in the human fecal samples. *Faecalibacterium*
997 *prausnitzii* is highly abundant in all four hominid species. *Prevotella copri* is highly abundant in
998 the chimpanzee and gorilla samples, has a lower abundance in the orangutan samples, and is not
999 present in the human samples.

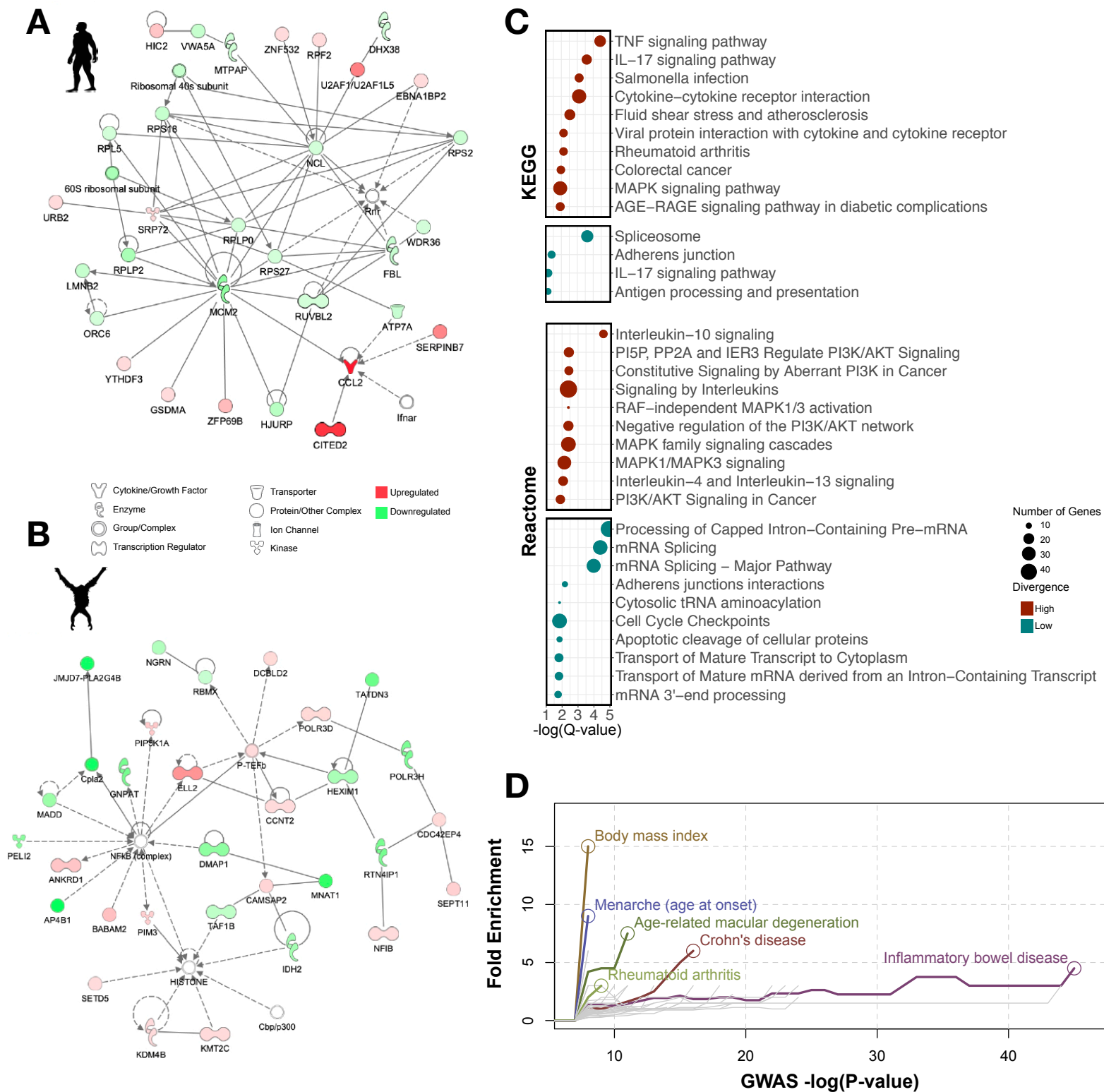


1000 **Figure 2. Patterns of host gene expression changes in response to hominid microbiome**
1001 **treatment**

1002 **(A)** Log₂ fold change for all differentially expressed genes (rows), grouped by expression
1003 pattern. The colored bar on the right hand side indicates in response to which hominid
1004 microbiome these genes change their expression.

1005 **(B)** UpSet plot visualizing the intersections among the sets of host genes that respond to hominid
1006 microbiomes.

1007 **(C)** Examples of the expression pattern of 8 differentially expressed genes. Each panel shows the
1008 change in expression (y-axis) in response to the four hominid microbiomes (x-axis) of a single
1009 host gene, with the gene name listed at the top of each panel. Error bars indicate 1X the standard
1010 error.



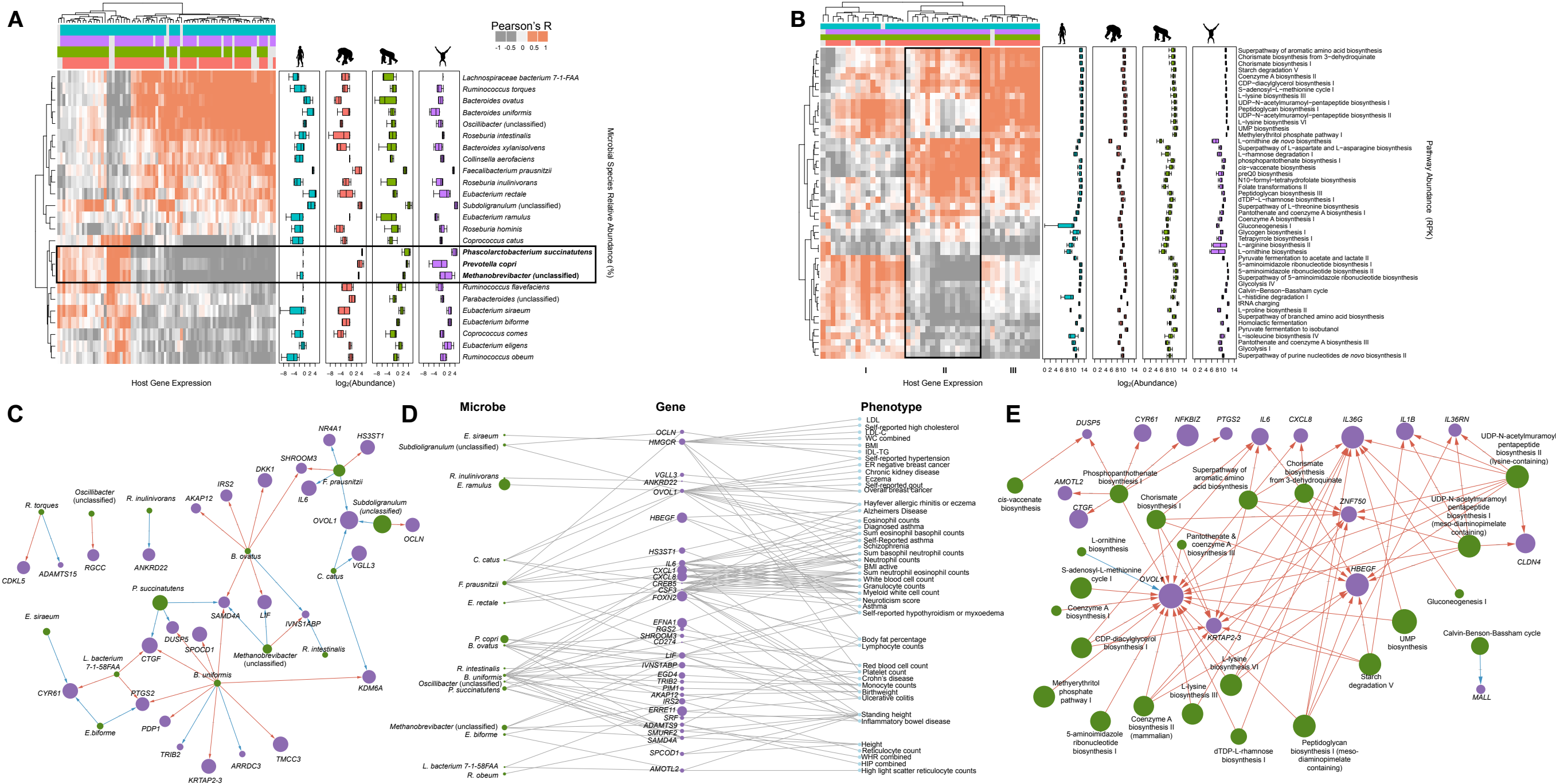
1011 **Figure 3. Interaction networks and functional enrichment categories for host genes**
1012 **responding to hominid gut microbiomes**

1013 **(A)** Interaction network showing host genes that respond only to human microbiomes, generated
1014 using Ingenuity Pathway Analysis.

1015 **(B)** Similar to (A), but including host genes that respond only to orangutan microbiomes.

1016 **(C)** Functional categories in the KEGG (top) and Reactome (bottom) databases enriched among
1017 high-divergence genes (red) and low-divergence genes (blue). X-axis indicates the statistical
1018 significance of enrichment, and the circle size corresponds to the number of genes in each
1019 category.

1020 **(D)** Complex disease enriched among genes that respond to hominid microbiomes. Fold
1021 enrichment (y-axis) is shown for a given P value threshold (x-axis) to define genes that are
1022 associated with each complex disease in the GWAS Catalog. Each colored line represents a
1023 different complex disease with an enrichment of at least three-fold, with a circle indicating the
1024 most significant P-value threshold. Diseases that did not reach significance are shown in grey
1025 lines.



1026 **Figure 4. Relationship between host gene expression and specific microbiome features**

1027 **(A)** Heatmap showing correlations between microbial species (rows) and host genes (columns).

1028 Colors at the top indicate to which hominid microbiome a gene responds to. Boxplots to the right

1029 show the abundance of each microbial species in each hominid microbiome (microbial

1030 abundance transformed by \log_2).

1031 **(B)** Similar to (A), but showing microbial pathways instead of species.

1032 **(C)** Network visualization of microbial species (green) and high-divergence host genes (purple)

1033 that respond to each species connected with an arrow. Node size of microbial species and host

1034 genes corresponds to species abundance and \log_2 fold change of the differential expression,

1035 respectively. Arrow colors indicate whether a microbial species increases (blue) or decreases

1036 (red) the expression of the connected host gene.

1037 **(D)** Three tier network showing microbial species (left column), the host genes they each

1038 regulate (middle column), and TWAS phenotypes these genes are associated with (right column).

1039 Microbial species and host gene node size indicates microbial abundance and differential

1040 expression, respectively, correlated with high-divergence genes and TWAS phenotypes.

1041 **(E)** Similar to (C), but showing microbial pathways instead of species.