RNAseq studies reveal distinct transcriptional response to vitamin A deficiency in small intestine versus colon, discovering novel VA-regulated genes

Zhi Chai[1,2], Yafei Lyu[3,6], Qiuyan Chen[2], Cheng-Hsin Wei[2,7], Lindsay Snyder[4,8], Veronika Weaver[4], Qunhua Li[5], Margherita T. Cantorna[4], A. Catharine Ross[2*].

[1]Intercollege Graduate Degree Program in Physiology, [2]Department of Nutritional Sciences, [3]Intercollege Graduate Degree Program in Bioinformatics and Genomics, [4]Department of Veterinary and Biomedical Sciences, [5]Department of Statistics. The Pennsylvania State University, University Park, PA. [6]Present address: Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Present address: Frederick National Laboratory for Cancer Research, Frederick, MD, USA. [8]Present address: Center for Evolutionary and Theoretical Immunology, The University of New Mexico, Albuquerque, NM, USA.

***Corresponding author**

acr6@psu.edu

110 Chandlee lab

University Park. PA 16802

Key words: vitamin A deficiency, small intestine, colon, gene expression profile, RNAseq, differential expression, WGCNA.

Abstract

Vitamin A (VA) deficiency remains prevalent in resource limited countries, affecting over 250 million preschool aged children. Vitamin A deficiency is associated with reduced intestinal barrier function and increased risk of mortality due to mucosal infection. *Citrobacter rodentium* (*C. rodentium*) infection in mice is a model for diarrheal diseases in humans. During *C. rodentium* infection, vitamin A deficient (VAD) mice displayed reduced survival rate and pathogen clearance compared to their vitamin A sufficient (VAS) counterparts. **Objectives:** To characterize and compare the impact of vitamin A (VA) deficiency on gene expression patterns in the small intestine (SI) and the colon, and to discover novel target genes in VA-related biological pathways. **Methods:** vitamin A deficient (VAD) mice were generated by feeding VAD diet to pregnant C57/BL6 dams and their post-weaning offspring. Total mRNA extracted from SI and colon were sequenced using Illumina HiSeq 2500 platform. Differentially Expressed Gene (DEG), Gene Ontology (GO) enrichment, and Weighted Gene Co-expression Network Analysis (WGCNA) were performed to characterize expression patterns and co-expression patterns. **Results:** The comparison between VAS and VAD groups detected 49 and 94 DEGs in SI and colon, respectively. According to GO information, DEGs in the SI demonstrated significant enrichment in categories relevant to retinoid metabolic process, molecule binding, and immune function. Immunity related pathways, such as "humoral immune response" and "complement activation," were positively associated with VA in SI. On the contrary, in colon, "cell division" was the only enriched category and was negatively associated with VA. Three co-expression modules showed significant

46 correlation with VA status in SI and were identified as modules of interest. Those
47 modules contained four known retinoic acid targets. Therefore we have prioritized the
48 other module members (e.g. Mbl2, Mmp9, Cxcl14, and Nr0b2) to be investigated as
49 candidate genes regulated by VA. Also in SI, markers of two cell types, mast cell and
50 Tuft cell, were found altered by VA status. Comparison of co-expression modules
51 between SI and colon indicated distinct regulatory networks in these two organs.

52 **Introduction**

53 Vitamin A (VA) deficiency remains a global public health issue, particularly in
54 resource-limited areas. It is estimated that over 20% of preschool aged children are
55 clinically VA deficient [1]. VA deficiency causes xerophthalmia and is positively
56 associated with the severity of infectious disease [2]. It has been shown that high-dose VA
57 supplementation reduced childhood mortality and morbidity [3]. The World Health
58 Organization (WHO) reported that VA supplementation reduced diarrhea-related
59 mortality by 33% and measles-related mortality by 50% [4].

60 VA is a pleiotropic regulator of gut immunity. On one hand, VA regulates the
61 development, function, and homing of immune cells in the gut. RA balances
62 inflammatory and regulatory immune responses through suppression of IFN-γ and IL-17
63 production and induction of FoxP3 and IL-10 secreting regulatory T cells [5,6]. RA induces
64 gut homing of lymphocytes by upregulating CCR9 and α4β7 expressions [7,8]. RA also
65 functions as a specific IgA isotype switching factor that facilitates the differentiation of
66 IgA^+ antibody secreting cells and enhances IgA production in the presence of TGF-β.

Finally, VA has complicated interactions with innate lymphoid cells and intestinal microbiota, two other players regulating mucosal immunity [9]. On the other hand, VA affects the cell differentiation and function of the intestinal epithelium [10]. Intestinal epithelial cells (ECs) are not only important for nutrient absorption but also act as a barrier between the host and the intestinal microbiota [11]. In murine models, VA deficiency has been associated with a higher goblet cell population, increased mucin production, as well as reduced defensin production and Paneth cell marker expression [12-18]. Retinoic acid (RA) is the most bioactive form of VA. RA is reported to enhance barrier integrity by inducing tight junction-associated genes such as occludin, claudins and zonula occludens [19]. In sum, VA is a versatile regulator of both the innate and adaptive immunity in the gut.

Citrobacter rodentium (*C. rodentium*) is a gram negative, natural mouse intestinal pathogen mimicking enteropathogenic *Escherichia coli* (EPEC) infection in humans [20]. Non-susceptible mouse strains are able to clear the infection by day 30 post-infection (32). Production of IL-22 and IL-17 by type 3 lymphoid cells (ILC3) are essential factors in gut protection at early stages of infection, whereas the clearance of infection requires robust Th17 responses in the later stages [21,22]. Host VA status has been reported to affect the host responses to *C. rodentium* infection. VA sufficient (VAS) mice were able to survive and clear the infection, whereas VAD animals suffered from a 40% mortality rate and survivors become non-symptomatic carrier of *C. rodentium* [23].

To determine the key genes that mediated the divergent host responses between VAS and VAD mice, we compared the gene expression levels in the intestines of the naïve VAS and VAD mice. SI is a major site of VA metabolism. In the steady state,

90 retinol concentrations are significantly higher in the SI and mesenteric lymph nodes
91 (mLNs) compared with other tissues except for liver, the major storage site of retinol [24,25].
92 ECs and dendritic cells (DCs) in SI are RA producers. Intestinal RA concentration
93 generally follows a gradient from proximal to distal, correlating with the imprinting
94 ability of DCs [24]. Additionally, Colon is the major site of infection for *C. rodentium*.
95 Therefore, both SI and colon were the organs of interest for our study. As more than 500
96 genes have been found to be regulated by RA, whole tissue RNAseq analysis was
97 performed to provide a comprehensive and exploratory view of the expression profiles in
98 both SI and colon [26]. Differential expression and co-expression analyses were used to
99 characterize the VA effect on the transcriptomes in those two organs. This process
100 identified several immune-related genes (Mbl2, Mmp9, and Cxcl14), transcription factor
101 (Nr0b2), and cell types (mast cell and Tuft cell) as novel targets under the control of VA
102 in SI, and showed that co-expression networks of the two organs had different patterns.

103 **Results**

104 **Model validation: Serum retinol is lower in VAD mice than VAS mice.**

105 VAS and VAD mice were generated by dietary means as previously described
106 [27,28]. In order to validate the VA status of our animal model, we collected serum samples
107 at the end of the SI study. As expected, VAD mice had significantly lower serum retinol
108 levels than their VAS counterparts ($P<0.001$). This indicated that VA deficiency was

successfully induced in our animal model, enabling us to further analyze the VA effect on the intestinal gene transcription (**Fig. 1**).

Overview of the results derived from Differential expression and Co-expression analyses

The two RNAseq datasets reported in this article are available in NCBI Gene Expression Omnibus (GEO) under the accession number GSE143290. Mapping revealed 24421 genes; these genes were subsequently screened for low expression transcripts and normalized by DESeq2 (**Fig. 2**). Among the 14368 SI-expressed genes, 49 DEGs corresponding to VA effect, and 9 co-expression modules were identified. Similarly, 94 DEGs and 13 co-expression modules were identified among the 15340 genes in the colon dataset in the colon study. Of these, 14059 genes were shared between the two datasets and these were used for the module preservation analysis (**Fig. 2**).

In the complete two-by-two factorial design, the effect of *C. rodentium* infection and the Interaction effect (VA x *C. rodentium* infection) were observed in colon, the outcomes of which will be presented in a different article. The present article focuses on the VA effects in the intestines.

Data pre-processing

The raw count of 24421 transcripts was obtained after mapping in each of the two studies. Transcripts with low expression levels were first removed, for the purpose of assuring that each analysis would be focused on the tissue-specific mRNAs and would

129 not be disrupted by the genes not expressed in that organ. As a result, 14368 and 15340
130 genes were remained in the SI and colon datasets, respectively. Both datasets, following
131 normalization according to DESeq2 package, were used for the following DE analysis
132 and co-expression network analysis (**Fig. 2**).

133 **Identification of DEGs**

134 The criteria of the DEGs were set as follows: i) |Fold Change|>2, to ensure
135 biological significance. ii) Adjusted p value calculated by DESeq2 (padj) < 0.05, to
136 ensure statistical significance.

137 For the SI study, a total of 49 genes were identified as DEGs of VA effect.
138 Among the 49 DEGs, 18 were upregulated (**Supplementary Table 1**) and 31 were
139 downregulated (**Supplementary Table 2**) in the SI of VAS mice comparing with the
140 VAD mice. Visualization of the statistical significance and Fold Change of all the DEGs
141 on the same plot is helpful for sorting and prioritizing genes of interest. Therefore, the
142 volcano plot of the DEGs in the SI study is depicted (**Fig. 3a**). Points of interest in
143 volcano plots are found toward the top of the plot that are far to either left-hand side (e.g.
144 Cd207, Mcpt4, and Fcgr2b) or right-hand side (Isx, Cyp26b1, Mmp9, and Mbl2). These
145 DEGs display large magnitude of fold changes (hence being left or right of center) as
146 well as high statistical significance (hence being toward the top). A heatmap of the DEGs
147 in the SI study is shown in **Supplementary Fig. 1**.

148 For the colon study, under the criteria of i) and ii), 94 DEGs were identified, of
149 which 34 genes had significantly higher expressions (**Supplementary Table 3**) and 60

genes had lower expressions (**Supplementary Table 4**) in the VAS group. To facilitate visualization, the volcano plot and heatmap are shown in **Fig. 3b** and **Supplementary Fig. 2**. Some of the most significantly upregulated DEGs are also well-known markers for goblet cell activities. Tff3 (trefoil factor 3) and Muc1 (mucin 1) are two goblet cell markers [29]. Clca1 (chloride channel accessory 1) and Clca4b (chloride channel accessory 4B), two genes known to regulate mucus production and neutrophil recruitment [30-34]. Among all the upregulated DEGs under the VAS condition, Isx, Prdm1 (PR domain containing 1), and Zxdb (zinc finger X-linked duplicated B) are three transcription factors. Among all the downregulated DEGs, Klk1 (kallikrein 1) is a mucus regulator and Foxm1 (forkhead box M1) is a transcription factor.

GO and KEGG enrichment of DEGs

Enrichment analysis were performed separately on all DEGs, upregulated DEGs, and downregulated DEGs of the VAS group. In the SI study, the GO analysis on all 49 DEGs suggested significant enrichment in categories relevant to immune function, molecule binding, and retinoid metabolic process (**Supplementary Fig. 3**). In agreement with our anticipation, the 49 DEGs between VAD and the VAS groups were significantly enriched in "retinoid metabolic process." Bco1, Cyp26b1, and Isx are the three DEGs belonging to this enriched category. The 18 upregulated DEGs in the SI study were significantly enriched in five Biological Processes (**Fig. 3c**) and one KEGG pathway (*Staphylococcus aureus* infection, mmu05150). The 31 downregulated DEGs in the SI study were significantly enriched in 10 categories of Molecular Function (**Fig. 3d**).

171 In the colon study, downregulated DEGs (n=60) in VAS group comparing with
172 VAD group were significantly enriched in "Cell Division" (GO term) and "Cell Cycle"
173 (KEGG term, mmu04110). Upregulated DEGs (n=34) in the colon study were not
174 enriched in any GO or KEGG pathways.

175 **Co-expression network constructed by WGCNA**

176 Gene co-expression network constructed by WGCNA provides a systems biology
177 approach to understand the gene networks instead of individual genes. To make the result
178 comparable to that of the differential expression analysis, the same screening and
179 normalization methods were used on both datasets (**Fig .2**).
180 For the SI study, according to the scale-free topology model fit and mean
181 connectivity, soft threshold power β was set to 7 (**Fig. 4a**). Minimum module size was set
182 at 30. A total of 12 samples and 14368 genes were utilized for co-expression network
183 construction, acquiring nine modules with module size (MS) ranging from 202 to 5187.
184 These modules in the SI network will be referred to by their color labels henceforth (e.g.
185 SI(Color)), as the standard of this approach. The modules labeled by colors were depicted
186 in the hierarchical clustering dendrogram in **Fig. 4b**. Grey module is used to hold all
187 genes that do not clearly belong to any other modules. The gene expression patterns of
188 each module across all samples were demonstrated as heatmaps (**Supplementary Fig. 4**).
189 The same analysis was performed on the colon transcriptomes. Soft threshold
190 power β was set to 26 to produce an approximate scale-free network with appropriate
191 mean connectivity (**Fig. 4d**). With the input of 16 samples and 15340 genes, 13 modules

192 were identified and color-coded. The MS ranged from 87 to 5799 genes. Clustering
193 dendrogram of genes with module membership in colors is shown in **Fig. 4e**. Note the
194 colors are merely labels arbitrarily assigned to each models. For example, the Blue
195 Module in the SI study and the Blue Module in the colon study, although both color-
196 coded as blue, are not related to each other in any other aspects. The expression patterns
197 of each of the 13 modules is displayed in **Supplementary Fig. 5**. The modules in the
198 Colon network will hereby referred by Colon(Color).

199 **Modules significantly correlated with different VA status**

200 In the SI study, to determine if any of the nine modules of co-expressed genes
201 were associated with VA, we tested the correlations between the module eigengenes and
202 the VA status (**Fig. 4c).** Eigengene is defined as the first principal component of a given
203 module and may be considered as a representative of the gene expression profiles in that
204 module. Three modules were found significantly correlated with VA status. They are the
205 SI(Yellow), SI(Pink), and SI(Red) modules, with p values= 6×10^{-6}, 8×10^{-4} and
206 0.001, respectively. The SI(Yellow) module contains key players in RA metabolism
207 (Bco1, Scarb1, Gsta4, Aldh1a2, and Aldh1a3), potential RA targets (Rdh1, Aldh3a2,
208 Aldh7a1, and Aldh18a1), along with some immune cell markers (e.g. Cd1d1, Cd28,
209 Cd44, Cd86, and Cd207). In the SI(Red) module, genes previously reported as VA
210 targets were included (e.g. Isx, Ccr9, and Rarb). Some immunological genes in this
211 module may also be of interest (e.g. Lypd8, Cxcl14, Mmp9, Mbl2, Noxa1, Duoxa2,
212 Muc13, Itgae, Itgam, Madcam1, Il22ra2, Pigr, Clca1, Clca3a2, and Clca4b). In order to

213 visualize the modules of interest, in each of the three modules, the top 50 genes selected
214 by connectivity were depicted as networks in **Fig. 5**.
215 A similar diagram (**Fig. 4f**) was derived from the same analysis on the colon
216 dataset. Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) modules showed
217 significant correlation with VA status, with p values= 5×10^{-4}, 0.002, 0.004, and 0.004,
218 respectively.

219 **Module characterization**

220 Among the three SI modules significantly correlated with the VA status,
221 SI(Yellow) module is the largest one (MS=924). SI(Yellow) has a correlation coefficient
222 of -0.94 with VA status (**Fig. 4c**), meaning this module contains gene with low
223 expression levels in the VAS group (**Supplementary Fig. 4**). According to
224 ClusterProfiler, SI(Yellow) module mainly exhibited functional enrichment in adaptive
225 immunity and cytokine activity (**Supplementary Table 5**), meaning these activities were
226 reduced in the SI of VAS mice, compared with their VAD counterparts. In contrast,
227 SI(Red) module is positively correlated with VA status, and enriched for functional
228 categories involved in innate immune response (**Supplementary Table 5**). SI(Pink)
229 module is also positively correlated with VA status but no functional annotation is
230 enriched in this module. Overall, comparing VAS SI with VAD SI, the innate immune
231 activity is stronger, while adaptive immune response is weaker. On the other hand, in
232 colon, the Colon(Pink) and Colon(Yellow) were enriched in categories related to immune
233 functions (**Supplementary Table 6**). For the modules significantly correlated with VA

234 status, Colon(Green) was enriched for "cytoplasmic translation," "ribosome biogenesis,"
235 and "ncRNA processing." Colon(Red) module was enriched for "anion transmembrane
236 transporter activity."

237 To query whether the expression pattern of each module was mainly driven by
238 VA's regulatory effect on cell differentiation, a cell type marker enrichment was
239 performed. Criteria were set as the Bonferroni corrected p value (CP) of hypergeometric
240 test<0.05. This test identified four modules that significantly overlap with the list of SI
241 cell markers (**Fig. 6**). Yellow module, which is negatively correlated with VA (**Fig. 4c**),
242 contains 24 Tuft cell markers, reaching statistical significance (CP=4×10^{-7}), suggesting
243 a negative regulation of VA on the Tuft cell population. Green module also significantly
244 overlapped with Tuft cell markers (CP=0.01). Turquoise module, which is the largest
245 module with MS=5187 (**Supplementary Table 5**), was significantly enriched for Goblet
246 cell (CP=1×10^{-6}) and Paneth cell markers (CP=0.01).

247 **Module preservation between SI and colon**

248 14059 genes were robustly expressed in both the SI and colon datasets and were used for
249 the module preservation analysis (**Fig. 2**). Network construction were re-done in SI and
250 colon separately. All the modules from this analysis were referred as mpSI(Color) or
251 mpColon(Color). Module preservation analysis was performed on mpSI and mpColon
252 networks. The overall significance of the observed preservation statistics can be assessed
253 using Zsummary (**Supplementary Table 7**). Zsummary combines multiple preservation
254 Z statistics into a single overall measure of preservation [35]. All mpSI modules

255 demonstrated evidence of preservation in mpColon, except for the mpSI(Grey) module.
256 In addition, mpSI(Green) and mpSI(Turquoise) modules showed strong evidence of
257 preservation.

258 **Module overlap between SI and colon**

259 The contingency table (**Supplementary Fig. 6**) reports the number of genes that fall into
260 the modules of the SI network versus modules of the colon network, showing certain
261 level of agreement of the module assignment between the two organs. The SI modules
262 mpSI(Blue), mpSI(Red), mpSI(Turquoise) and mpSI(Green) have well-defined colon
263 counterparts highlighted in the red shades. However, the mpSI(Black) module appeared
264 not to be preserved in colon since it did not overlap significantly with any mpColon
265 module. Furthermore, many modules of interest in SI no longer have a single counterpart
266 module in the colon, but were instead mapped to multiple modules in the mpColon
267 network. For example, mpSI(Yellow), a module negatively correlated with VAS status,
268 appeared to split into three parts in the mpColon network: 55 genes were part of the
269 mpColon(Yellow) module, 9 fell in the mpColon(Tan) module, 202 genes were Blue in
270 the mpColon network.

271 **Discussion**

272 **Differential expression analysis confirmed known targets under the control of VA in**
273 **SI**

274 As expected, the 49 DEGs corresponding to VA effect in SI showed significant
275 enrichment in "retinoid metabolic process" (**Supplementary Fig. 3**), driven by three
276 DEGs Isx, Cyp26b1, and Bco1.

277 Cyp26b1 was one of the most significantly upregulated DEGs in SI in regards to
278 padj and Fold Change (**Fig. 3a** and **Supplementary Table 1**). CYP26 is a family of
279 cytochrome P450 enzymes. They are RA hydroxylases that catabolize RA and are
280 induced transcriptionally by RA. Therefore, they can determine the magnitude of cellular
281 exposure to RA by inactivating intracellular RA. In the gastrointestinal tract of rodents,
282 the regulation of Cyp26b1 gene expression is less well studied compared with that in the
283 liver and lung. Regarding autoregulation of RA concentration in rodent liver, animal
284 studies using diets containing different levels of VA have shown that RA induces hepatic
285 CYP26B1 expression, which in turn increases the rate of RA catabolism [36-38]. In neonatal
286 rat lung, CYP26B1 expression was higher and more responsive to retinoid supplements
287 compared with CYP26A1 [39]. Although many VA homeostasis genes are known to be
288 regulated by RA in a tissue-specific manners, in the SI study, a similar induction of
289 Cyp26b1 (14.12 fold) by VAS diet was observed. This is in line with the previous
290 discovery in our lab, where RA treatment strongly induced CYP26 mRNA expression in
291 SI [40]. Taken together, the induction of Cyp26b1 by RA has been shown in rodent liver,
292 lung, and SI.

293 Isx, an intestine-specific homeobox transcription factor, is another upregulated
294 DEGs ranked high in terms of padj and Fold Change in SI (**Fig. 3a** and **Supplementary**
295 **Table 1**). RA induces the expression of Isx via RARs [41]. ISX is also a transcriptional
296 repressor for BCO1 and SCARB1 in intestine. ISX binds to a conserved DNA-binding
297 motif upstream of the BCO1 gene and suppresses the expression of BCO1 [42], a VA
298 forming enzyme which converts absorbed beta-carotene to retinal. Additionally, ISX also
299 downregulates intestinal SCARB1 expression [42,43]. SCARB1 protein is a receptor
300 localized on the apical membrane of absorptive epithelial cells and can facilitate the
301 intestinal absorption of various essential dietary lipids including carotenoids and fat-
302 soluble vitamins [44]. In sum, the observation in the SI study confirmed the well-
303 established negative feedback regulation of dietary VA on the intestinal VA
304 concentration. In VAS groups, Isx mRNA was upregulated, while Scarb1 and Bco1 (two
305 of the DEGs with highest BaseMean values) were downregulated (**Fig. 3a** and
306 **Supplementary Table 2**), suggesting that the absorption and conversion of carotenoids
307 were reduced in response to ample dietary VA provided in the VAS diet.
308 Dendritic cells (DCs) are a heterogeneous population of antigen presenting cells
309 that differ in surface markers, localizations and functions. Cd207, also known as langerin,
310 is a C-type lectin mainly found in Langerhans cells and langerin^{+} DCs, two subtypes of
311 tissue-derived DCs [45]. Langerin^{+} DCs are normally only abundant in skin but not in
312 GALT. However, it has been reported that RA deficiency can change DC subtypes in
313 GALT and cause dramatic increase of langerin^{+} DCs [46,47], which is in concordance with
314 the DEG result of the SI study, where Cd207 expression was significantly higher in the
315 VAD groups (**Fig. 3a** and **Supplementary Table 2**).

316 **Potential new target genes regulated by VA in SI**

317 Mmp9 is an upregulated gene in the VAS group ranked top four in terms of both
318 padj value and Fold Change (**Fig. 3a** and **Supplementary Table 1**). This gene encodes a
319 member of the matrix metalloproteinase family of extracellular matrix-degrading
320 enzymes that are involved in numerous biological processes [48]. MMP9 has gelatinase
321 activity and plays important immunological roles in DCs and macrophages. For DCs, this
322 activity is essential for their migration to lymph nodes in order to activate T lymphocytes.
323 Mouse myeloid DC cultured with atRA displayed increased gelatinase activity and
324 reduced adherence [49]. In macrophages, MMP9 was found essential for forming
325 multinucleated giant cells, which are believed to enhance the defensive capacity of
326 macrophages [50]. In murine RAW264.7 and peritoneal macrophages, secreted MMP9
327 enhanced phagocytic activity of macrophages [51]. In ulcerative colitis models, MMP9 was
328 shown to regulate the release of pro-inflammatory cytokine such as KC [52,53], and mediate
329 increased colonic tight junction permeability via effect on myosin light chain kinase [54]. In
330 *C. rodentium* infection models, MMP9-/- mice had elevated levels of protective
331 segmented filamentous bacteria and interleukine-17, lower levels of *C. rodentium*, and a
332 smaller reduction in fecal microbial diversity in response to the enteric infection,
333 although the colonic epithelial cell hyperplasia and barrier dysfunction were similar
334 between genotypes [55]. The results of another study suggested that MMP9 initially
335 mediates inflammation, but later exerts a protective role on *C. rodentium* infection [56].
336 Despite the clear evidence on Mmp9's modulatory role on mucosal immunity, this is the
337 first report on the upregulation of Mmp9 by VA on the whole tissue level in the SI.

338 Previous studies have mainly focused on how VA regulate Mmp9 expression in
339 macrophages and DCs only, but not the epithelial cells. Increased expression of MMPs
340 −1, -2, -3, -8, -9, and −12, have all been associated with IBD [55]. In murine macrophages,
341 the carotenoid lutein was shown to enhance Mmp9 transcription through intracellular
342 ROS generation, ERK1/2, p38 MAPK, and RARbeta activation [51]. Furthermore,
343 luciferase assays demonstrated that different vitamin A-related substances induced Mmp9
344 production via different mechanisms- atRA induced MMP-9 production via RARalpha
345 activation whereas retinol and beta-carotene caused MMP-9 production via RARalpha
346 and beta activation. In mouse myeloid DCs, it was discovered by chromatin
347 immunoprecipitation assays that Mmp-9 expression was enhanced by atRA through a
348 transcriptional mechanism involving greater RARalpha promoter binding, recruitment of
349 p300, and subsequent histone H3 acetylation, despite the absence of a canonical RARE in
350 the promoter region of the Mmp9 gene [49]. On the other hand, studies based on ulcerative
351 colitis models suggested that MMP9 can be expressed by both infiltrating leukocytes and
352 by the colonic epithelium [53,54,57,58]. Therefore, the induction of Mmp9 by VA in intestinal
353 epithelial cells warrants further study.
354 Mbl2 is another active player in gut immunity, the expression of which may also
355 be controlled by VA in SI. Mannose-binding lectin (MBL) is an important pattern-
356 recognition receptor (PRR) in the innate immune system. MBLs are members of the
357 collectin gene family which belongs to the C-type lectin superfamily [59]. MBLs contain a
358 C-terminal carbohydrate recognition domain, a collagen-like region, a cysteine-rich N-
359 terminal domain and a neck region. CRD has been shown to bind different sugars, such
360 as mannan, mannose, and fucose. With those biochemical entities, MBLs are protective

361 against gut infection by selectively recognizing and binding sugars presented on the
362 pathogens [60]. This binding can neutralize the pathogen and inhibit infection by
363 complement activation through opsonization by collectin receptors or through the lectin
364 pathway. MBL may also protect the intestinal mucosa independently of the complement
365 pathway by direct interaction with *Shigella flexneri 2a* [61]. Low levels of MBL have been
366 related to increased susceptibility to several pathogens such as *Neisseria meningitidis,*
367 *Staphylococcus aureus, Streptococcus pneumoniae,* and *S. aureus* [62-65]. In regards of the
368 transcriptional regulation of MBLs, this protein is mainly synthesized by hepatocytes and
369 present in liver and serum. SI is a predominant site of extrahepatic expression of MBL.
370 Allelic variants of MBL2 have been associated with deficiencies in MBL protein
371 production [66], which may increase the risk of mother-to-child HIV transmission.
372 However, among infants with MBL2 variants, VA plus β-carotene supplementation
373 partially counteracted the increased risk of transmission [67]. Currently, no mechanistic
374 study has investigated how VA affects MBL2 gene expression; however, since Mbl2 is
375 the fourth upregulated gene in the VAS group in terms of significance (padj value) and
376 the third upregulated DEG in terms of Fold Change (**Fig. 3a** and **Supplementary Table**
377 **1**), we postulate that VA may be a regulator of MBL2 gene expression. If so, it will
378 provide another line of evidence that VA participates in the regulation of the innate
379 mucosal immunity.

380 Nr0b2 (nuclear receptor subfamily 0, group B, member 2), also known as small
381 heterodimer partner (SHP), is an atypical nuclear receptor that lacks a conventional DNA
382 binding domain and binds to a variety of nuclear receptors [68]. In liver, SHP plays an
383 important role in controlling cholesterol, bile acid, and glucose metabolism. Hepatic SHP

384 has been shown to be retinoid responsive both *in vitro* and *in vivo* [69,70]. In AML12 cells (a
385 hepatocyte cell line), SHP was upregulated by atRA in a dose-dependent manner,
386 independent of farnesoid X receptor (FXR) [69,71]. However, in the SI study, SHP appears
387 to be downregulated by VA in SI (**Fig. 3a** and **Supplementary Table 2**). This may
388 reflect the organ specificity of the transcriptional regulation on SHP gene, and may
389 reflect the different roles played by liver and SI in cholesterol metabolism. Nr0b2 ranks
390 relatively high regarding connectivity in SI(Yellow) module (**Fig. 5a**), which is in line
391 with its annotation as a TF.

392 **Inhibition of mast cell population by VA in SI**

393 Mcpt4 (mast cell protease 4) is a downregulated DEG ranked top three in terms of
394 padj and Fold Change (**Fig. 3a** and **Supplementary Table 2**). Mcpt4 is a murine
395 functional homologue to human chymase [72,73]. Mcpt4 cleaves several substrates
396 important for ECM degradation and tissue remodeling [74]. Mcpt4 is expressed
397 predominantly in mast cells. Mast cells are effector cells important for both innate and
398 adaptive immunity. Mast cell activation induces degranulation and release of
399 inflammatory mediators, including histamine, cytokines, and proteases. Many of those
400 inflammatory effectors play key roles in various allergic reactions. Intestinal mast cells
401 are localized primarily to the mucosal epithelium and lamina propria (LP) and express the
402 chymases Mcpt1 and Mcpt2. On the other hand, the connective tissue subtype of mast
403 cells localizes primarily within the submucosa and expresses the chymases (Mcpt4 and
404 Mcpt5), the tryptases (Mcpt6 and Mcpt7), and CPA3 (carboxypeptidase A3) [75]. Mcpt1, 2

405 and 6 (also known as Tpsb2 or tryptase beta 2), three additional downregulated DEGs
406 (**Supplementary Table 2**), together with Mcpt4, not only drove the significant
407 enrichment in "serine-type endopeptidase activity," "serine-type peptidase activity" and
408 "serine hydrolase activity" (**Fig. 3d**), but were also clustered in the same SI(Yellow)
409 module (**Supplementary Table 5**). One of the features of robust modules of co-
410 expressed genes is that a module may reflect specific cell types and biological functions.
411 Thus it is possible to obtain cell type-specific information from the analysis of whole
412 tissue without isolating homogeneous population of cells [76,77]. Altogether, this, for the
413 first time, suggests that mast cells were inhibited by VAS status in SI. This is in line with
414 previous observations. Retinol and RA were shown to inhibit the growth and
415 differentiation of human leukemic mast cells (HMC-1) [78], and the mast cells derived from
416 peripheral blood or cord blood samples [79,80]. Peritoneal mast cell numbers were also
417 reported to be 1.5 times higher in VAD rat [81].

418 **Co-expression analysis has a different emphasis compared with differential**
419 **expression**

420 The three modules significantly correlated with VA status (Yellow, Red, and
421 Pink) in SI were identified as modules of interest. Those modules were not only enriched
422 for immune functions, but also contained an enriched number of DEGs (**Supplementary**
423 **Table 5**). However, DEGs may not always be central in the co-expression network.
424 Those three modules contain 40 DEGs, 81.6% of the 49 DEGs. Taking the SI(Yellow)
425 module as an example, it contains 924 genes, only 6.4% of the total 14368 genes, but it

426 contains 30 out of the 33 downregulated DEGs (**Supplementary Table 5**). This can be
427 anticipated because SI(Yellow) was found to be negatively correlated with VAS
428 condition (**Fig. 4c**), with expression patterns separated clearly according to VA status
429 (**Supplementary Fig. 4**). However, those DEGs that were identified under relatively
430 stringent criteria (adj $p<0.05$, $|FoldChange|>2$), are not necessarily of high connectivity.
431 In fact, none of the DEGs visualized in **Fig. 5** was a DEG ranked towards the top in terms
432 of p value or Fold Change, indicating that differential expression and co-expression
433 analyses emphasize on two distinct aspects of gene expression. In other words, the
434 identification of DEGs with high significance do not necessarily indicate that they play
435 important regulatory roles.

436 **Implications of results for an indirect inhibition of VA on the SI Tuft cell population**

437 Tuft cells (also known as brush or caveolated cells) have distinctive apical bristles
438 that form a highly organized brush border, giving the cells their eponymous tufted
439 morphology. Tuft cells represent about 0.5% of the epithelial cells in the murine small
440 and large intestines, with a slightly more prevalent distribution in the distal SI [82].
441 Although the existence of Tuft cell has been known in multiple organs for close to a
442 century, not until recently were they distinguished from enteroendocrine, Paneth, and
443 goblets cells, and raising from three to four the number of secretory cell types in the SI
444 epithelium [83].
445 Tuft cell markers were found significantly enriched in SI(Yellow) and SI(Green)
446 modules (**Fig. 6**). SI(Yellow) negatively correlates with VA status, suggestive of gene

447 downregulation by VA in this module (**Fig.4c** and **Supplementary Fig. 4**). Among all of
448 the 24 Tuft cell markers contained in SI(Yellow), Krt18, also known as CK18, encodes a
449 structural protein. Also in the SI(Yellow), Trpm5 is a well-known Tuft cell marker,
450 encoding a chemosensory protein that regulates ion transport [82]. Lastly, another Tuft
451 marker Ptpn6 [84] is a top 5% hub gene of the SI(Yellow) module (**Fig 5a** and
452 **Supplementary Table 5**). In sum, it is inferred that Tufts cell population in the SI is
453 higher under VAD status.

454 As an understudied cell type, no literature can be found on the direct regulation of
455 VA on Tuft cell differentiation. Based on our literature review, the speculation is that the
456 VAD status is associated with higher ILC2 population, which stimulates Tuft cell
457 differentiation. VA regulates the balance between ILC2s and ILC3s to protect the
458 intestinal mucosa. In the VA deficiency state, the proliferation and function of the ILC3
459 subset are suppressed and there are marked increases of ILC2 cell proliferation and type 2
460 cytokine production [9]. IL-13 is one of those cytokines that can stimulate the
461 differentiation of epithelial precursor cells towards Tuft cells [82].

462 **A comparison between SI and colon regarding the transcriptional profiles**
463 **influenced by VA**

464 Compared with the SI, the colon had higher number of DEGs but fewer
465 enrichment categories corresponding to the VA effect. In the colon, 94 DEGs were
466 identified but they were only enriched for GO terms related to cell division, suggesting
467 that the primary role VA played in colon was inhibiting cell proliferation and inducing

468 epithelial differentiation. For example, Foxm1 (forhead box M1) is a downregulated
469 DEG in colon (**Supplementary Table 4**). It is a typical proliferation-associated
470 transcription factor which plays a key role in the cell cycle. FOXM1 stimulates
471 proliferation by promoting S-phase entry and M-phase entry. FOXM1 also regulates
472 genes essential for proper mitosis [85]. Foxm1 was present in two of the GO categories
473 "regulation of cell cycle process," and "negative regulation of cell cycle" where
474 downregulated DEGs in colon were significantly enriched. This, together with the
475 upregulation of goblet cell markers (Tff3, Muc1, Clca1, and Clca4b), is consistent with
476 the differentiation-promoting effect of VA (**Supplementary Table 3**).
477 Unlike the colon, there were many more enriched GO terms and DEGs with
478 changes in extremely high magnitudes in the SI. The 49 DEGs in SI, although much
479 fewer than that in colon, not only included differentiated epithelial cell markers
480 (**Supplementary Table 2**), but also enriched in immunological and retinoid metabolic
481 pathways (**Fig. 3 c,d** and **Supplementary Fig. 3**). Some of the DEGs responded to VA
482 status in extreme magnitudes because they were involved in VA metabolism. For
483 example, Isx (Fold Change=3231.03), Cyp26b1 (Fold Change=14.12) and Bco1 (Fold
484 Change= -8.66) (**Supplementary Table 1** and **2**). Additionally, even though Isx is a
485 DEG corresponding to VA effect in colon as well as SI, Bco1 and Scarb1 were not
486 differentially expressed in colon, suggesting that colon is not a major site for dietary VA
487 to regulate carotenoid conversion and absorption; this is in line with the fact that the
488 digestion and absorption of carotenoids primarily takes place in SI. In sum, besides cell
489 differentiation, the immune functions and retinoid metabolism are also tightly controlled
490 by VA in the SI.

From the co-expression network perspective, the transcriptional profile of the two organs showed moderate levels of similarities. Most genes were commonly expressed in both SI and colon. Module preservation and module overlap analyses demonstrated that counterpart co-expression modules can be traced between the two organs. However, most of those modules only had moderate preservations due to their distinct structures (**Supplementary Fig. 6** and **Supplementary Table 7**). This is in line with the fact that even though SI and colon are consecutive organs with similar developmental origins, their biological functions distinct with each other.

In terms of the similarities on the DEG level, CXCL14 is an upregulated DEG under the VAS condition in both SI and colon (**Fig. 3a, b** and **Supplementary Table 1, 3**). CXCL14, originally identified as BRAK, BMAC, or Mip-2γ, belongs to the CXC chemokine family. CXCL14 is related to anti-microbial and chemoattractive activities. On one hand, in skin epithelial cells, CXCL14 can maintain the skin integrity through its broad-spectrum bactericidal activity, which can regulate the balance between commensal and pathogenic bacteria [86,87]. However, in lungs, CXCL14 did not play a pivotal role during *E.coli* or influenza infections [88]. Although it is expressed in high level in SI [89], an antibacterial role for CXCL14 in the SI has not yet been investigated. On the other hand, CXCL14 is a chemoattractant for activated macrophages, immature dendritic cells, and natural killer cells [89]. *In vitro*, eukaryotic CXCL14 was found as a potent chemoattractant also for neutrophils [90]. Snyder et al. [91] discovered that RA signaling is required in macrophages ($CD4^-$, $CD11b^+$, $F480^+$) and neutrophils ($CD4^-$, $CD11b^+$, $F480^-$) for the complete clearance of *C. rodentium*. Therefore we postulate that RA may be able to

513 induce Cxcl14 expression, and subsequently change the migration and/or function
514 (especially IL-17 production) of the gut macrophages and neutrophils.

515 **Conclusions**

516 Taken together, the results of the analyses presented in this paper suggest the
517 following: DEGs corresponding to VA effect were present in both the SI and colon. In SI,
518 DEGs altered by VA were mainly involved in the retinoid metabolic pathway and
519 immunity-related pathways. Novel target genes (e.g. Mbl2, Mmp9, Cxcl14, and Nr0b2)
520 and cell types (based on mast cell and Tuft cell markers) under the regulation of VA were
521 suggested by differential expression analysis coupled with the co-expression network
522 analysis. In regard to the colon, VA demonstrated an overall suppressive effect on the cell
523 division pathway. Finally, the comparison of co-expression modules between SI and
524 colon indicated distinct regulatory networks in these two organs.

525 **Methods**

526 **Animals.** C57BL/6 mice, originally purchased from Jackson Laboratories (Bar
527 Harbor, MN), were bred and maintained at the Pennsylvania State University (University
528 Park, PA) for experiments. Mice were exposed to a 12 hour light, 12 hour dark cycle with
529 *ad libitum* access to food and water. All animal experiments were performed according to
530 the Pennsylvania State University's Institutional Animal Care and Use Committee
531 (IACUC) guideline (IACUC # 43445). Vitamin A deficient (VAD) and vitamin A

532 sufficient (VAS) mice were generated as previously described [23,27,28,91]. Briefly, VAS or
533 VAD diets were fed to pregnant mothers and the weanlings. VAS diet contained 25 μg of
534 retinyl acetate per day whereas VAD diet did not contain any VA. At weaning, mice were
535 maintained on their respective diets until the end of the experiments. The SI study lasted
536 5 days following *C. rodentium* infection, whereas the colon study ended during the peak
537 of infection (post-infection day 10). Serum retinol concentrations were measured by
538 UPLC (Ultra Performance Liquid Chromatography) to verify the VA status of
539 experimental animals.

540 **UPLC serum retinol quantification.** Serum samples were saponified and
541 quantified for total retinol concentration by ultra-performance liquid chromatography
542 (UPLC) [92]. Briefly, serum aliquots (30–100 μl) were incubated for 1 h in ethanol. Then
543 5% potassium hydroxide and 1% pyrogallol were added. After saponification in a 55°C
544 water bath, hexanes (containing 0.1% butylated hydroxytoluene, BHT) and deionized
545 water were added to each sample for phase separation. The total lipid extract was
546 transferred and mixed with a known amount of internal standard,
547 trimethylmethoxyphenyl-retinol (TMMP-retinol). Samples were dried under nitrogen and
548 re-constituted in methanol for UPLC analysis. The serum total retinol concentrations
549 were calculated based on the area under the curve relative to that of the internal standard.

550 **Statistical analysis.** Statistical analyses were performed using GraphPad Prism
551 software (GraphPad, La Kolla, CA, USA). Two-way analysis of variance (ANOVA) with
552 Bonferroni's *post-hoc* tests were used to compare serum retinol levels. $P<0.05$ was used
553 as the cut off for a significant change.

554 **Tissue collection and RNA extraction.** Tissue were collected at the end of each
555 study. During tissue collection, the dissected colons were first cut open to remove fecal
556 contents. Both lower SI and colon tissues were snap frozen in liquid nitrogen
557 immediately upon tissue harvest. In the SI study, lower SI tissue was used for total RNA
558 extraction. In the colon study, total RNA was extracted from the whole colon tissue free
559 of colonic content. Total RNA extractions were performed using the Qiagen RNeasy
560 Midi Kit according to the manufacturer's protocol. RNA concentrations were measured
561 by Nanodrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE).
562 TURBO DNA-free kit (Ambion, Austin, USA) were used to remove genomic DNA.
563 RNA quality, assessed by RNA Integrity Number (RIN), were determined by Agilent
564 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). RNA samples of sufficient
565 quality (RIN > 8) were used to construct RNA-seq libraries.
566 **RNAseq library preparation, sequencing and mapping**. Library preparation
567 and sequencing were done in the Penn State Genomics Core Facility (University Park,
568 PA). 200ng of RNA per sample were used as the input of TruSeq Stranded mRNA
569 Library Prep kit to make barcoded libraries according to the manufacturer's protocol
570 (Illumina Inc. San Diego. CA. USA). The library concentration was determined by qPCR
571 using the Library Quantification Kit Illumina Platforms (Kapa Biosystems. Boston. MA.
572 USA). An equimolar pool of the barcoded libraries was made and the pool was
573 sequencing on a HiSeq 2500 system (Illumina) in Rapid Run mode using single ended
574 150 base-pair reads. 30-37 million reads were acquired for each library. Mapping were
575 done in the Penn State Bioinformatics Consulting Center. Quality trimming and adapter
576 removal was performed using trimmomatic [93]. The trimmed data were then mapped to the

577 mouse genome (mm10, NCBI) using hisat2 alignment program. Hisat2 alignment was
578 performed by specifying the 'rna-strandness' parameter. Default settings were used for all
579 the other parameters [94]. Coverages were obtained using bedtools [95]. Mapped data was
580 visualized and checked with IGV [96]. Reads mapped to each transcript were counted using
581 featureCounts [97].

582 **Differential expression**. Transcripts with low expression levels were first
583 removed, to assure that data analysis would be focused on tissue-specific mRNAs and not
584 disrupted by genes not expressed in the organ. In the SI study, if the expression level of a
585 gene was lower than 10 in more than eight of the samples, or if the sum of the expression
586 level in all 12 samples was lower than 200, this transcript was removed from the later
587 analysis. In the colon study, if the expression level was lower than 10 in more than 10 of
588 the samples, or if the sum of the expression level in all 16 samples was lower than 220,
589 the transcript was regarded as lowly expressed gene and removed. Differential expression
590 (DE) analysis were performed using the DESeq2 package [98]. Because of the two-by-two
591 factorial design, the effect of VA, *C. rodentium* infection and VA x *C. rodentium*
592 Interaction were examined separately. This article focused on the VA effect, with |Fold
593 change|> 2 and adjusted $p<0.05$ as the criteria for differentially expressed genes (DEGs).
594 Heatmaps and volcano plots were depicted to visualize the result through the R packages
595 pheatmap, ggplot, and ggrepel.

596 **WGCNA.** For consistency and comparability, the same data matrix in differential
597 expression were used to build signed co-expression networks using WGCNA package in
598 R [99]. In other words, the dataset went through the same screening and normalization
599 process as the differential expression analysis (**Fig. 2**). For each set of genes, a pair-wise

600 correlation matrix was computed, and an adjacency matrix was calculated by raising the
601 correlation matrix to a power. A parameter named Softpower was chosen using the scale-
602 free topology criterion [100]. An advantage of weighted correlation networks is the fact that
603 the results are highly robust with respect to the choice of the power parameter. For each
604 pairs of genes, a robust measure of network interconnectedness (topological overlap
605 measure) was calculated based on the adjacency matrix. The topological overlap (TO)
606 based dissimilarity was then used as input for average linkage hierarchical clustering.
607 Finally, modules were defined as branches of the resulting clustering tree. To cut the
608 branches, a hybrid dynamic tree-cutting algorithm was used [101]. To obtain moderately
609 large and distinct modules, we set the minimum module size to 30 genes and the
610 minimum height for merging modules at 0.25. As per the standard of WGCNA tool,
611 modules were referred to by their color labels henceforth and marked as SI(Color) or
612 Colon(Color). Genes that did not belong to any modules were assigned to the Grey
613 module. Color assignment together with hierarchical clustering dendrogram were
614 depicted. To visualize the expression pattern of each module, heatmap were drawn using
615 pheatmap package in R. Eigengenes were computed and used as a representation of
616 individual modules. Eigengenes are the first principal component of each set of module
617 transcripts, describing most of the variance in the module gene expression. Eigengenes
618 provides a coarse-grained representation of the entire module as a single entity [102]. To
619 identify modules of interest, module-trait relationships were analyzed. The modules were
620 tested for their associations with the traits by correlating module eigengenes with trait
621 measurements, i.e., categorical traits (VA status, infection status, and gender) and
622 numerical traits (shedding, body weight, body weight change percentage, and RIN score).

623 The correlation coefficient and significance of each correlation between module
624 eigengenes and traits were visualized in a labeled heatmap, color-coded by blue (negative
625 correlation) and red (positive correlation) shades. Because the present paper has a focus
626 on the VA effect, the only trait reported was the VA status. The correlation results from
627 all other traits were beyond the scope of this article and will be discussed elsewhere.

628 **Functional enrichment analyses.** Gene ontology and KEGG enrichment were
629 assessed using clusterProfiler (v3.6.0) in R [103]. For DEG list and co-expression modules,
630 the background was set as the total list of genes expressed in the SI (14368 genes) or
631 colon (15340 genes), respectively. The statistical significance threshold level for all GO
632 enrichment analyses was p.adj (Benjamini and Hochberg adjusted for multiple
633 comparisons) <0.05.

634 **Cell type marker enrichment.** The significance of cell type enrichment was
635 assessed for each module in a given network with a cell type marker list of the same
636 organ, using a one-sided hypergeometric test (as we were interested in testing for over-
637 representation). To account for multiple comparisons, a Bonferroni correction was
638 applied on the basis of the number of cell types in the organ-specific cell type marker list.
639 The criterion for a module to significantly overlap with a cell marker list was set as a
640 corrected hypergeometric P value < 0.05 [76,77,104]. The cell type marker lists of SI and
641 colon were obtained from CellMarker database (http://biocc.hrbmu.edu.cn/CellMarker/),
642 a manually curated resource of cell markers [105].

643 **Module visualization.** A package named igraph was used to visualize the co-
644 expression network of genes in each individual modules based on the adjacency matrix
645 computed by WGCNA. The connectivity of a gene is defined as the sum of connection

646 strength (adjacency) with the other module members. In this way, the connectivity
647 measures how correlated a gene is with all other genes in the same module [35]. The top 50
648 genes in each module ranked by connectivity were used for the visualization. In order to
649 further improve network visibility, adjacencies equal to 1 (adjacencies of a gene to itself)
650 or lower than 0.7 (less important co-expression relationships) were not shown in the
651 diagram. Each node represents a gene. The width of the lines between any two nodes was
652 scaled based on the adjacency of this gene pair. The size of the node was set proportional
653 to the connectivity of this gene. For the shape of the nodes, squares were used to
654 represent all the mouse transcription factors (TFs), based on AnimalTFDB 3.0 database
655 (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/) [106]. Circles represent all the non-TFs. In
656 terms of the color of the nodes, upregulated DEGs in the VAS condition were marked
657 with red color and downregulated DEGs in blue. For the font of the gene labels, red bold
658 italic was used to highlight genes which are cell type markers according to CellMarker
659 database (http://biocc.hrbmu.edu.cn/CellMarker/) [105].

660 **Module preservation.** Genes with low expression (based on the criteria stated in
661 Differential Expression session) in the SI or Colon datasets were excluded from this
662 analysis. In other words, only the probes that were robustly expressed in both organs
663 were used for module preservation analysis. Therefore, reconstruction of co-expression
664 networks were done in both datasets separately, with minimum module size = 30 and the
665 minimum height for merging modules = 0.25 [99-101]. All the modules from this analysis
666 were marked as mpSI(Color) or mpColon(Color).

667 Color assignment was remapped so that the corresponding modules in the SI
668 network and the re-constructed mpSI network sharing the similar expression pattern were

669 assigned with the same color. Similarly, the modules in the re-constructed mpColon
670 network were also assigned same colors according to the expression pattern resemblance
671 with their corresponding Colon modules. In order to quantify network preservation
672 between mpSI and mpColon networks at the module level, permutation statistics is
673 calculated in R function modulePreservation, which is included in the updated WGCNA
674 package [35]. Briefly, module labels in the test network (mpColon network) were randomly
675 permuted for 50 times and the corresponding preservation statistics were computed,
676 based on which Zsummary, a composite preservation statistic, were calculated. The
677 higher the value of Zsummary, the more preserved the module is between the two
678 datasets. As for threshold, if Zsummary>10, there is strong evidence that the module is
679 preserved. If Zsummary is >2 but <10, there is weak to moderate evidence of
680 preservation. If Zsummary<2, there is no evidence that the module is preserved [35]. The
681 Grey module contains uncharacterized genes while the Gold module contains random
682 genes. In general, those two modules should have lower Z-scores than most of the other
683 modules.

684 **Module overlap.** Module overlap is the number of common genes between one
685 module and a different module. The significance of module overlap can be measured
686 using the hypergeometric test. In order to quantify module comparison, in other words, to
687 test if any of the modules in the test network (mpColon network) significantly overlaps
688 with the modules in the reference network (mpSI network), the overlap between all
689 possible pairs of modules was calculated along with the probability of observing such
690 overlaps by chance. The significance of module overlap was assessed for each module in
691 a given network with all modules in the comparison network using a one-sided

692 hypergeometric test (as we were interested in testing for over-representation) [77]. To
693 account for multiple comparisons, a Bonferroni correction was applied on the basis of the
694 number of modules in the comparison network (mpColon network). Module pair with
695 corrected hypergeometric P value < 0.05 was considered a significant overlap.
696
697
698
699
700 **Acknowledgements**

701 We would like to acknowledge the sequencing service performed by C. Praul and
702 colleagues in the Genomic Core Facility at the Pennsylvania State University. We are
703 grateful to A. Sebastian and I. Albert from the Bioinformatics Consulting Center at Penn
704 State for their help on mapping and data pre-processing. The study was supported by the
705 research funding (R56 AI114972) provided by the National Institute of Health (NIH) and
706 the training grant (T32GM108563) provided by National Institute of General Medical
707 Sciences (NIGMS). The funders had no role in the study design, data collection and
708 analysis, decision to publish, or preparation the manuscript.
709
710 **Author Contributions**

711 A.C.R., M.T.C and Z.C. conceived the study, designed the experiments, and made major
712 contributions to drafting the manuscript. A.C.R. and M.T.C supervised the projects and
713 organized the coworkers. Z.C. performed the most of the experiments and data analyses.
714 Y.L. and Q.L. provided guidance on the selection and application of bioinformatics tools.

715 C.W. measured the serum retinol level using UPLC. Q.C., L.S. and V.W. assisted on the
716 animal studies and provided valuable research thoughts. All authors participated in the
717 revision and approved the final manuscript.
718
719 **Competing interests**
720 The authors declare no competing interests.
721

References

1 West, K. P., Jr. Extent of vitamin A deficiency among preschool children and women of reproductive age. *The Journal of nutrition* **132**, 2857s-2866s (2002).

2 Glasziou, P. P. & Mackerras, D. E. Vitamin A supplementation in infectious diseases: a meta-analysis. *BMJ (Clinical research ed.)* **306**, 366-370 (1993).

3 Coutsoudis, A., Broughton, M. & Coovadia, H. M. Vitamin A supplementation reduces measles morbidity in young African children: a randomized, placebo-controlled, double-blind trial. *Am J Clin Nutr* **54**, 890-895 (1991).

4 Glasziou, P. & Mackerras, D. Vitamin A supplementation in infectious diseases: a meta-analysis. *BMJ (Clinical research ed.)* **306**, 366-370 (1993).

5 Xiao, S. *et al.* Retinoic acid increases Foxp3+ regulatory T cells and inhibits development of Th17 cells by enhancing TGF-β-driven Smad3 signaling and inhibiting IL-6 and IL-23 receptor expression. *The Journal of Immunology* **181**, 2277-2284 (2008).

6 Cantorna, M. T., Nashold, F. E. & Hayes, C. E. In vitamin A deficiency multiple mechanisms establish a regulatory T helper cell imbalance with excess Th1 and insufficient Th2 function. *The Journal of Immunology* **152**, 1515-1522 (1994).

7 Hammerschmidt, S. I. *et al.* Retinoic acid induces homing of protective T and B cells to the gut after subcutaneous immunization in mice. *The Journal of clinical investigation* **121**, 3051-3061 (2011).

8 Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527-538 (2004).

9 Sirisinha, S. The pleiotropic role of vitamin A in regulating mucosal immunity. *Asian Pac J Allergy Immunol* **33**, 71-89 (2015).

10 Choi, J. Y., Cho, K. N. & Yoon, J. H. All-trans retinoic acid induces mucociliary differentiation in a human cholesteatoma epithelial cell culture. *Acta Otolaryngol* **124**, 30-35 (2004).

11 Goto, Y. & Ivanov, I. I. Intestinal epithelial cells as mediators of the commensal–host immune crosstalk. *Immunology and cell biology* **91**, 204-214 (2013).

12 Cha, H. R. *et al.* Downregulation of Th17 cells in the small intestine by disruption of gut flora in the absence of retinoic acid. *Journal of immunology (Baltimore, Md. : 1950)* **184**, 6799-6806, doi:10.4049/jimmunol.0902944 (2010).

13 Amit-Romach, E., Uni, Z., Cheled, S., Berkovich, Z. & Reifen, R. Bacterial population and innate immunity-related genes in rat gastrointestinal tract are altered by vitamin A-deficient diet. *J Nutr Biochem* **20**, 70-77, doi:10.1016/j.jnutbio.2008.01.002 (2009).

14 Koch, M. J. *et al.* Crystalloid lysozyme inclusions in Paneth cells of vitamin A-deficient rats. *Cell Tissue Res* **260**, 625-628 (1990).

15 Olson, J. A., Rojanapo, W. & Lamb, A. J. THE EFFECT OF VITAMIN A STATUS ON THE DIFFERENTIATION AND FUNCTION OF GOBLET CELLS IN THE RAT INTESTINE*. *Annals of the New York Academy of Sciences* **359**, 181-191, doi:10.1111/j.1749-6632.1981.tb12746.x (1981).

16 De Luca, L., Little, E. P. & Wolf, G. Vitamin A and protein synthesis by rat intestinal mucosa. *J Biol Chem* **244**, 701-708 (1969).

17 Ahmed, F., Jones, D. B. & Jackson, A. A. The interaction of vitamin A deficiency and rotavirus infection in the mouse. *Br J Nutr* **63**, 363-373 (1990).

18 Uni, Z. *et al.* Vitamin A deficiency interferes with proliferation and maturation of cells in the chicken small intestine. *Br Poult Sci* **41**, 410-415, doi:10.1080/713654958 (2000).

19 Osanai, M. *et al.* Cellular retinoic acid bioavailability determines epithelial integrity: Role of retinoic acid receptor alpha agonists in colitis. *Molecular pharmacology* **71**, 250-258, doi:10.1124/mol.106.029579 (2007).
20 Collins, J. W. *et al.* Citrobacter rodentium: infection, inflammation and the microbiota. *Nat Rev Microbiol* **12**, 612-623, doi:10.1038/nrmicro3315 (2014).
21 Mangan, P. R. *et al.* Transforming growth factor-β induces development of the T H 17 lineage. *Nature* **441**, 231 (2006).
22 Sonnenberg, G. F., Monticelli, L. A., Elloso, M. M., Fouser, L. A. & Artis, D. CD4+ lymphoid tissue-inducer cells promote innate immunity in the gut. *Immunity* **34**, 122-134 (2011).
23 McDaniel, K. L. *et al.* Vitamin A-deficient hosts become nonsymptomatic reservoirs of Escherichia coli-like enteric infections. *Infection and immunity* **83**, 2984-2991 (2015).
24 Villablanca, E. J. *et al.* MyD88 and retinoic acid signaling pathways interact to modulate gastrointestinal activities of dendritic cells. *Gastroenterology* **141**, 176-185, doi:10.1053/j.gastro.2011.04.010 (2011).
25 Jaensson-Gyllenback, E. *et al.* Bile retinoids imprint intestinal CD103+ dendritic cells with the ability to generate gut-tropic T cells. *Mucosal Immunol* **4**, 438-447, doi:10.1038/mi.2010.91 (2011).
26 Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *Journal of Lipid Research* **43**, 1773-1808, doi:10.1194/jlr.R100015-JLR200 (2002).
27 Carman, J., Smith, S. & Hayes, C. Characterization of a helper T lymphocyte defect in vitamin A-deficient mice. *The Journal of Immunology* **142**, 388-393 (1989).
28 Smith, S. M., Levy, N. S. & Hayes, C. E. Impaired immunity in vitamin A-deficient mice. *The Journal of nutrition* **117**, 857-865 (1987).
29 Wiede, A. *et al.* Localization of TFF3, a new mucus-associated peptide of the human respiratory tract. *American journal of respiratory and critical care medicine* **159**, 1330-1335 (1999).
30 Dietert, K., Reppe, K., Mundhenk, L., Witzenrath, M. & Gruber, A. D. mCLCA3 modulates IL-17 and CXCL-1 induction and leukocyte recruitment in murine Staphylococcus aureus pneumonia. *PLoS One* **9**, e102606 (2014).
31 Erickson, N. A. *et al.* The Goblet Cell Protein Clca1 (Alias mClca3 or Gob-5) Is Not Required for Intestinal Mucus Synthesis, Structure and Barrier Function in Naive or DSS-Challenged Mice. *PLoS One* **10**, e0131991, doi:10.1371/journal.pone.0131991 (2015).
32 Erickson, N. A. *et al.* Soluble mucus component CLCA1 modulates expression of leukotactic cytokines and BPIFA1 in murine alveolar macrophages but not in bone marrow-derived macrophages. *Histochemistry and Cell Biology* **149**, 619-633, doi:10.1007/s00418-018-1664-y (2018).
33 Ching, J. C., Lobanova, L. & Loewen, M. E. Secreted hCLCA1 is a signaling molecule that activates airway macrophages. *PLoS One* **8**, e83130 (2013).
34 Nyström, E. E. *et al.* Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. *EBioMedicine* (2018).
35 Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS computational biology* **7**, e1001057 (2011).
36 Yamamoto, Y., Zolfaghari, R. & Ross, A. C. Regulation of CYP26 (cytochrome P450RAI) mRNA expression and retinoic acid metabolism by retinoids and dietary vitamin A in liver of mice and rats. *The FASEB Journal* **14**, 2119-2127 (2000).

37 Ross, A. C. Retinoid production and catabolism: role of diet in regulating retinol esterification and retinoic acid oxidation. *The Journal of nutrition* **133**, 291S-296S (2003).

38 Ross, A. C., Cifelli, C. J., Zolfaghari, R. & Li, N.-q. Multiple cytochrome P-450 genes are concomitantly regulated by vitamin A under steady-state conditions and by retinoic acid during hepatic first-pass metabolism. *Physiological genomics* **43**, 57-67 (2010).

39 Wu, L. & Ross, A. C. Acidic retinoids synergize with vitamin A to enhance retinol uptake and STRA6, LRAT, and CYP26B1 expression in neonatal lung. *Journal of lipid research* **51**, 378-387 (2010).

40 Wang, Y., Zolfaghari, R. & Ross, A. C. Cloning of rat cytochrome P450RAI (CYP26) cDNA and regulation of its gene expression by all-trans-retinoic acid in vivo. *Archives of biochemistry and biophysics* **401**, 235-243 (2002).

41 Lobo, G. P. *et al.* Genetics and diet regulate vitamin A production via the homeobox transcription factor ISX. *Journal of Biological Chemistry* **288**, 9017-9027 (2013).

42 Lobo, G. P. *et al.* ISX is a retinoic acid-sensitive gatekeeper that controls intestinal β, β-carotene absorption and vitamin A production. *The FASEB Journal* **24**, 1656-1666 (2010).

43 Choi, M. Y. *et al.* A dynamic expression survey identifies transcription factors relevant in mouse digestive tract development. *Development (Cambridge, England)* **133**, 4119-4129 (2006).

44 Reboul, E. & Borel, P. Proteins involved in uptake, intracellular transport and basolateral secretion of fat-soluble vitamins and carotenoids by mammalian enterocytes. *Progress in lipid research* **50**, 388-402 (2011).

45 Merad, M., Ginhoux, F. & Collin, M. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. *Nature Reviews Immunology* **8**, 935 (2008).

46 Chang, S. Y. *et al.* Lack of retinoic acid leads to increased langerin-expressing dendritic cells in gut-associated lymphoid tissues. *Gastroenterology* **138**, 1468-1478. e1466 (2010).

47 Chang, S. Y. & Kweon, M. N. Langerin‐expressing dendritic cells in gut‐associated lymphoid tissues. *Immunological reviews* **234**, 233-246 (2010).

48 Hijova, E. Matrix metalloproteinases: their biological functions and clinical implications. *Bratislavske lekarske listy* **106**, 127-132 (2005).

49 Lackey, D. E. & Hoag, K. A. Vitamin A upregulates matrix metalloproteinase-9 activity by murine myeloid dendritic cells through a nonclassical transcriptional mechanism. *The Journal of nutrition* **140**, 1502-1508, doi:10.3945/jn.110.122556 (2010).

50 MacLauchlan, S. *et al.* Macrophage fusion, giant cell formation, and the foreign body response require matrix metalloproteinase 9. *Journal of leukocyte biology* **85**, 617-626, doi:10.1189/jlb.1008588 (2009).

51 Lo, H. M. *et al.* The carotenoid lutein enhances matrix metalloproteinase-9 production and phagocytosis through intracellular ROS generation and ERK1/2, p38 MAPK, and RARbeta activation in murine macrophages. *Journal of leukocyte biology* **93**, 723-735, doi:10.1189/jlb.0512238 (2013).

52 Munakata, S. *et al.* Inhibition of Plasmin Protects Against Colitis in Mice by Suppressing Matrix Metalloproteinase 9–Mediated Cytokine Release From Myeloid Cells. *Gastroenterology* **148**, 565-578. e564 (2015).

53 Liu, H. *et al.* Constitutive expression of MMP9 in intestinal epithelium worsens murine acute colitis and is associated with increased levels of pro-inflammatory cytokine Kc. *American Journal of Physiology-Heart and Circulatory Physiology* (2013).

54 Nighot, P. K. *et al.* Matrix metalloproteinase 9-induced increase in intestinal epithelial tight junction permeability contributes to the severity of experimental DSS colitis. *American Journal of Physiology-Heart and Circulatory Physiology* (2015).

55 Rodrigues, D. M. *et al.* Matrix metalloproteinase 9 contributes to gut microbe homeostasis in a model of infectious colitis. *BMC microbiology* **12**, 105 (2012).

56 Jeppsson, S. *et al.* Mo2048 Dual Roles Played by Matrix Metalloproteinase (MMP9) in Citrobacter rodentium-induced Hemorrhagic Colitis. *Gastroenterology* **142**, S-728 (2012).

57 Garg, P. *et al.* Matrix metalloproteinase-9-mediated tissue injury overrides the protective effect of matrix metalloproteinase-2 during colitis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **296**, G175-G184 (2009).

58 Santana, A. *et al.* Attenuation of dextran sodium sulphate induced colitis in matrix metalloproteinase-9 deficient mice. *World journal of gastroenterology: WJG* **12**, 6464 (2006).

59 Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. Phylogenetic perspectives in innate immunity. *Science (New York, N.Y.)* **284**, 1313-1318 (1999).

60 Neth, O. *et al.* Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infection and immunity* **68**, 688-693 (2000).

61 Zuo, D.-M., Zhang, L.-Y., Lu, X., Liu, Y. & Chen, Z.-L. Protective role of mouse MBL-C on intestinal mucosa during Shigella flexneri invasion. *International immunology* **21**, 1125-1134, doi:10.1093/intimm/dxp078 (2009).

62 Hibberd, M. L. *et al.* Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. *The Lancet* **353**, 1049-1053 (1999).

63 Kars, M., Van Dijk, H., Salimans, M., Bartelink, A. & Van de Wiel, A. Association of furunculosis and familial deficiency of mannose‐binding lectin. *European journal of clinical investigation* **35**, 531-534 (2005).

64 Roy, S. *et al.* MBL genotype and risk of invasive pneumococcal disease: a case-control study. *The Lancet* **359**, 1569-1573 (2002).

65 Shi, L. *et al.* Mannose-binding lectin-deficient mice are susceptible to infection with Staphylococcus aureus. *Journal of Experimental Medicine* **199**, 1379-1390 (2004).

66 Turner, M. W. Mannose-binding lectin (MBL) in health and disease. *Immunobiology* **199**, 327-339 (1998).

67 Coutsoudis, A. *et al.* Synergy between mannose-binding lectin gene polymorphisms and supplementation with vitamin A influences susceptibility to HIV infection in infants born to HIV-positive mothers. *The American Journal of Clinical Nutrition* **84**, 610-615, doi:10.1093/ajcn/84.3.610 (2006).

68 Lu, T. T. *et al.* Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Molecular cell* **6**, 507-515 (2000).

69 Mamoon, A., Ventura-Holman, T., Maher, J. F. & Subauste, J. S. Retinoic acid responsive genes in the murine hepatocyte cell line AML 12. *Gene* **408**, 95-103 (2008).

70 Maguire, M. *et al.* Diet-dependent retinoid effects on liver gene expression include stellate and inflammation markers and parallel effects of the nuclear repressor Shp. *The Journal of nutritional biochemistry* **47**, 63-74 (2017).

71 Mamoon, A., Subauste, A., Subauste, M. C. & Subauste, J. Retinoic acid regulates several genes in bile acid and lipid metabolism via upregulation of small heterodimer partner in hepatocytes. *Gene* **550**, 165-170 (2014).

72 Andersson, M. K., Karlson, U. & Hellman, L. The extended cleavage specificity of the rodent β-chymases rMCP-1 and mMCP-4 reveal major functional similarities to the human mast cell chymase. *Molecular immunology* **45**, 766-775 (2008).

73 Tchougounova, E., Pejler, G. & Åbrink, M. The chymase, mouse mast cell protease 4, constitutes the major chymotrypsin-like activity in peritoneum and ear tissue. A role for mouse mast cell protease 4 in thrombin regulation and fibronectin turnover. *Journal of Experimental Medicine* **198**, 423-431 (2003).

74 Tchougounova, E. *et al.* A key role for mast cell chymase in the activation of pro-matrix metalloprotease-9 and pro-matrix metalloprotease-2. *Journal of biological chemistry* **280**, 9291-9296 (2005).

75 Friend, D. S. *et al.* Mast cells that reside at different locations in the jejunum of mice infected with Trichinella spiralis exhibit sequential changes in their granule ultrastructure and chymase phenotype. *The Journal of cell biology* **135**, 279-290 (1996).

76 Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380 (2011).

77 Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature neuroscience* **11**, 1271 (2008).

78 Alexandrakis, M. G. *et al.* Inhibitory Effect of Retinoic Acid on Proliferation, Maturation and Tryptase Level in Human Leukemic Mast Cells (HMC-1). *International Journal of Immunopathology and Pharmacology* **16**, 43-47, doi:10.1177/039463200301600106 (2003).

79 Kinoshita, T. *et al.* Retinoic acid is a negative regulator for the differentiation of cord blood-derived human mast cell progenitors. *Blood* **95**, 2821-2828 (2000).

80 Ishida, S., Kinoshita, T., Sugawara, N., Yamashita, T. & Koike, K. Serum inhibitors for human mast cell growth: possible role of retinol. *Allergy* **58**, 1044-1052 (2003).

81 Wiedermann, U. *et al.* Vitamin A deficiency increases inflammatory responses. *Scandinavian Journal of Immunology* **44**, 578-584 (1996).

82 Banerjee, A., McKinley, E. T., von Moltke, J., Coffey, R. J. & Lau, K. S. Interpreting heterogeneity in intestinal tuft cell structure and function. *The Journal of clinical investigation* **128**, 1711-1719 (2018).

83 Gerbe, F. *et al.* Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium. *The Journal of cell biology* **192**, 767-780 (2011).

84 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339, doi:10.1038/nature24489 (2017).

85 Wierstra, I. & Alves, J. FOXM1, a typical proliferation-associated transcription factor. *Biological chemistry* **388**, 1257-1274 (2007).

86 Frick, I.-M. *et al.* Constitutive and inflammation-dependent antimicrobial peptides produced by epithelium are differentially processed and inactivated by the commensal Finegoldia magna and the pathogen Streptococcus pyogenes. *The Journal of Immunology* **187**, 4300-4309 (2011).

87 Maerki, C. *et al.* Potent and broad-spectrum antimicrobial activity of CXCL14 suggests an immediate role in skin infections. *The Journal of Immunology* **182**, 507-514 (2009).

88 Sidahmed, A. M. *et al.* CXCL14 deficiency does not impact the outcome of influenza or Escherichia coli infections in mice. *The Journal of Infection in Developing Countries* **8**, 1301-1306 (2014).

89 Hara, T. & Tanegashima, K. Pleiotropic functions of the CXC-type chemokine CXCL14 in mammals. *The Journal of Biochemistry* **151**, 469-476, doi:10.1093/jb/mvs030 (2012).

90 Cao, X. *et al.* Molecular cloning and characterization of a novel CXC chemokine macrophage inflammatory protein-2γ chemoattractant for human neutrophils and dendritic cells. *The Journal of Immunology* **165**, 2588-2595 (2000).
91 Snyder, L. M. *et al.* Retinoic Acid Mediated Clearance of Citrobacter rodentium in Vitamin A Deficient Mice Requires CD11b+ and T Cells. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.03090 (2019).
92 Hodges, J. K., Tan, L., Green, M. H. & Ross, A. C. Vitamin A supplementation transiently increases retinol concentrations in extrahepatic organs of neonatal rats raised under vitamin A–marginal conditions. *The Journal of nutrition* **146**, 1953-1960 (2016).
93 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
94 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357 (2015).
95 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
96 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24 (2011).
97 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2013).
98 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).
99 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559 (2008).
100 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4** (2005).
101 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720 (2007).
102 Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC systems biology* **1**, 54 (2007).
103 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* **16**, 284-287 (2012).
104 Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proceedings of the National Academy of Sciences* **107**, 12698-12703 (2010).
105 Shi, A. *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research* **47**, D721-D728, doi:10.1093/nar/gky900 (2018).
106 Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Research* **47**, D33-D38, doi:10.1093/nar/gky822 (2018).

Figure legends

Fig. 1 VAD mice have lower serum retinol levels than VAS mice. Serum retinol concentrations were measured by UPLC on post-infection day 5. Values are the mean ± SEM of n=3-6 mice per group. Two-way ANOVA test. *** P<0.001. **Abbreviation:**

1005 vitamin A (VA), vitamin A deficient (VAD), vitamin A sufficient (VAS), Ultra
1006 Performance Liquid Chromatography (UPLC), standard error of the mean (SEM),
1007 analysis of variance (ANOVA)
1008

1009 **Fig. 2 Overview of the results of bioinformatics analyses.** Initial mapping identified
1010 24421 genes. After data pre-processing, 14368 genes in the SI study and 15340 genes in
1011 the colon study were maintained in the two datasets. Starting from the 14368 SI-
1012 expressed genes, 49 DEGs corresponded to VA effect, and 9 co-expression modules were
1013 identified. Similarly for the colon study, 94 DEGs and 13 modules were identified among
1014 the 15340 colon-expressed genes in the colon study. A total of 14059 genes were shared
1015 between the two datasets and were used to for the module preservation analysis.
1016

1017 **Fig. 3 Transcriptional changes in the intestines corresponding to VA status. a**, **b**,
1018 Volcano plots of DEGs in SI (**a**) and in colon (**b**). Red: DEGs upregulated in VAS group
1019 comparing with VAD group; Blue: DEGs downregulated in VAS group comparing with
1020 VAD group; Black: non-DEGs. Criteria: |Fold Change|=2 is marked out by two black
1021 vertical lines x=1 and x=-1. adjp<0.05 is marked out by a horizontal black line y= -
1022 Log10(0.05). **c**, **d**, GO enrichment terms upregulated (**c**) or downregulated (**d**) by VA in
1023 SI. Criteria: p.adj<0.05. **Abbreviations:** padj: adjusted p value according to DESeq2,
1024 differentially expressed gene (DEG), SI (small intestine), VAS (vitamin A sufficient),
1025 VAD (vitamin A deficient), Gene ontology (GO), p.adj (BH-adjusted p value from the
1026 hypergeometric test conducted by ClusterProfiler).
1027

1028 **Fig. 4 Co-expression network construction in the intestines via WGCNA. a**, **b**, **c**,
1029 Results from SI study. **d**, **e**, **f**, Results from colon study. **a**, **d**, Soft threshold power beta
1030 were chosen for SI (**a**) and colon (**d**) networks according to the scale-free topology
1031 criteria (horizontal red lines). **b**, **e**, Clustering dendrograms of co-expression modules in
1032 SI (**b**) and colon (**e**). Transcripts are represented as vertical lines and are arranged in
1033 branches according to topological overlap similarity using the hierarchical clustering. The
1034 dynamic treecut algorithm was used to automatically detect the nine highly connected
1035 modules, referred to by a color designation. Module color designation are shown below
1036 the dendrogram. **c**, **f**, Correlation of module eigengenes with VA status in SI (**c**) and colon
1037 (**f**) networks. Upper numbers in each cell: Correlation coefficient between the module and
1038 the trait. Lower numbers in parentheses: p-value of the correlation. Significant
1039 correlations were obtained between VA status and the SI(Yellow), SI(Pink), SI(Red) (**c**),
1040 Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) (**f**) modules.
1041 **Abbreviation:** weighted gene co-expression network analysis (WGCNA), small intestine
1042 (SI), vitamin A (VA).
1043

1044 **Fig. 5 Visualization of the co-expression network in three modules of interest in SI**.
1045 **a**, SI(Yellow), **b**, SI(Red), and **c**, SI(Pink). In each of the modules of interest, the top 50

1046 genes ranked by connectivity are depicted using igraph in R. Each node represents a
1047 gene. The size of the node is proportional to the connectivity of the gene. The shape of
1048 the nodes is coded squares=TFs, circles=non-TFs. Color of the nodes denote: Blue=
1049 upregulated DEGs in VAS condition, Red=downregulated DEGs. Font: red bold
1050 italic=cell type markers. Thickness of lines between nodes: proportional to the adjacency
1051 value of the gene pair. In order to improve network visibility, adjacencies equal to 1 or
1052 lower than 0.7 were omitted. **Abbreviations**: SI (small intestine), TF (transcription
1053 factor), VAS (vitamin A sufficient).
1054

1055 **Fig. 6 Overlap between module genes and SI cell markers**. The cell type marker lists
1056 of SI were obtained from CellMarker, a manually curated database. Upper numbers in
1057 each cell: Number of probes shared between the designated module (y axis) and the cell
1058 marker list (x axis). Lower numbers in parentheses: Bonferroni corrected p-value (CP) of
1059 the one-sided hypergeometric test, which evaluates if the gene list of a module
1060 significantly over-represents any cell type markers. Criteria: CP<0.05. Color shade of
1061 each cell is according to –log10(CP). Four significant overlaps were highlighted with red
1062 shades.
1063

1064 **Tables**

1065 The only tables are enclosed as supplementary tables.

1066

1067 **Data availability**

1068 The RNAseq datasets generated for this study can be found in NCBI Gene Expression

1069 Omnibus (GEO) database with the accession number GSE143290 at

1070 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143290.

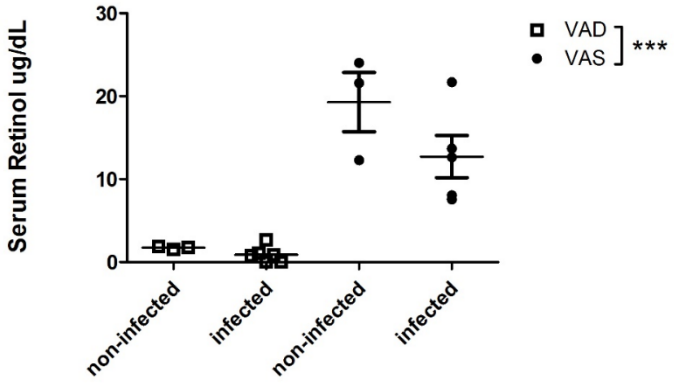


Fig. 1 VAD mice have lower serum retinol levels than VAS mice. Serum retinol concentrations were measured by UPLC on post-infection day 5. Values are the mean ± SEM of n=3-6 mice per group. Two-way ANOVA test. *** $P<0.001$. **Abbreviation:** vitamin A (VA), vitamin A deficient (VAD), vitamin A sufficient (VAS), Ultra Performance Liquid Chromatography (UPLC), standard error of the mean (SEM), analysis of variance (ANOVA)

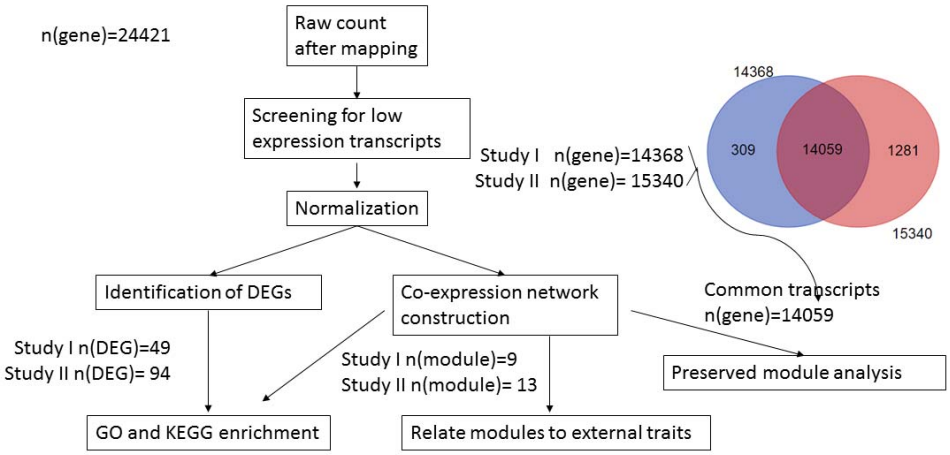


Fig. 2 Overview of the results of bioinformatics analyses. Initial mapping identified 24421 genes. After data pre-processing, 14368 genes in the SI study and 15340 genes in the colon study were maintained in the two datasets. Starting from the 14368 SI-expressed genes, 49 DEGs corresponded to VA effect, and 9 co-expression modules were identified. Similarly for the colon study, 94 DEGs and 13 modules were identified among the 15340 colon-expressed genes in the colon study. A total of 14059 genes were shared between the two datasets and were used to for the module preservation analysis.

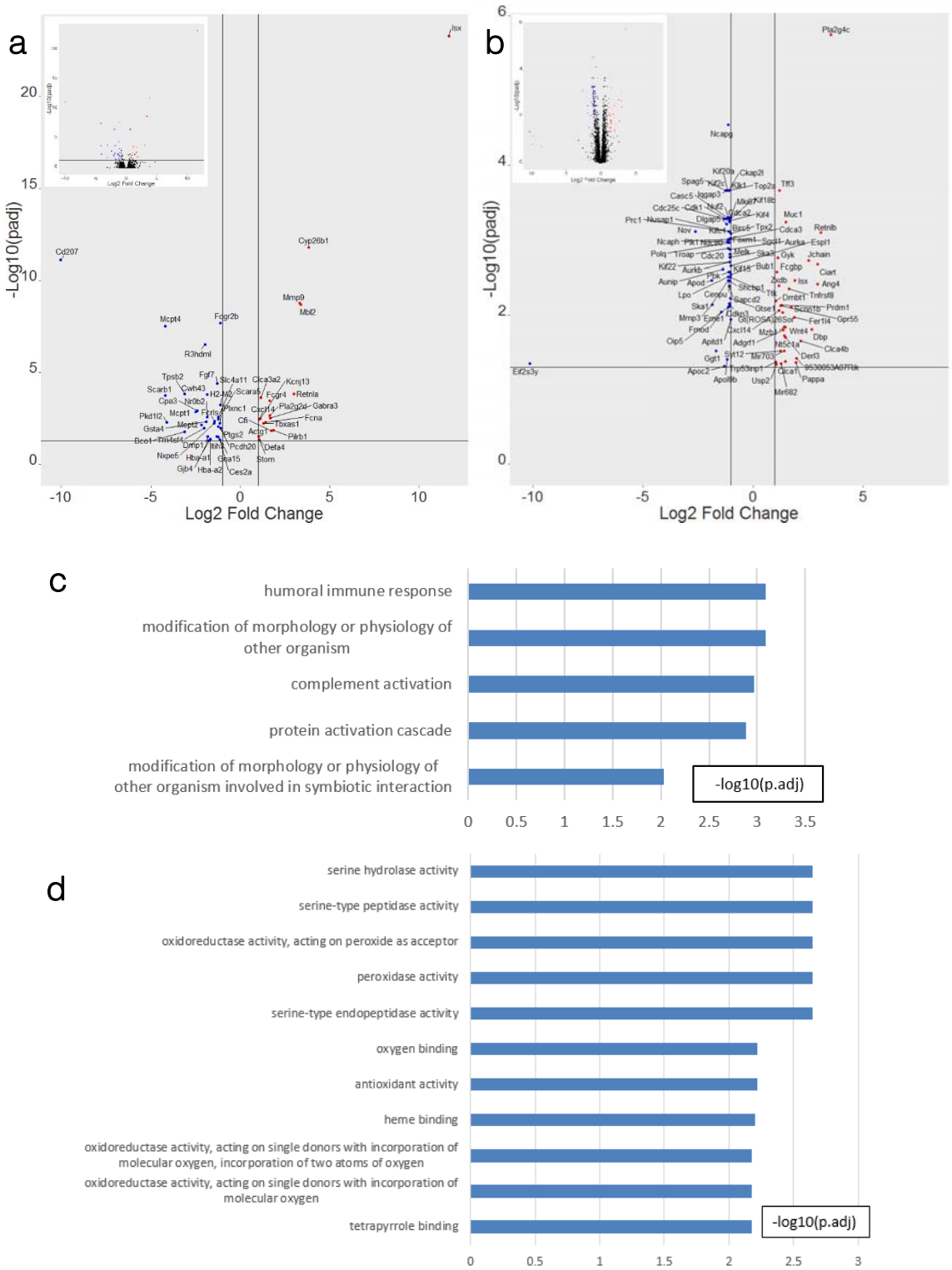


Fig. 3 Transcriptional changes in the intestines corresponding to VA status. a, **b**, Volcano plots of DEGs in SI (**a**) and in colon (**b**). Red: DEGs upregulated in VAS group comparing with VAD group; Blue: DEGs downregulated in VAS group comparing with VAD group; Black: non-DEGs. Criteria: |Fold Change|=2 is marked out by two black vertical lines x=1 and x=-1. adjp<0.05 is marked out by a horizontal black line y= -Log10(0.05). **c**, **d**, GO enrichment terms upregulated (**c**) or downregulated (**d**) by VA in SI. Criteria: p.adj<0.05. **Abbreviations:** padj: adjusted p value according to DESeq2, differentially expressed gene (DEG), SI (small intestine), VAS (vitamin A sufficient), VAD (vitamin A deficient), Gene ontology (GO), p.adj (BH-adjusted p value from the hypergeometric test conducted by ClusterProfiler).

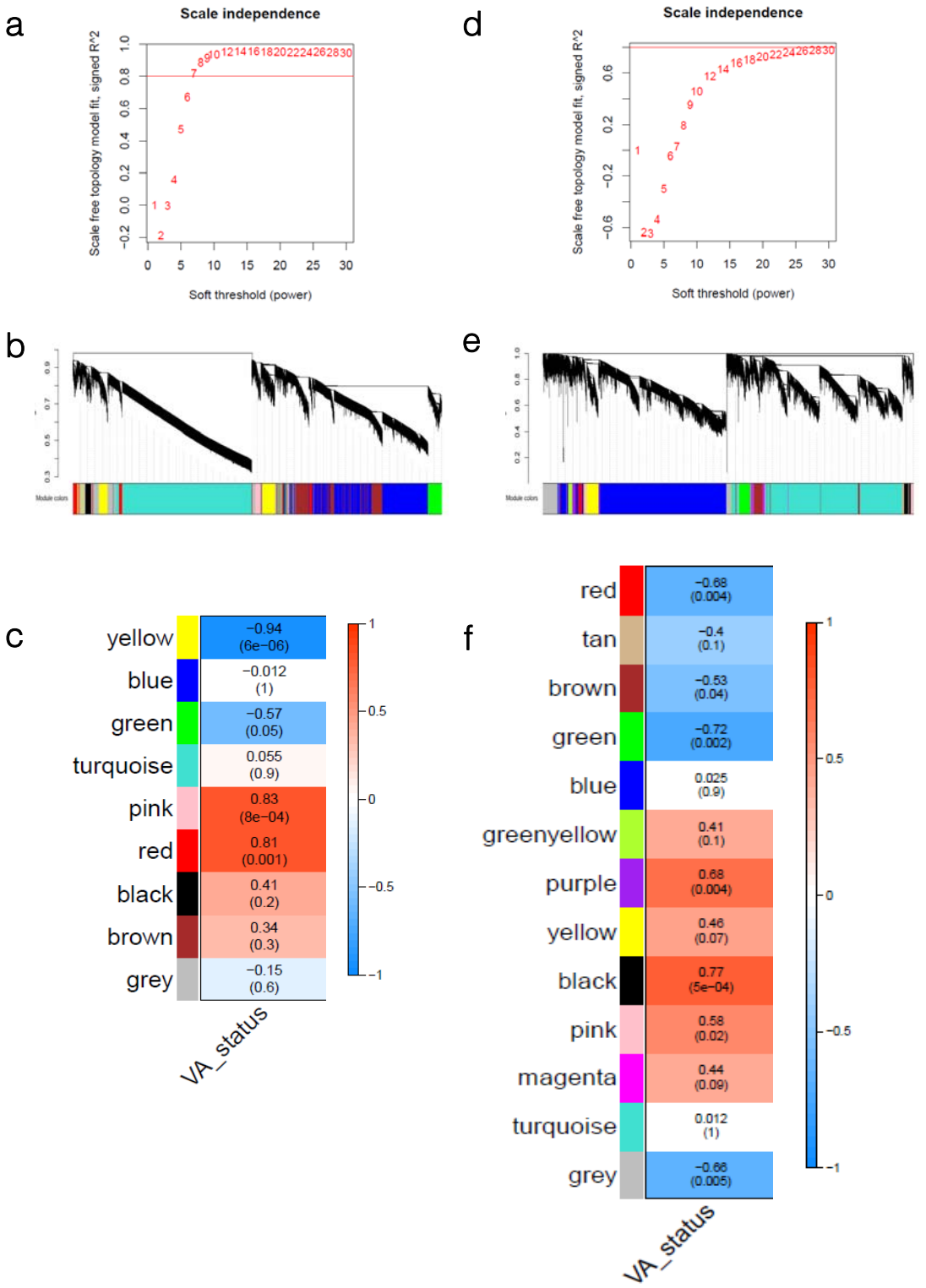


Fig. 4 Co-expression network construction in the intestines via WGCNA. a, **b**, **c**, Results from SI study. **d**, **e**, **f**, Results from colon study. **a**, **d**, Soft threshold power beta were chosen for SI (**a**) and colon (**d**) networks according to the scale-free topology criteria (horizontal red lines). **b**, **e**, Clustering dendrograms of co-expression modules in SI (**b**) and colon (**e**). Transcripts are represented as vertical lines and are arranged in branches according to topological overlap similarity using the hierarchical clustering. The dynamic treecut algorithm was used to automatically detect the nine highly connected modules, referred to by a color designation. Module color designation are shown below the dendrogram. **c**, **f**, Correlation of module eigengenes with VA status in SI (**c**) and colon (**f**) networks. Upper numbers in each cell: Correlation coefficient between the module and the trait. Lower numbers in parentheses: p-value of the correlation. Significant correlations were obtained between VA status and the SI(Yellow), SI(Pink), SI(Red) (**c**), Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) (**f**) modules. **Abbreviation:** weighted gene co-expression network analysis (WGCNA), small intestine (SI), vitamin A (VA).

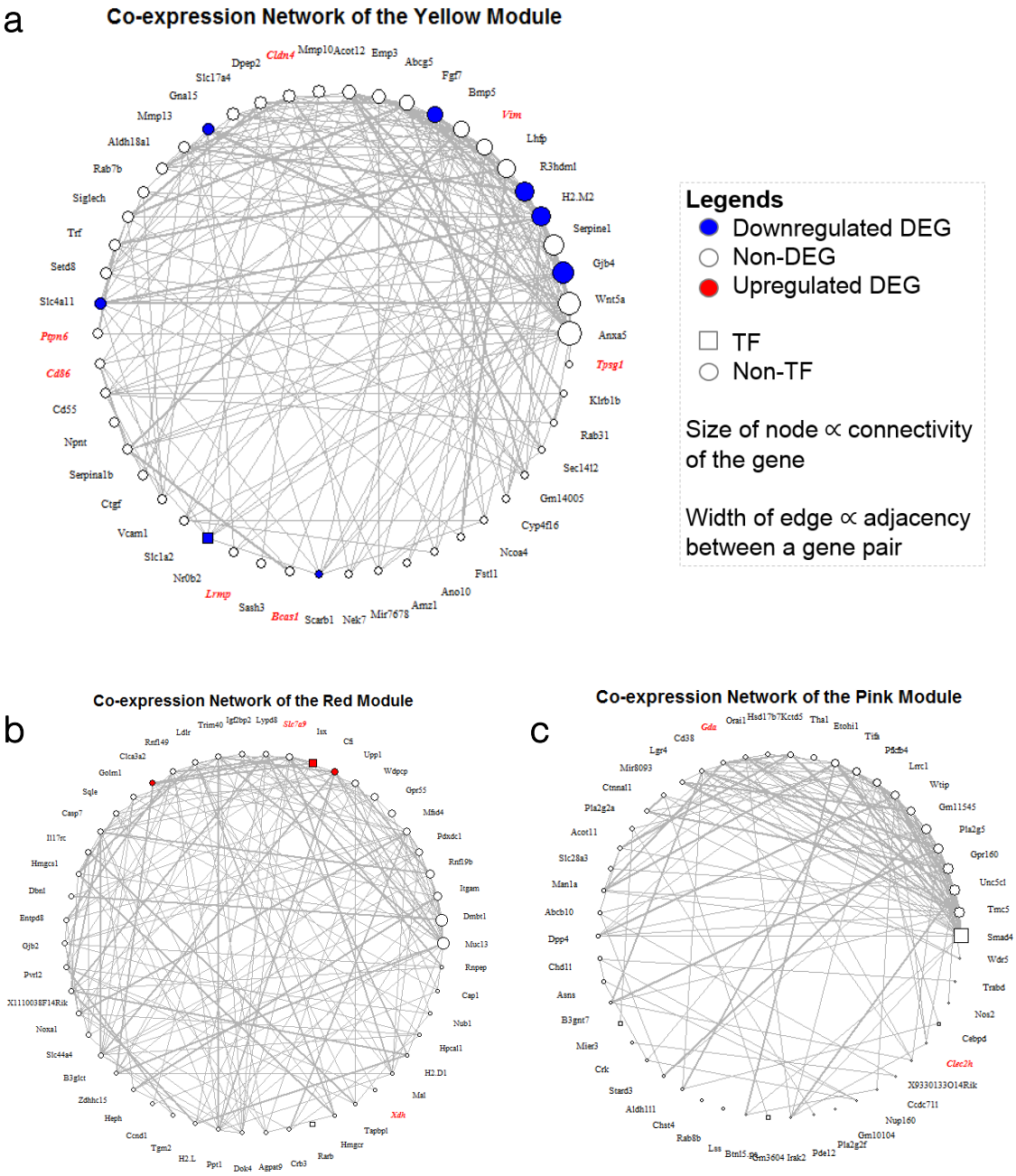


Fig. 5 Visualization of the co-expression network in three modules of interest in SI. **a**, SI(Yellow), **b**, SI(Red), and **c**, SI(Pink). In each of the modules of interest, the top 50 genes ranked by connectivity are depicted using igraph in R. Each node represents a gene. The size of the node is proportional to the connectivity of the gene. The shape of the nodes is coded squares=TFs, circles=non-TFs. Color of the nodes denote: Blue= upregulated DEGs in VAS condition, Red=downregulated DEGs. Font: red bold italic=cell type markers. Thickness of lines between nodes: proportional to the adjacency value of the gene pair. In order to improve network visibility, adjacencies equal to 1 or lower than 0.7 were omitted. **Abbreviations**: SI (small intestine), TF (transcription factor), VAS (vitamin A sufficient).

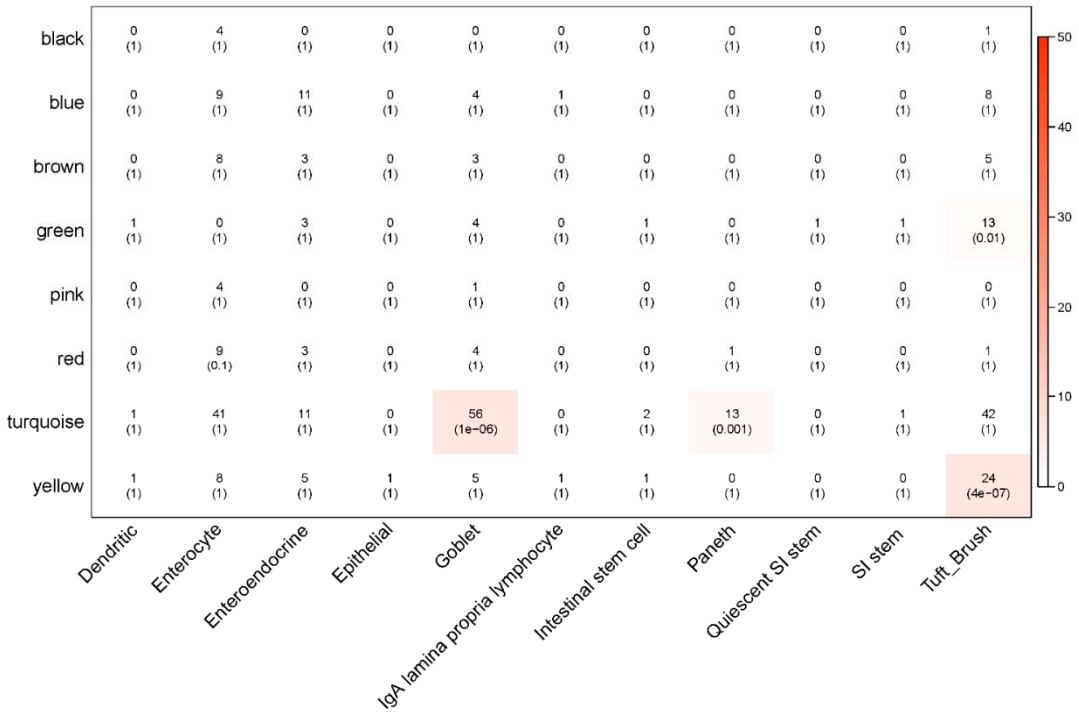
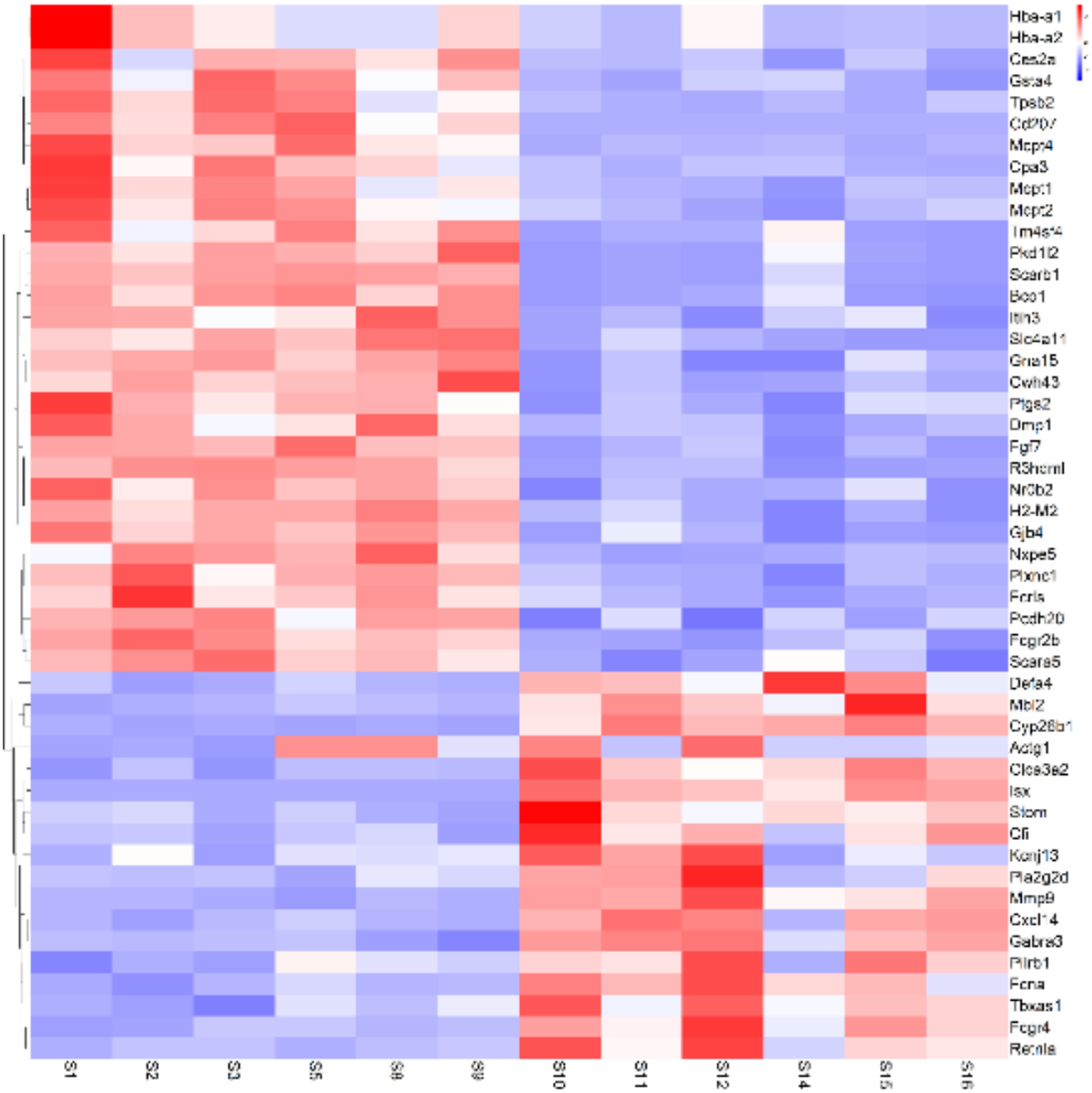


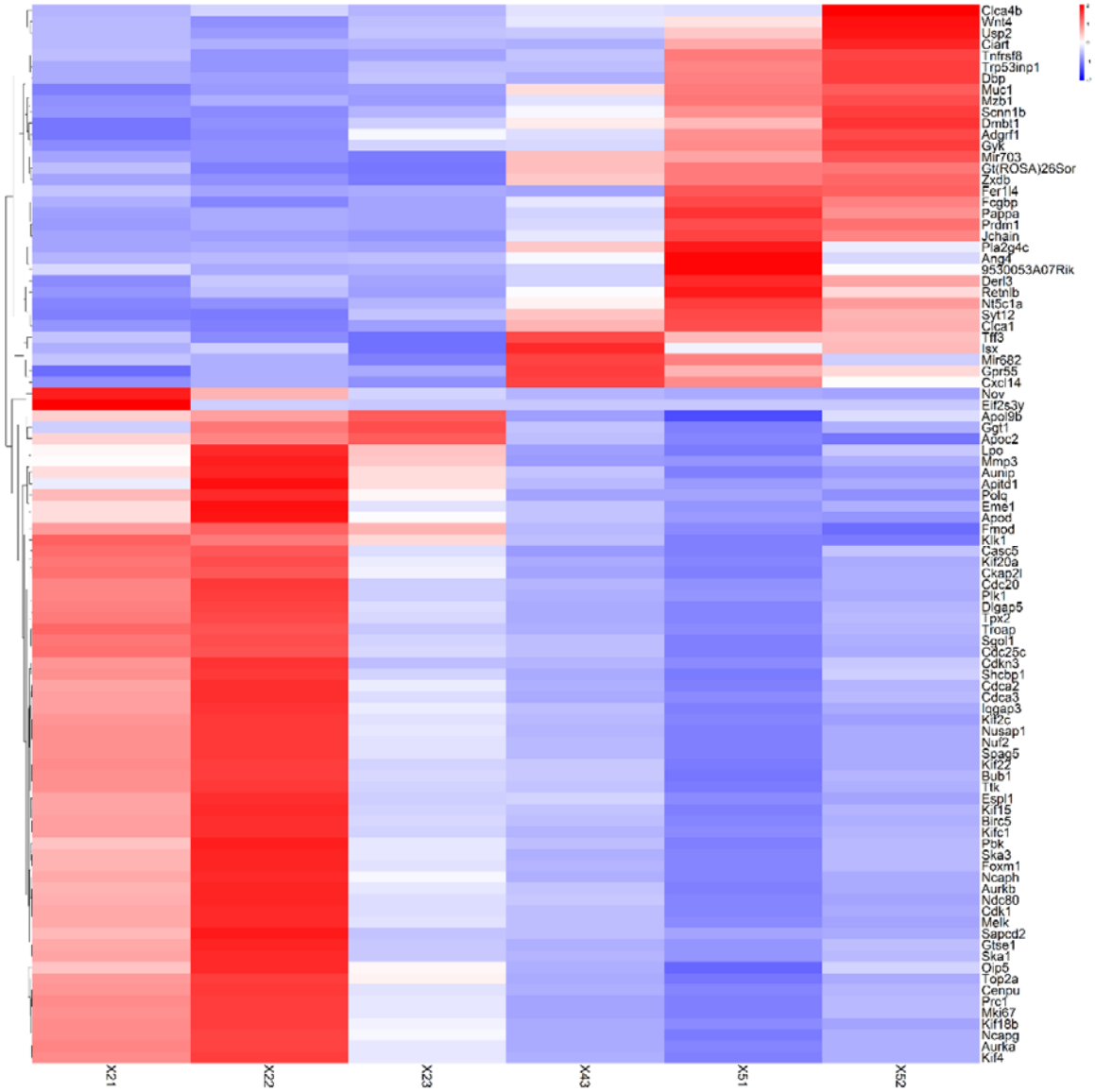
Fig. 6 Overlap between module genes and SI cell markers. The cell type marker lists of SI were obtained from CellMarker, a manually curated database. Upper numbers in each cell: Number of probes shared between the designated module (y axis) and the cell marker list (x axis). Lower numbers in parentheses: Bonferroni corrected p-value (CP) of the one-sided hypergeometric test, which evaluates if the gene list of a module significantly over-represents any cell type markers. Criteria: CP<0.05. Color shade of each cell is according to –log10(CP). Four significant overlaps were highlighted with red shades.

Supplementary material



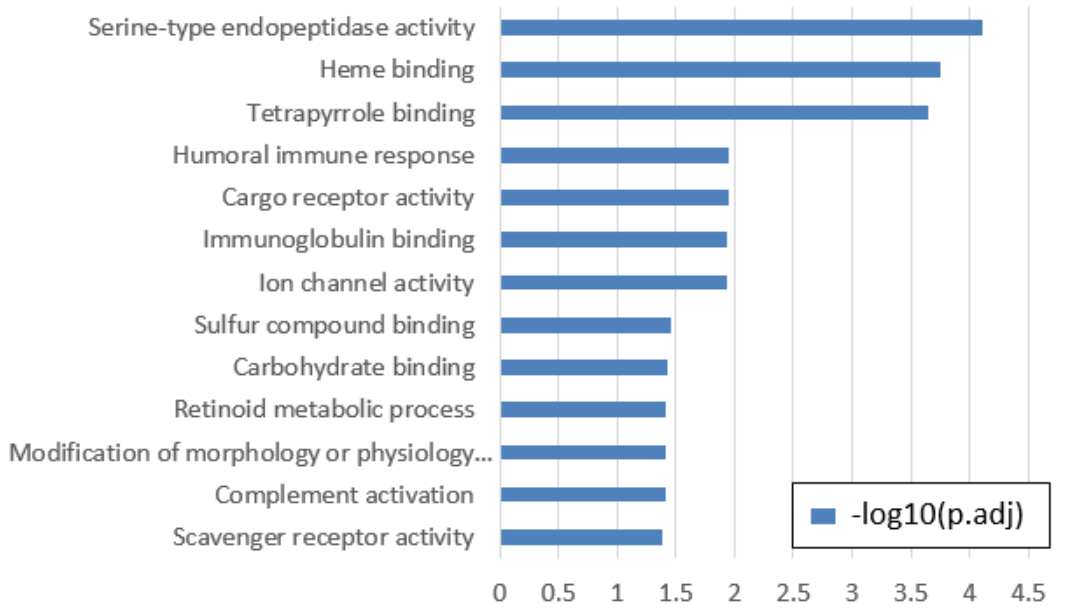
Supplementary Figure 1. Heat map for DEGs corresponding to VA effect in the SI study.

Notes: Red: group mean expression significantly higher; Blue: group mean expression significantly lower. VAD: columns S1, S2, S3, S5, S8, and S9. VAS: columns S10, S11, S12, S14, S15, and S16. **Abbreviations:** differentially expressed gene **(**DEG), vitamin A deficient (VAD), vitamin A sufficient (VAS).



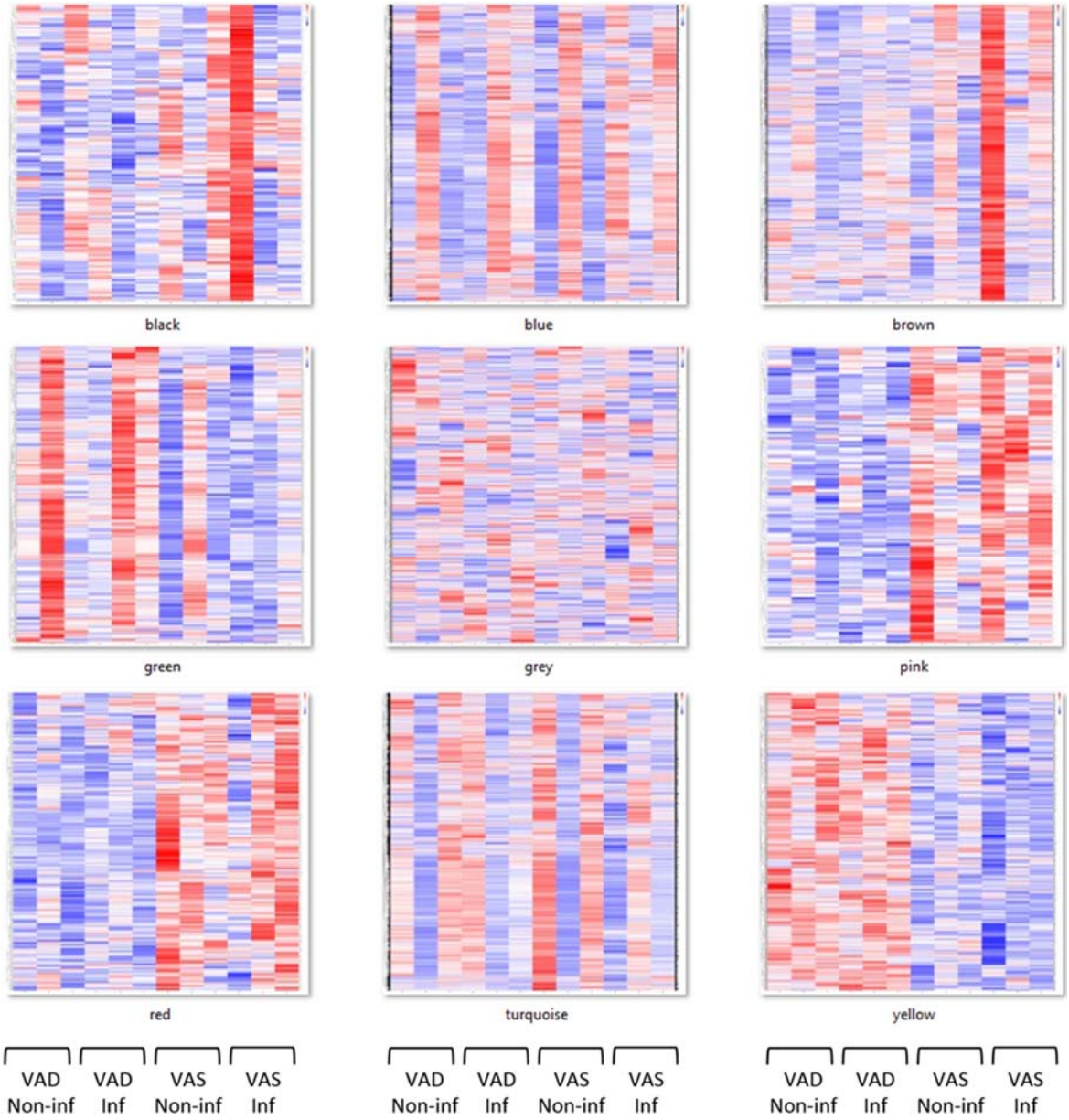
Supplementary Figure 2. Heat map for DEGs corresponding to VA effect in the colon study.

Notes: Red: group mean expression significantly higher; Blue: group mean expression significantly lower. VAD: columns X21, X22 and X23. VAS: columns X43, X51 and X52.
Abbreviations: differentially expressed gene **(**DEG), vitamin A deficient (VAD), vitamin A sufficient (VAS).



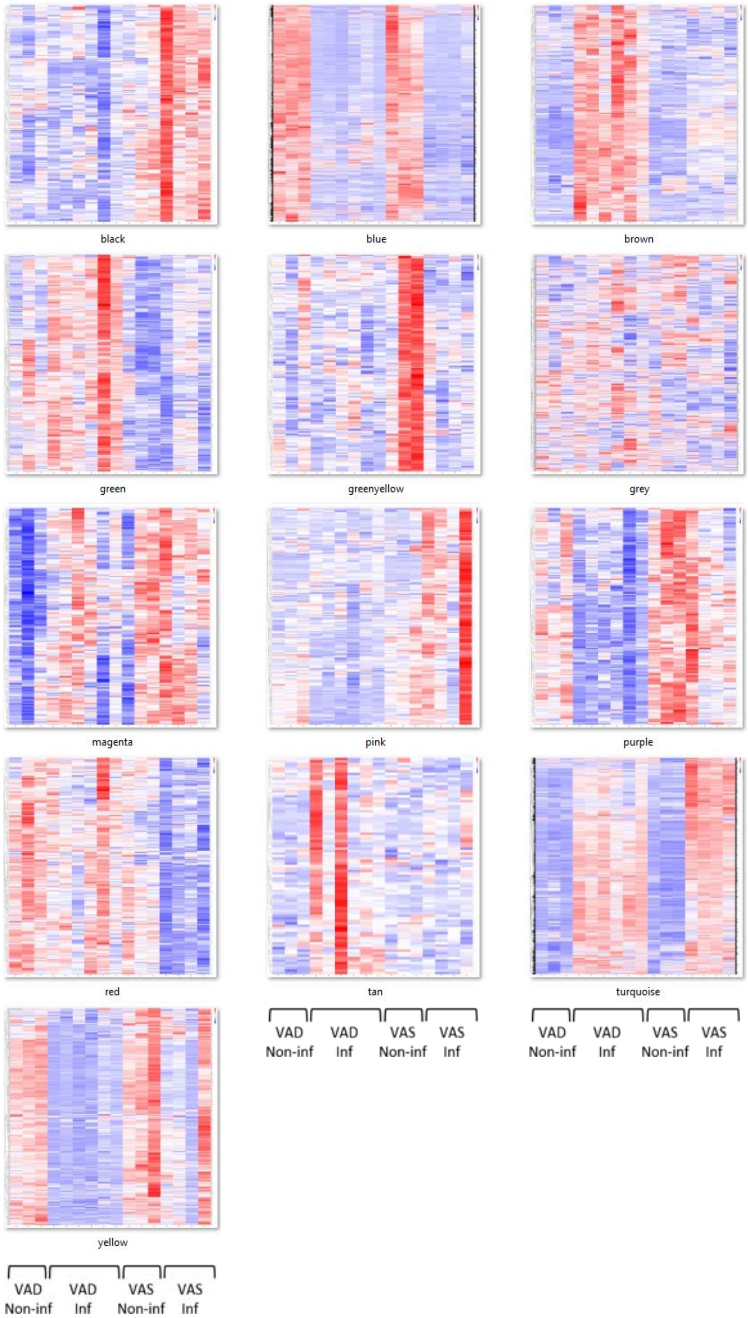
Supplementary Figure 3. GO enrichment terms of DEGs corresponding to VA effect in the SI study.

Criteria: p.adj<0.05.
Abbreviations: p.adj (BH-adjusted p value from the hypergeometric test conducted by ClusterProfiler), Gene ontology (GO), differentially expressed gene (DEG)



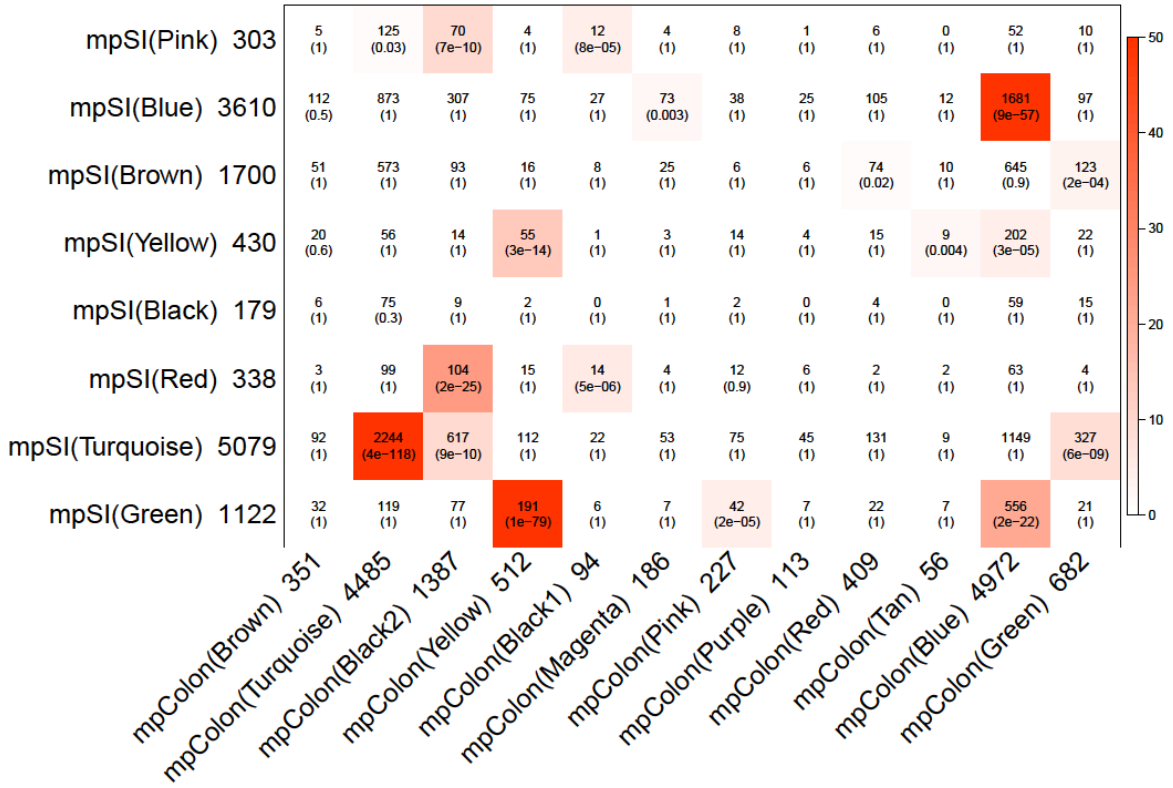
Supplementary Figure 4. Gene co-expression patterns identified by WGCNA in the SI study.

Notes: Red: High expression level; Blue: Low expression level.
VAD: columns 1, 2, 3, 4, 5, and 6. VAS: columns 7, 8, 9, 10, 11, and 12.
Abbreviations: vitamin A deficient (VAD), vitamin A sufficient (VAS), weighted gene co-expression network analysis (WGCNA).



Supplementary Figure 5. Gene co-expression patterns identified by WGCNA in the colon study.

Notes: Red: High expression level; Blue: Low expression level.
VAD: columns 1, 2, 3, 4, 5, 6, 7, 8, and 9. VAS: columns 10, 11, 12, 13, 14, 15, and 16.
Abbreviations: vitamin A deficient (VAD), vitamin A sufficient (VAS), weighted gene co-expression network analysis (WGCNA).



Supplementary Figure 6. Cross-tabulation of SI modules (rows) and colon modules (columns).

Notes: Each row and column is labeled by (color assignment__module size). In each cell of the table, upper numbers give counts of genes in the intersection of the corresponding row and column modules. Lower numbers in parentheses: Bonferroni corrected p-value (CP) of the one-sided hypergeometric test, which evaluates if a significant overlap is reached between the corresponding modules. Criteria: CP<0.05. The table is color-coded by –log10(CP), according to the color legend on the right.

Supplementary Table 1. Upregulated DEGs in the VAS group compared with VAD group in the SI study. The order is ranked by the adjusted p values (padj).

Abbreviations:

BaseMean: Mean expression level (normalized by DESeq2) across all 12 libraries

Fold Change: Normalized mean expression level of VAS group / Normalized mean expression level of VAD group

padj: adjusted p value calculated by DESeq2

	BaseMean	Fold Change	padj	EntrezID	Fullname
Isx	957.15	3231.03	4.92E-24	71597	intestine specific homeobox
Cyp26b1	39.63	14.12	1.52E-12	232174	cytochrome P450, family 26, subfamily b, polypeptide 1
Mmp9	53.98	10.06	1.60E-09	17395	matrix metallopeptidase 9
Mbl2	164.18	10.37	2.02E-09	17195	mannose-binding lectin (protein C) 2
Retnla	32.95	7.93	0.000147	57262	resistin like alpha
Clca3a2	464.34	2.22	0.000235	80797	chloride channel accessory 3A2
Fcgr4	81.24	3.17	0.000337	246256	Fc receptor, IgG, low affinity IV
Pla2g2d	58.32	3.13	0.002081	18782	phospholipase A2, group IID
Gabra3	32.47	3.18	0.002919	14396	gamma-aminobutyric acid (GABA) A receptor, subunit alpha 3
Kcnj13	105.76	2.18	0.003246	100040591	potassium inwardly-rectifying channel, subfamily J, member 13
Cxcl14	400.56	2.08	0.003349	57266	chemokine (C-X-C motif) ligand 14
Fcna	30.21	2.67	0.004892	14133	ficolin A
Tbxas1	50.77	2.46	0.00581	21391	thromboxane A synthase 1, platelet
Pilrb1	36.20	3.63	0.013114	170741	paired immunoglobin-like type 2 receptor beta 1
Cfi	106.97	3.33	0.014409	12630	complement component factor i
Actg1	42795.18	2.04	0.030234	11465	actin, gamma, cytoplasmic 1
Defa4	661.50	2.08	0.042915	13238	defensin, alpha 4

Stom	6202.03	2.09	0.045444	13830	stomatin

Supplementary Table 2. Downregulated DEGs in the VAS group compared with VAD group in the SI study. The order is ranked by the adjusted p values (padj).

Abbreviations:

BaseMean: Mean expression level (normalized by DESeq2) across all 12 libraries

Fold Change: -Normalized mean expression level of VAD group/ Normalized mean expression level of VAS group

padj: adjusted p value calculated by DESeq2

	BaseMean	Fold Change	padj	EntrezID	Fullname
Cd207	90.13	-1055.17	7.41E-12	246278	CD207 antigen
Fcgr2b	197.54	-2.17	1.95E-08	14130	Fc receptor, IgG, low affinity IIb
Mcpt4	42.09	-18.32	3.06E-08	17227	mast cell protease 4
R3hdml	41.80	-3.91	3.01E-07	1E+08	R3H domain containing-like
Fgf7	125.51	-2.44	3.78E-05	14178	fibroblast growth factor 7
Tpsb2	20.66	-8.61	0.000147	17229	tryptase beta 2
Cwh43	47.30	-3.56	0.000159	231293	cell wall biogenesis 43 C-terminal homolog
Scarb1	15409.55	-18.32	0.000173	20778	scavenger receptor class B, member 1
H2-M2	195.61	-2.14	0.000579	14990	histocompatibility 2, M region, locus 2
Cpa3	25.50	-5.29	0.001197	12873	carboxypeptidase A3, mast cell
Mcpt1	201.78	-5.57	0.001313	17224	mast cell protease 1
Slc4a11	73.43	-2.35	0.002602	269356	solute carrier family 4 sodium bicarbonate, transporter-like, member 11
Nr0b2	46.58	-3.60	0.002602	23957	nuclear receptor subfamily 0, group B, member 2
Scara5	64.88	-2.35	0.003349	71145	scavenger receptor class A member 5
Fcrls	75.39	-2.69	0.004543	80891	Fc receptor-like S, scavenger receptor
Gsta4	2001.56	-3.71	0.004722	14860	glutathione S-transferase, alpha 4
Pkd1l2	82.38	-17.15	0.00499	76645	polycystic kidney disease 1 like 2
Plxnc1	92.51	-2.19	0.00581	54712	plexin C1
Nxpe5	38.78	-2.74	0.00581	381680	neurexophilin and PC-esterase domain family, member 5
Mcpt2	79.94	-4.45	0.006874	17225	mast cell protease 2
Ptgs2	184.93	-2.35	0.00872	19225	prostaglandin-endoperoxide synthase 2
Pcdh20	75.28	-2.13	0.010642	219257	protocadherin 20

Tm4sf4	5881.59	-4.07	0.010642	229302	transmembrane 4 superfamily member 4
Bco1	18783.84	-8.66	0.017006	63857	beta-carotene oxygenase 1
Itih3	56.45	-2.50	0.029795	16426	inter-alpha trypsin inhibitor, heavy chain 3
Dmp1	174.24	-3.49	0.030234	13406	dentin matrix protein 1
Gna15	32.73	-2.33	0.032603	14676	guanine nucleotide binding protein, alpha 15
Hba-a1	1116.80	-3.15	0.042345	15122	hemoglobin alpha, adult chain 1
Hba-a2	1116.81	-3.15	0.042345	110257	hemoglobin alpha, adult chain 2
Ces2a	7062.11	-2.19	0.044776	102022	carboxylesterase 2A
Gjb4	18.73	-3.44	0.049916	14621	gap junction protein, beta 4

Supplementary Table 3. Upregulated DEGs in the VAS group compared with VAD group in the colon study. The order is ranked by the adjusted p values (padj).

Abbreviations:

BaseMean: Mean expression level (normalized by DESeq2) across all 16 libraries

Fold Change: Normalized mean expression level of VAS group/ Normalized mean expression level of VAD group

padj: adjusted p value calculated by DESeq2

	BaseMean	Fold change	padj	EntrezID	Fullname
Pla2g4c	153.87	11.61	1.77E-06	232889	phospholipase A2 group IVC (cytosolic calcium-independent)
Tff3	14507.68	2.31	0.000218	21786	trefoil factor 3 intestinal
Muc1	1208.43	2.84	0.000576	17829	mucin 1
Retnlb	4272.92	8.44	0.000801	57263	resistin like beta
Gyk	5346.19	2.19	0.001731	NA	NA
Jchain	8052.94	5.80	0.001902	16069	immunoglobulin joining chain
Ciart	70.23	7.68	0.002115	229599	circadian associated repressor of transcription
Fcgbp	58520.79	2.14	0.002649	215384	Fc fragment of IgG binding protein
Isx	478.87	3.75	0.003429	71597	intestine specific homeobox
Ang4	737.71	7.73	0.003836	219033	Angiogenin ribonuclease A family member 4
Zxdb	128.03	2.29	0.004079	668166	zinc finger X-linked duplicated B
Tnfrsf8	253.70	3.13	0.004473	21941	tumor necrosis factor receptor superfamily member 8
Dmbt1	256019.57	2.05	0.006472	12945	deleted in malignant brain tumors 1
Prdm1	365.97	2.46	0.007372	12142	PR domain containing 1 with ZNF domain
Gpr55	62.18	2.39	0.007529	227326	G protein-coupled receptor 55
Scnn1b	171.80	3.36	0.007837	20277	sodium channel nonvoltage-gated 1 beta
Cxcl14	237.56	2.27	0.008616	57266	chemokine (C-X-C motif) ligand 14
Gt(ROSA)26Sor	640.78	2.57	0.009265	14910	gene trap ROSA 26 Philippe Soriano
Fer1l4	1111.67	3.67	0.010737	74562	fer-1-like 4 (C. elegans)
Mzb1	499.91	2.70	0.014338	69816	marginal zone B and B1 cell-specific protein 1

Wnt4	153.06	2.76	0.014572	22417	wingless-type MMTV integration site family member 4
Dbp	598.43	6.36	0.015664	13170	D site albumin promoter binding protein
Adgrf1	1333.45	2.60	0.015726	77596	adhesion G protein-coupled receptor F1
Nt5c1a	241.89	2.69	0.018873	230718	5'-nucleotidase cytosolic IA
Derl3	132.87	2.79	0.019995	70377	Der1-like domain family member 3
Clca4b	35258.66	4.52	0.022243	99709	chloride channel accessory 4B
Syt12	117.12	2.39	0.030138	171180	synaptotagmin XII
Mir703	311.12	2.70	0.030376	735265	microRNA 703
9530053A07Rik	150.87	3.94	0.037762	319482	RIKEN cDNA 9530053A07 gene
Clca1	33895.13	2.76	0.041848	23844	chloride channel accessory 1
Trp53inp1	2245.74	2.04	0.042511	60599	transformation related protein 53 inducible nuclear protein 1
Pappa	95.62	3.85	0.043324	18491	pregnancy-associated plasma protein A
Mir682	1134.07	2.44	0.044732	751556	microRNA 682
Usp2	574.80	2.10	0.045265	53376	ubiquitin specific peptidase 2

Supplementary Table 4. Downregulated DEGs in the VAS group compared with VAD group in the colon study. The order is ranked by the adjusted p values (padj).

Abbreviations:

baseMean: Mean expression level (normalized by DESeq2) across all 16 libraries

Fold Change: -Normalized mean expression level of VAD group / Normalized mean expression level of VAS group

padj: adjusted p value calculated by DESeq2

	baseMean	Fold Change	padj	EntrezID	Fullname
Ncapg	980.69	-2.17	2.86E-05	54392	non-SMC condensin I complex subunit G
Spag5	1195.54	-2.36	0.000218	54141	sperm associated antigen 5
Iqgap3	1246.37	-2.27	0.000218	404710	IQ motif containing GTPase activating protein 3
Kif2c	723.68	-2.27	0.000218	73804	kinesin family member 2C
Kif20a	1461.39	-2.23	0.000218	19348	kinesin family member 20A
Ckap2l	1089.29	-2.16	0.000218	70466	cytoskeleton associated protein 2-like
Klk1	15380.80	-2.16	0.000218	16612	kallikrein 1
Top2a	7175.88	-2.12	0.000218	21973	topoisomerase (DNA) II alpha
Nuf2	639.83	-2.18	0.000499	66977	NUF2 NDC80 kinetochore complex component
Mki67	12477.05	-2.02	0.000499	17345	antigen identified by monoclonal antibody Ki 67
Cdca2	759.61	-2.04	0.000519	108912	cell division cycle associated 2
Cdc25c	374.52	-2.49	0.000523	12532	cell division cycle 25C
Cdk1	2132.71	-2.30	0.000523	12534	cyclin-dependent kinase 1
Casc5	732.79	-2.13	0.000523	NA	NA
Kif18b	609.51	-2.03	0.000523	70218	kinesin family member 18B
Kif4	1299.61	-2.12	0.000563	16571	kinesin family member 4
Birc5	2002.24	-2.07	0.00058	11799	baculoviral IAP repeat-containing 5
Dlgap5	750.87	-2.29	0.000617	218977	DLG associated protein 5
Nov	912.12	-6.18	0.000767	18133	nephroblastoma overexpressed gene
Nusap1	1305.47	-2.11	0.000767	108907	nucleolar and spindle associated protein 1
Kifc1	746.69	-2.15	0.000773	1.01E+08	kinesin family member C1
Prc1	1904.60	-2.12	0.000778	233406	protein regulator of cytokinesis 1
Tpx2	2102.92	-2.02	0.000825	72119	TPX2 microtubule-associated
Cdca3	2361.73	-2.03	0.000948	14793	cell division cycle associated 3
Foxm1	1480.44	-2.02	0.000948	14235	forkhead box M1

Ncaph	975.29	-2.19	0.000988	215387	non-SMC condensin I complex subunit H
Sgol1	571.16	-2.08	0.000988	NA	NA
Plk1	1908.24	-2.21	0.00105	18817	polo like kinase 1
Ndc80	574.42	-2.22	0.001081	67052	NDC80 kinetochore complex component
Polq	396.62	-2.12	0.001081	77782	polymerase (DNA directed) theta
Troap	353.71	-2.17	0.001304	78733	trophinin associated protein
Aurka	1097.58	-2.06	0.001349	20878	aurora kinase A
Cdc20	2113.67	-2.08	0.001556	107995	cell division cycle 20
Melk	1197.45	-2.07	0.001717	17279	maternal embryonic leucine zipper kinase
Ska3	356.45	-2.04	0.001717	219114	spindle and kinetochore associated complex subunit 3
Kif22	1333.00	-2.08	0.00198	110033	kinesin family member 22
Aurkb	1236.08	-2.08	0.001983	20877	aurora kinase B
Espl1	972.03	-2.02	0.00223	105988	extra spindle pole bodies 1 separase
Aunip	130.98	-2.57	0.002439	69885	aurora kinase A and ninein interacting protein
Pbk	1080.14	-2.14	0.002649	52033	PDZ binding kinase
Kif15	994.29	-2.05	0.002649	209737	kinesin family member 15
Bub1	836.75	-2.01	0.00288	12235	BUB1 mitotic checkpoint serine/threonine kinase
Lpo	743.77	-2.17	0.003056	76113	lactoperoxidase
Ttk	549.82	-2.04	0.003116	22137	Ttk protein kinase
Apod	208.52	-3.66	0.003429	11815	apolipoprotein D
Ska1	184.75	-2.17	0.003429	66468	spindle and kinetochore associated complex subunit 1
Cenpu	230.75	-2.14	0.003443	71876	centromere protein U
Shcbp1	616.00	-2.09	0.003478	20419	Shc SH2-domain binding protein 1
Sapcd2	577.91	-2.10	0.004117	72080	suppressor APC domain containing 2
Gtse1	689.81	-2.02	0.005935	29870	G two S phase expressed protein 1
Cdkn3	433.13	-2.10	0.007089	72391	cyclin-dependent kinase inhibitor 3
Mmp3	181.18	-3.59	0.007272	17392	matrix metallopeptidase 3
Oip5	185.33	-2.11	0.00762	70645	Opa interacting protein 5
Eme1	260.03	-2.17	0.007927	268465	essential meiotic structure-specific endonuclease 1
Fmod	75.62	-2.70	0.009127	14264	fibromodulin
Apitd1	140.12	-2.02	0.011464	NA	NA
Ggt1	59.81	-3.18	0.030087	14598	gamma-glutamyltransferase 1
Apol9b	61.14	-2.27	0.039556	71898	apolipoprotein L 9b
Eif2s3y	863.06	-1128.39	0.044934	26908	eukaryotic translation initiation factor 2 subunit 3 structural gene Y-linked
Apoc2	134.33	-2.46	0.048726	11813	apolipoprotein C-II

Supplementary Table 5. SI Modules and their sizes (MS), functional annotations, hub genes, and overlap with DEGs

Abbreviations:

MS: Module Size. Number of gene in certain modules.

DEG count: number of DEGs (corresponding to VA effect) that also belong to certain module

DEG ratio: DEG count/MS

Module color	MS	Representative functional annotations	Top 5% Hub Genes (ranked by connectivity)	DEG count	DEG ratio %	DEGs
black	223	alcohol dehydrogenase (NADP+) activity, bile acid binding, monooxygenase activity, glycoside metabolic process	Cml5 Akr1c19 Cyp2d22 Cyp2b10 Dus1l Nat8 Dgat1 Adh1 Aldh1a1 2010109I03Rik B930041F14Rik	0	0.00%	-
blue	3593	cell morphogenesis involved in neuron differentiation, cell surface receptor signaling pathway involved in cell-cell signaling, embryonic organ development, axon development, regulation of small GTPase mediated signal transduction, regulation of cell morphogenesis, Wnt signaling pathway, epithelial tube morphogenesis, positive regulation of cell projection organization, cell junction organization, adherens junction organization, Notch signaling pathway, actin cytoskeleton	Polr2a Spen Prrc2a Golga4 Smarcc2 Arid1a Sbf1 Brd4 Sptan1 Kmt2b Lrp1 Cep170b Atn1 Ctnna1 Birc6 Cherp Zmiz2 Usp19 Tln2 Scaf1 Wipf2 Hectd1 Wnk2 Ep400 Ubr4 Dctn1 Dync1h1 Rere Myh9 Fam160b2 Htt Ehmt2 Smarca4 Prpf8 Arhgef11 Camsap3 Tln1 Nisch Tpr Pcnxl3 Rapgef1 Zfp592 Myo18a Zc3h7b Acin1 Atxn2l Trrap R3hdm2 Vps13d Dst Larp1 Bcl9l Ptgfrn Hcfc1 Cmip Mtmr3 Gse1 Ankrd52 Arhgef12 Ankrd11 Kmt2d Hnrnpul2 Setd1b Nup98 Cep250 Tns3 Nek9 Iqsec2 Bahcc1 Itgb4 Ski Elf4 Ttyh3 Pitpnm1 Ints1 Eif4g3 Plec Ube3b Cic Casz1 Dot1l Ubr5 Xpo6 Farp1 Kat6a Baz1b Dpp9 Tbc1d10b Bcorl1 Sipa1l3 Mapkbp1 Man2a1 Ncoa3 Crebbp Nsd1 Arhgef5 Atxn2 Plekha5 Mroh1 Ep300 Zmiz1 Setd2 Zswim8 Piezo1 2900026A02Rik Ccdc97 2310067B10Rik Lmtk2 Lpp Wasf2 Plekha6 Prrc2b Med15 Atp6v0a1 Parp4 Cabin1 Gltscr1 Srcap Hnrnpul1 Gcn1l1 Nup210 Tmem131 C2cd3 Fam168b Sec16a C77080 Wiz Pom121 Fosl2 Mark2 Dock6 Zc3h4 Macf1 Notch2 Fus Coro7 Sema4g Cdc42bpb Zzef1 Wbp1l Numa1 Rai1 5830417I10Rik Glg1 Ubap2 Mtcl1 Golga3 Dgkz Heatr5b Dab2ip Nup188 Ndst2 Epas1 Baz2a Wnk1 Arap1 Adgrl1 Ncoa6 Ago2 Myo9b Tns2 Fbrs Tnrc18 Megf8 Kdm5a Atf7 Sap130 Cramp1l Supt6 Med12 Sipa1l2 Dock5 Huwe1 Rnf40 Hspg2 Myh14 Wdfy3 Tnrc6b Akap12	0	0.00%	-
brown	1909	ncRNA metabolic process, ribosome	Pfas Ahctf1 Sfpq Camk1d Dnm1 Bms1 Shank3 Brpf1 Kifc2 Morc2a	1	0.05%	Cyp26b1

		biogenesis, regulation of mRNA metabolic process, mRNA processing, circadian rhythm, active transmembrane transporter activity, neurotransmitter transporter activity	Timp3 Iars2 Nos3 Nxf1 Nckap5 Ncl Stk35 Prpf4b Mybbp1a Snrnp70 B4galnt3 Nolc1 Akap1 Sart3 Cdk11b Adamtsl5 Nckipsd Ampd2 Tnks1bp1 Nrbp2 Nfasc Gart Slc6a6 Gigyf2 Slc7a5 Ercc6 Per2 Neat1 Taf1 Mdm4 Noc2l Mavs Adam1a Secisbp2 Usp36 Dcaf7 Per1 Rrp1b Rasal2 Ftsj3 Thrap3 Plekhh1 Myom1 Snapc4 Vegfa Ddx21 Dhx9 Kat2a Gnl2 Uvssa Asxl2 Pdcd11 Dynll2 A430105I19Rik Peg3 Vamp2 Hnrnpu Tnrc6a Pld2 Meg3 Sh2b3 Kif5b Ano6 Zfp579 Kdm6b Dennd6b Lars Nprl3 Igf2 Fkbp5 Trim41 Sox7 Firre Lrig3 Nol6 Grip2 Zfp384 Urb2 Emc1 Npr1 Zfp106 Pprc1 Hspa12a Ankrd10 Ryr3			
green	593	extracellular matrix organization, leukocyte migration, positive regulation of cell motility, inflammatory response, integrin-mediated signaling pathway, basement membrane, transmembrane signaling receptor activity	Fn1 Col6a4 Lamc1 Col6a3 Atp2a3 Col4a1 Nid1 Col4a2 Pdgfrb Csf1 Eng Fam129b Igsf10 Col5a1 Col6a1 Lama4 Kcnj10 Inpp5d Col6a2 Ano7 Tspan9 Daam2 Ltbp4 Pxdn Gfra2 Rasal3 Runx3 Zfp385a Acvrl1	1	0.17%	Fcrls
grey	1315	-	-	1	0.08%	Pla2g2d
pink	202	-	Smad4 Tmc5 Unc5cl Gpr160 Pla2g5 Gm11545 Wtip Lrrc1 Pfkfb4 Tifa	1	0.50%	Defa4
red	422	defense response to other organism, regulation of innate immune response, toll-like receptor signaling pathway, pattern recognition receptor signaling pathway	Muc13 Dmbt1 Itgam Rnf19b Pdxdc1 Mfsd4 Gpr55 Wdpcp Upp1 Cfi Isx Slc7a9 Lypd8 Igf2bp2 Trim40 Ldlr Rnf149 Clca3a2 Golm1 Sqle Casp7	9	2.13%	Stom Isx Mbl2 Cfi Mmp9 Tbxas1 Clca3a2 Cxcl14 Gabra3
turquoise	5187	mitochondrion organization, ATP synthesis coupled electron transport, oxidative phosphorylation, mitochondrial gene expression, ncRNA processing, protein folding, tRNA metabolic process, rRNA processing, antimicrobial humoral response, translation factor activity, RNA binding, translation initiation factor activity	2010107E04Rik Uqcrc2 Uqcrfs1 Emc2 Atp5c1 Mrps12 Dnajc8 Atp5f1 Snrpc Psma2 Sdhb Cnbp Rpl19 Cox5a Fam213b Atp5o Chmp2a Cox7a2 Psmb6 Cript Mrps14 Esd Cox6b1 Mrpl30 Psmd6 Clrn3 Smim7 Ak6 Mrpl53 Rps15 Ndufa2 Cox4i1 Mrpl50 Pmpcb Mrps18a Prdx5 Psmb4 Vamp8 Ndufb9 Lamtor5 Ran Ndufb6 Mrpl22 Ndufs4 Snrpd2 Smu1 Commd2 Mpc2 Brk1 Bud31 Fis1 Ndufb2 Gadd45gip1 Atp5j Sdhaf2 Pon2 Myl12b Ndufa8 Cox6c Glrx5 Ywhaq Ube2a Hmgcl Mrpl54 Eif3i Mrpl51 Rpl22l1 Hprt Cox7b Zcrb1 Uqcrb Rer1 Nme2 Chmp2b Atp5g3 2700060E02Rik Echs1 Naca Ccdc53 Aprt Nts Rpl36al Mrpl43 Rpsa Hmbs Minos1 Spryd7 Mrpl41 Cox14 Ndufc2 Tmem243 Ndufa4 Emc7 4933434E20Rik Ndufb8 Igbp1 Prdx1 Psma7 Atox1 Rpl15 Atp5k Polr1d Banf1 Ndufa9 Rnf181	6	0.12%	Fcgr4 Fcna Actg1 Pilrb1 Retnla Kcnj13

			Rpl18 0610012G03Rik Uqcrq Atp5g1 Ndufb11 Rps9 Cryzl1 Atp5l Golt1a Nop10 15-Sep Uqcr10 BC004004 Trappc3 Fh1 Cox19 Ndufab1 Alg5 Sdhd Snrpd3 Cox8a Cnih1 Mrpl27 Psma6 Timm10 Bsg Pdcd6 Vamp3 Fam104a Tmem11 Nudt21 Tmem256 Tmem14c Tmem126a Rbx1 Cdc26 Ndufb5 Carkd Jtb Snrpg Gm561 Mrpl10 Vps29 Psma1 Sar1b Rps2 Rap1b Abhd11os Cox16 Malsu1 Npm1 Use1 Naa20 Ergic3 Tmem183a Nit1 Txnl1 Adipor1 1700123O20Rik Mrpl14 Chchd2 Ppib Ndufb3 Nsmce4a Hsbp1 Txn1 Eef1d B3gat3 Ndufa11 Psmd8 Mrps17 Commd4 Lamtor4 Psmb3 Mrps24 Ccndbp1 Commd1 Swi5 Slc16a1 Tm2d2 D8Ertd738e Higd2a Cdc42 Fam96a Nedd8 Pomp Ccdc32 Smim20 Lgals2 Atp5h Triap1 Myl12a Synj2bp-cox16 Srp14 Ccdc90b Actr10 Uqcr11 Dbi Coq9 Ssu72 Mrps16 Prdx2 Psmc2 Psmb2 Npm3 Cwc15 Atp6v0b Gnb2l1 Arpc5 Rps18 Psmb1 Ier3ip1 Ndufa7 Dgcr6 Atp6v1e1 Rarres2 Vdac1 Apoa1bp Vti1b Rnf7 Cd302 Rpl29 Lgals6 Tmem242 Lgals4 Smdt1 Cox5b Snrnp27 Gm12669 Aup1 Bcas2 Bri3 Commd9 Coa3 Commd6 Anapc16 Mcfd2 Eef1g Lsm3 Mrpl33 Eci3 Mrpl49 Psma4 Uqcrh Lsm5 Churc1 Cstb Ndufs5 Glrx3 Ndufb7 Nol7 Ptpmt1 1500011K16Rik Timm10b			
yellow	924	adaptive immune response, chemotaxis, positive regulation of protein secretion, positive regulation of cytokine production, cytokine receptor activity, G-protein coupled receptor activity, integrin binding	Anxa5 Wnt5a Gjb4 Serpine1 H2-M2 R3hdml Lhfp Vim Bmp5 Fgf7 Abcg5 Emp3 Acot12 Mmp10 Cldn4 Dpep2 Slc17a4 Gna15 Mmp13 Aldh18a1 Rab7b Siglech Trf Setd8 Slc4a11 Ptpn6 Cd86 Cd55 Npnt Serpina1b Ctgf Vcam1 Slc1a2 Nr0b2 Lrmp Sash3 Bcas1 Scarb1 Nek7 Mir7678 Amz1 Ano10 Fstl1 Ncoa4 Cyp4f16 Gm14005	30	3.25%	Nr0b2 Mcpt2 Gna15 Dmp1 Hba-a1 Fcgr2b Cpa3 Fgf7 Mcpt1 Scarb1 Gsta4 Ptgs2 Cd207 Ces2a Gjb4 Cwh43 Pkd1l2 R3hdml Scara5 Tm4sf4 Plxnc1 Nxpe5 Itih3 Mcpt4 Tpsb2 Hba-a2 H2-M2 Pcdh20 Bco1 Slc4a11

Supplementary Table 6. Colon Modules and their sizes (MS), functional annotations, hub genes, and overlap with DEGs

Abbreviations:

MS: Module Size. Number of gene in certain modules.

DEG count: number of DEGs (corresponding to VA effect) that also belong to certain module

DEG ratio: DEG count/MS

Module color	MS	Representative functional annotations	Top 5% Hub Genes (ranked by connectivity)	DEG count	DEG ratio %	DEGs
black	246	MHC protein complex	Tmem106a Cxcr3 Neat1 Gpr55 Rdh16 Ciita Dram2 Mgat1 Slc37a1 Ccdc116 H2-T23 Ppp6r1 Agpat1	2	0.81%	Usp2 Gpr55
blue	5674	chemical synaptic transmission, axon development, regulation of transmembrane transport, extracellular matrix, cation channel complex	Prrc2b Tbccd1 Hoxd12 Fhit Zmiz1 Tusc5 Syde2 Syn3 Gng3 1700001L05Rik Unc80 Ulk1 Chrna5 Tshz1 Zfp516 Raver2 Slc2a4rg-ps Trim44 Rprml Mrc2 Dnah5 Glis2 Ccdc157 Kazald1 Wdtc1 Dpysl3 Foxp1 Casp9 Plekha4 Ccdc15 Tmem237 Zscan18 Sesn3 Acadvl Ephx4 Camk2g 1700109H08Rik Cox7a2l Fendrr Fam193b Flt3l Rab11fip5 9330182L06Rik Fzd2 Kif3c Nkx2-2 C920006O11Rik Ddrgk1 Fabp2 Ccdc57 Adam19 Kxd1 Fn3k Gsdmc2 Gm11240 Lix1 Pcnx Fbn1 Vwa1 Tob2 Pcdhb16 Rab4a Fam189a2 Pou2f3 Igf1 Rab30 Nudt16 Tgfbi 1300002E11Rik Trpm5 Mapk8ip3 Ago3 Nr2c2 Car15 Col9a2 Anks1 Zcchc18 Slitrk1 Ism1 Usp20 Plekho2 Gm4262 Plxnb3 Lca5 Col25a1 Slu7 Nhlrc1 2610044O15Rik8 Eci3 Mmp2 Zfp846 Capn3 Cav3 Zfp346 Ppic Mir678 Tmem163 Arhgap20 Sema3d P4htm 9230114K14Rik Asphd2 Add2 Vezf1 Zfp592 Pds5b Adh1 D430019H16Rik Acad10 Klhl36 Stra6l Gpank1 Zfp595 Crb2 Casq2 Efhd1 Coro2b Ulk2 Fgfrl1 Pcgf1 Musk Htr3a Cacna1c Tmem132a Zfp94 Map3k3 4930511M06Rik Ypel1 Slc45a1 Ctnnd2 Tomm6os Emp3 1110002L01Rik Acss1 Rgs7 Rhebl1 1010001N08Rik Susd1 Rab36 Slc26a11 4930402H24Rik Hip1 Npy4r Resp18 Gm12359 Trim62 Rsl1 Hrg Col11a2 Sh3gl2 Wnt6 Pcdh20 Bcl9 Hint1 Atg10 Nr1h5 Recql5 Kcnv1 Tet1 1700123I01Rik Plekha5 Lrp12 Gm6277 Hoxa5 Adgrb3 Scamp5 Car3 Aamdc Slc7a2 Zfp748 Syt7 Fbxl16 Tbkbp1 Chd5 Sh3rf3 Zfp316 Cntln 2510009E07Rik Slc24a3 Meis1 Tceanc2 Jakmip3 Lrp3 Edn1 Itm2b Zfp629 Khk Flywch1 Cdyl2 Cyp2f2 Pank4 Dcdc2b	6	0.11%	Klk1 Cxcl14 Pla2g4c Apoc2 Ggt1 Fmod

			Arhgef37 Numbl Zfyve28 Slc10a4 Eppk1 Fam184a Ces2c Mettl7b Zmat1 Dcp1b Trpm3 Cmklr1 Sdc4 Spata2l Eng Cilp Tubal3 Gm10336 Atad3aos Cep131 Arhgap10 Zfp108 Tril Ucp3 6430573F11Rik Prune2 Fam163a Id4 C1qtnf4 Il17d Eif2ak4 Lrsam1 Fcgrt Sez6l Oaz3 Stk17b Ccbe1 Xpc Pard3b Cplx1 Zkscan8 Ggn Ptrf Sepp1 Abca8a 1700037H04Rik Socs7 Pde4a Zeb2os Tmem8b Fez1 Gabra3 Fbln2 Sulf1 Mapt Zfp141 Fat2 Map7d2 Sfrp1 Tprgl Angptl1 Ilvbl Tspyl5 Cap2 Inafm1 Hist1h1e Ddit4l Zfp157 Sepn1 Nsg2 Plcd4 Cml1 Maf Ndn Hcst Anxa6 Prmt2 Lrrc27 Zfp956 Rassf10 Gm20257 Ptgir Ugt2b35 Araf Arap3 Plekhb1 Bend4 2310001H17Rik Zfp638 Zfp329 Ryr3 Hexdc			
brown	577	-	Tmem43 Rnf39 Elmo3 Nfatc2 Pygo2 Fbxo11 Saa3 Mmp10 Fbxo11 Tlcd2 Timp1 Brap Mink1 Liph Tmem41a Wdr26 Med15 Timp1 Zbtb48 Dok7 Trim29 Ano6 Zfc3h1 Adnp Polk Zfp322a Stx19 Krt7	5	0.87%	Tff3 Fer1l4 Lpo Syt12 Nt5c1a
green	466	cytoplasmic translation, ribosome biogenesis, ncRNA processing, rRNA metabolic process, mRNA binding, nuclease activity	Mir7079 Pitrm1 Reps1 Slc38a10 Dtnbp1 Ptpro Rad51b Mfsd6l Rcl1 Ppp2cb Pdcd2 Wdr53 Polr3h Rps10 Hmga1-rs1 Snapc4 Snhg8 G6pc3 Ewsr1 Lsm11 Adh5 Ewsr1 Ccdc14 Fhl3	2	0.43%	Mmp3 Apod
greenyellow	174	transcription cofactor activity, transcription factor activity, protein binding, proximal promoter sequence-specific DNA binding	Dbp Ncoa4 Cipc Ccbl1 Trp53inp1 Rest Zswim6 Fam126b Fam76a	3	1.72%	Ciart Trp53inp1 Dbp
grey	897	-	-	7	0.78%	Nov Mir703 Ang4 Mir682 Scnn1b 9530053A07Rik Apol9b
magenta	191	-	Stard4 Phf20l1 Eif2ak1 Mob1b Amfr Brd1 Arl6ip6 Hpse Dyrk1a Zfp407	2	1.05%	Gt(ROSA)26Sor Zxdb
pink	241	leukocyte mediated immunity, adaptive immune response, lymphocyte mediated immunity, leukocyte migration, response to bacterium, T cell activation, neutrophil activation, phagocytosis, B cell mediated immunity	Mx2 Havcr2 Cxcl10 Plek Fam177a Tigit Ctrl Crlf3 Ocstamp Mzb1 2310045n01rik-mef2b Egr2	3	1.24%	Muc1 Pappa Mzb1
purple	184	anion transmembrane transporter activity	Xrn1 Rasef Ptgr2 Dgkd Tmtc2 Clic5 Akap10 Slc7a9 Ush1c Lrp6	1	0.54%	Clca1
red	258	-	Gjb3 Efna4 Cln6 Ttc26 Ngrn Smim5 S100g Fmo1 2010320M18Rik Cspp1 Rgcc Parp1 Smoc2	0	0.00%	
tan	87	positive regulation of secretion by cell, cytokine secretion	Gm14440 Il1b Galnt6 Xkr4 Arg2	0	0.00%	

turquoise	5799	mRNA processing, ncRNA metabolic process, chromosome segregation, ribosome biogenesis, RNA splicing, mitochondrion organization, establishment of protein localization to organelle, Golgi vesicle transport, tRNA metabolic process, ubiquitin-dependent protein catabolic process	Xpnpep1 Psmd2 Pmpca Psmd6 Pde12 Ywhae Tpm3 Metap2 Ppp1ca Phf5a Morf4l2 Lrrc47 Mphosph6 Nars Txnl1 Ddah1 Itpa Tor2a Lsm12 Ccdc43 Tars Pmm2 Ykt6 Mta2 Rad23b Etf1 Gnb2 Plaa Golt1b Ciapin1 Bzw1 Gars Ddx39 Eif5a Pgd Ap3m1 Med8 Rpl7l1 Ccdc124 Prelid1 Dera Cct6a Ybx1 Psme3 P4hb Hsd17b12 Mrpl4 Dnajc10 Tmem79 Chmp6 Nudcd2 Psmd13 Mrpl12 Taf2 H2afz Mrpl18 Psma3 Gorasp2 Prdx1 Tomm22 Cul2 Rab8a Cfl1 E2f4 Ube2n Shkbp1 Hmbs Tmed9 Ap2m1 Lamtor1 Kctd5 Vcp Ap1s1 Ube2l3 Smu1 Hn1l Zdhhc3 Gtf2f2 Psmc3 Actr1a Ttc13 Gspt1 Cdv3 Keap1 Fdps Rars Tmem14c Lsm6 Gmds Tfdp1 Farsa Psmd14 Dnaaf5 Cltb Slc35a4 Scamp4 Ralb Aars Ppih Ppp2r1b Rps6ka1 Gnb1 Crcp Psmd11 Tomm40 Srp72 Pmvk Psmc1 Psmc2 Lsm4 Slc30a7 Hnrnpm Mrpl36 Hspa14 Larp1b Psmd1 Afg3l1 Atpaf2 Dnajc11 Tmed2 Ccdc109b Psmd7 Pdap1 Dpp3 Anxa4 Nmt1 Hars Pkp2 Mrpl37 Sub1 Clic1 Ergic2 Arhgdia Cstf3 Srp9 Tuba1c Mydgf Ostc Dazap1 Syngr2 Nckap1 Ywhah Psmc6 Tuba4a Mrpl10 Exosc3 Med10 Eif4g1 Glo1 Eif4e Rqcd1 Ccar2 Mrpl47 Cdkn2aipnl Aggf1 Naa15 Mrps7 Dcaf13 Cdc34 Mrpl21 Bpgm Ywhaz Dpp9 Srsf9 Ccdc115 Klhdc4 Snrpb2 Psmc4 Mvp Pdcd6 Rbm14 Traf7 Nme1 Prpf38a Fam207a Ube2k Metap1 Eif2b5 Slc35c1 Rab1 Pigk Tmem126a Usp5 Ssr1 Ccdc25 Mrpl46 Hdac1 Ppp2r1a Nr2c2ap Lsm3 Ttc39a Ncbp1 Prelid2 Mrps25 Ythdf2 E2f2 Ptk2 Smarcb1 Gsr Sf3a1 Dync1li1 Bud31 Eif1ad Mrpl20 Pdia3 B3gnt3 Bbip1 Cog2 Akt1 Gdi2 Akr1a1 Pgs1 Cars Pcbp1 Rab35 Psmd8 Ppp4c Mrpl11 Mrps14 Nol7 Dolpp1 Ripk3 Mmadhc Acsl5 Strap Nadk Adrm1 Hspe1 Eno1b Znhit1 Casp8 Eno1 Psma2 Jagn1 Psmd3 Uchl3 Psmd12 Med19 Lrrc59 Asna1 Nudt5 Gnai3 Cpne2 Copb2 Prkrip1 Cops5 Atxn7l3 Ltbr Hspa4 Cdk16 Ubfd1 Kti12 Cacybp Map2k2 Dap Pak1ip1 Mrpl35 Trim27 Cks1b Fgfbp1 Dnajc25 Pnpt1 Arpc2 Gusb Gatad2a U2af2 Hnrnpf Vdac1 Adat3 Nars2 Coro1c Ap2a1 Pgk1 Tmsb10 Capzb Rsl24d1 Ercc3 Kcnk6 Gale Snd1 Ran Lsm5 Psmb7 Nploc4 Samm50 Arpc4 Rer1 Sidt1 Gnai2 Pla2g5	60	1.03%	Oip5 Melk Gtse1 Cdk1 Ncaph Nusap1 Pbk Cdca2 Cdc20 Aurka Clca4b Kif22 Ncapg Mki67 Kifc1 Dmbt1 Ckap2l Ska3 Espl1 Casc5 Aunip Kif18b Top2a Kif2c Retnlb Ttk Ska1 Cenpu Polq Cdca3 Foxm1 Cdkn3 Prc1 Prdm1 Shcbp1 Spag5 Ndc80 Kif4 Apitd1 Adgrf1 Kif15 Tnfrsf8 Birc5 Cdc25c Tpx2 Fcgbp Kif20a Isx Bub1 Iqgap3 Plk1 Sgol1 Eif2s3y Aurkb Dlgap5 Sapcd2 Eme1 Troap Gyk Nuf2
yellow	546	T cell activation, adaptive immune response, leukocyte differentiation, lymphocyte differentiation, B cell	Ppfia4 Atp8b2 Grap Ncf1 Gpnmb Celf2 Tespa1 Cybb Kcnab2 Fyb Lmo2 Pcdh15 Tnfsf8 P2ry13 Gpr171 Rasal3 Sit1 Dok3 Ceacam16 Zfp382 Ifi203 Pik3ap1 Ikzf1 Nup62-il4i1 Rorc Fermt3 Clec4n Magi2	3	0.55%	Jchain Wnt4 Derl3

	activation, mononuclear cell proliferation, leukocyte cell-cell adhesion, interferon-gamma production, positive regulation of cytokine biosynthetic process, gamma-delta T cell activation, interleukin-4 production, toll-like receptor signaling pathway, mast cell activation	

Supplementary Table 7. Preservation of SI modules according to Zsummary

Notes: if Zsummary>10, there is strong evidence that the module is preserved. If Zsummary>2 but Zsummary<10, there is weak to moderate evidence of preservation. If Zsummary<2, there is no evidence that the module preserved. The Grey module contains uncharacterized genes while the Gold module contains random genes. The table is ranked by Zsummary, from high to low.

Module color	Zsummary	Evidence of preservation
Green	13.546	Strong
Turquoise	11.471	Strong
Red	8.554	Weak to moderate
Gold	7.746	Weak to moderate
Pink	7.275	Weak to moderate
Blue	6.287	Weak to moderate
Yellow	3.519	Weak to moderate
Brown	2.544	Weak to moderate
Black	2.001	Weak to moderate
Grey	-0.619	No