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

Zhi Chai[1,2,*], Yafei Lyu[3,8], Qiuyan Chen[2], Cheng-Hsin Wei[2,9], Lindsay M. Snyder[4,10], Veronika Weaver[4], Aswathy Sebastian[5], István Albert[6], Qunhua Li[7], 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]Huck Institutes of the Life Sciences, [6]Department of Biochemistry and Molecular Biology, [7]Department of Statistics. The Pennsylvania State University, University Park, PA, USA. [8]Present address: Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [9]Present address: Frederick National Laboratory for Cancer Research, Frederick, MD, USA. [10]Present address: Center for Evolutionary and Theoretical Immunology, The University of New Mexico, Albuquerque, NM, USA.

***Corresponding authors**

A. Catharine Ross **(**acr6@psu.edu) and Zhi Chai **(**zhichaizju@gmail.com)

110 Chandlee Laboratory

University Park, PA, USA. 16802

Key words: vitamin A deficiency, small intestine, colon, gene expression profile, RNAseq, differential expression, Weighted Gene Co-expression Network Analysis (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. Using *Citrobacter rodentium* (*C. rodentium*) infection in mice as a model for diarrheal diseases in humans, previous reports showed reduced pathogen clearance and survival in vitamin A deficient (VAD) mice compared to their vitamin A sufficient (VAS) counterparts. **Objectives:** To characterize and compare the impact of preexisting 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:** VAD mice were generated by feeding VAD diet to pregnant C57/BL6 dams and their post-weaning offspring. RNAseq were performed using the total mRNAs extracted from SI and colon. Differentially Expressed Gene (DEG), Gene Ontology (GO) enrichment, and Weighted Gene Co-expression Network Analysis (WGCNA) were performed to characterize expression and co-expression patterns. **Results:** DEGs compared between VAS and VAD groups detected 49 SI and 94 colon genes. By GO information, SI DEGs were significantly enriched in categories relevant to retinoid metabolic process, molecule binding, and immune function. Immunity related pathways, including "humoral immune

response" and "complement activation" were positively associated with VA in SI. Three co-expression modules showed significant correlation with VA status in SI; these modules contained four known retinoic acid targets. In addition, other SI genes of interest (e.g. *Mbl2*, *Cxcl14*, and *Nr0b2*) in these modules were suggested as new candidate genes regulated by VA. Furthermore, our analysis showed that markers of two cell types in SI, mast cells and Tuft cells, were significantly altered by VA status. In colon, "cell division" was the only enriched category and was negatively associated with VA. Thus, comparison of co-expression modules between SI and colon indicated distinct networks under the regulation of dietary VA and suggest that preexisting VAD could have a significant impact on the host response to a variety of disease conditions.

Introduction

Vitamin A (VA), retinol and its metabolites, is an essential micronutrient for vertebrates of all ages, from early development throughout the life span. In humans, a nutritional deficiency of VA is both a major cause of xerophthalmia and a risk factor for infectious diseases [1]. VA deficiency (VAD) is estimated to affect nearly 20% of preschool aged children worldwide [2]. The impact of VA on child health is well demonstrated by World Health Organization (WHO) data showing that VA supplementation in preschool-age children has reduced diarrhea-related mortality by 33% and measles-related mortality by 50% [3].

Several lines of evidence indicate that VA is a pleiotropic regulator of gut immunity [4]. Retinoic acid (RA), the most bioactive metabolite of VA, has been shown to influence the differentiation [5] and gut homing of T cells [6], secretory IgA production [7], and the balance of different innate lymphoid cell (ILC) populations [4]. Regarding intestinal epithelial cells, VA affects both cell differentiation and function of the intestinal epithelium [8]. In murine models, VA deficiency has been associated with abnormality in goblet cell and Paneth cell ratios and activities [9], as well as weakened intestinal barrier integrity [10]. In sum, VA is a versatile regulator of both innate and adaptive immunity in the gut.

Preexisting VA status has been shown to affect the immune response in several animal models of gut infection, including infection with *Citrobacter rodentium* (*C. rodentium*), a gram negative natural mouse intestinal pathogen that mimics enteropathogenic *Escherichia coli* infection in humans [11]. Host VA status significantly affects the host response to *C. rodentium*, for vitamin A sufficient (VAS) mice cleared

the infection and survived, whereas vitamin A deficient (VAD) mice suffered a 40% mortality rate and those mice that survived the infection became non-symptomatic carriers of *C. rodentium* [12].

The initial VA status of mice at the time of infection could be an important factor in their ability to respond to gut infection. Previous studies did not comprehensively examine gene expression in the small intestine (SI) or colon of mice according to their VA status. In the present study, we sought to determine key genes that may mediate the divergent host responses between VAS and VAD mice, by comparing intestinal gene expression in naïve VAS and VAD mice prior to *C. rodentium* infection. Although *C. rodentium* infects primarily the colon [13], the SI is also relevant because it is a major site of retinol uptake into enterocytes, formation of retinyl esters for absorption into the body, and local conversion to metabolites such as RA [14]. In the steady state, retinol concentrations are significantly higher in the SI and mesenteric lymph nodes compared with most other tissues, excepting the liver, the major storage site of VA [15]. Intestinal epithelial cells and dendritic cells (DCs) in SI are producers of RA and the intestinal RA concentration generally follows a gradient from proximal to distal SI, correlating with the imprinting ability of resident DCs [15]. Therefore, both SI and colon were of interest for our present study in which we have used differential expression and co-expression analyses to characterize the effect of VA on the transcriptomic profiles in these two organs. Additional experimental groups and analyses focused on the effect of infection and interaction with VA status were also performed, and will be reported separately. Here, we have focus on the effect of VA status alone as a presumptive predetermining risk factor for the host response to gut infection.

100 **Results**

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

102 VAS and VAD mice were generated by feeding diets that were either adequate or
103 deficient in VA to pregnant mice and to their offspring for eight to ten weeks. The VA
104 status of our animal model was validated by determining serum retinol concentrations at
105 the end of the SI study [16]. Serum retinol levels were significantly lower in VAD vs. VAS
106 mice ($P<0.001$) (**Fig. 1**), showing that VA deficiency was successfully induced in our
107 model and enabling us to further analyze the effect of VA status on intestinal gene
108 expression.

109 **Distribution of differentially expressed genes in SI and colon and their co-expression**

110 We elected to use whole tissue RNAseq analysis to provide a comprehensive
111 exploratory view of the expression profiles in both SI and colon [17]. Initial mapping for the
112 SI revealed 24,421 genes, which were subsequently screened for low expression
113 transcripts and normalized by DESeq2 (**Fig. 2**). Among the 14,368 SI-expressed genes,
114 49 differentially expressed genes (DEGs) differed with VA status, and 9 co-expression
115 modules were identified. Similarly, among the 15,340 genes mapped in the colon dataset,
116 94 DEGs differed with VA status and 13 co-expression modules were identified. Of
117 these, 14,059 genes were shared between the two datasets and these were used for the
118 module preservation analysis (**Fig. 2**).

Identification of DEGs

Of the 49 DEGs in the SI that differed significantly by VA status, 18 were upregulated (higher in VAS than VAD, **Supplementary Table 1**) and 31 were downregulated (lower in VAS than VAD, **Supplementary Table 2**). A heatmap of the DEGs in the SI study is shown in **Supplementary Fig. 1**. A volcano plot of the SI DEGs (**Fig. 3a**) shows that genes with the largest negative or positive fold changes as well as highest statistical significance included *Cd207*, *Mcpt4*, and *Fcgr2b*, which were lower in VAS than VAD mice, and *Isx*, *Cyp26b1*, *Mmp9*, and *Mbl2*, which were higher in VAS than VAD mice.

Of the 94 DEGs in the colon, 34 were more highly expressed in VAS than VAD mice (**Supplementary Table 3**) and 60 genes were of lower expression (**Supplementary Table 4**). The heatmap for colon (**Supplementary Fig. 2)** and volcano plot (**Fig. 3b**) show that some of the most significantly upregulated DEGs are also well-known players in goblet cell activities, for example, *Tff3* and *Muc1* [18], known as two goblet cell markers; and *Clca1* and *Clca4b*, known to regulate mucus production and neutrophil recruitment [19]. Several genes with known regulatory functions were among the most strongly regulated in our datasets: for example, three transcription factors (TFs), *Isx*, *Prdm1*, and *Zxdb* were more highly expressed in VAS mice, while *Klk1*, a mucus regulator and *Foxm1*, a TF, were downregulated.

138 **Gene ontology (GO) and KEGG enrichment of DEGs**

139 We next performed enrichment analysis separately on all DEGs, upregulated
140 DEGs, and downregulated DEGs to compare the VAS vs VAD condition in the two
141 datasets (**Fig. 2**). Not surprisingly, the 49 SI DEGs were significantly enriched in
142 "retinoid metabolic process" (**Supplementary Fig. 3**), including in VAD mice a higher
143 expression of *Bco1*, a beta-carotene cleavage enzyme, and in VAS mice a higher
144 expression of *Cyp26b1*, a retinoic acid degrading enzyme, and *Isx*, a master regulator and
145 repressor of *Bco1* and other retinoid-pathway genes. The 18 upregulated SI DEGs were
146 significantly enriched in five Biological Processes (**Fig. 3c**), whereas the 31
147 downregulated SI DEGs were significantly enriched in 10 categories of Molecular
148 Function (**Fig. 3d**).

149 In the colon, downregulated DEGs (*n*=60) in VAS vs VAD mice were
150 significantly enriched in "Cell Division" (GO term) and "Cell Cycle" (KEGG term,
151 mmu04110), while upregulated colon DEGs (*n*=34) were not enriched in either GO or
152 KEGG pathways.

153 **WGCNA modules significantly correlated with VA status**

154 We next applied Weighted Gene Co-expression Network Analysis (WGCNA) as
155 a systems biology approach to understanding gene networks in contrast to individual
156 genes (**Fig. 2**). In the SI, setting a soft threshold power β to 7 (**Fig. 4a**), WGCNA
157 identified nine modules with module size (MS) ranging from 202 to 5,187 genes
158 (**Supplementary Table 5**). These modules in the SI network will be referred to

159 henceforth by their color labels [e.g., SI(Color)], which is a standard approach wherein
160 the colors are merely labels arbitrarily assigned to each module. Modules labeled by
161 colors are depicted in the hierarchical clustering dendrogram in **Fig. 4b**. The Grey
162 module was used to hold all genes that do not clearly belong to any other modules. The
163 gene expression patterns of each module across all samples are demonstrated as heatmaps
164 (**Supplementary Fig. 4**).

165 To determine if any of the nine modules of co-expressed SI genes were associated
166 with VA status, we tested the correlations between the module eigengenes and VA status
167 (**Fig. 4c**). Eigengene is defined as the first principal component of a given module and
168 may be considered as a representative of the gene expression profiles in that module [20].
169 Three modules were significantly correlated with VA status: the SI(Yellow), SI(Pink),
170 and SI(Red) modules, with *P* values of 6×10^{-6}, 8×10^{-4} and 0.001, respectively.

171 Among these three modules, Yellow module is the largest (MS=924), with a
172 correlation coefficient of -0.94 with VA status (**Fig. 4c**), thus, this module contains
173 transcripts that were lower in the SI of VAS mice vs. their VAD counterparts
174 (**Supplementary Fig. 4**). According to ClusterProfiler, the SI(Yellow) module is
175 functionally enriched in genes for adaptive immunity and cytokine activity
176 (**Supplementary Table 5**). The SI(Yellow) module contains known key players in RA
177 metabolism (*Bco1*, *Scarb1*, *Gsta4*, *Aldh1a2*, and *Aldh1a3*), as well as potential RA
178 targets (*Rdh1*, *Aldh3a2*, *Aldh7a1*, and *Aldh18a1*), along with several immune cell
179 markers (e.g. *Cd1d1*, *Cd28*, *Cd44*, *Cd86*, and *Cd207*).

180 The SI(Pink) module was positively correlated with VA status, however no
181 functional annotation was enriched in this module.

In contrast, the SI(Red) module was positively correlated with VA status and enriched for functional categories involved in innate immune response (**Supplementary Table 5**). SI(Red) module includes genes previously reported as VA targets *Isx*, *Ccr9*, and *Rarb*, as well as some immunological genes of interest, e.g., *Lypd8*, *Cxcl14*, *Mmp9*, *Mbl2*, *Noxa1*, *Duoxa2*, *Muc13*, *Itgae*, *Itgam*, *Madcam1*, *Il22ra2*, *Pigr*, *Clca1*, *Clca3a2*, and *Clca4b*, as the product of these genes could mediate the effects of VA status on infection. Overall, comparing VAS and VAD groups, our analysis suggests that innate immune activity is stronger, while adaptive immune response is weaker in the VAS SI.

In order to visualize the co-expression relationships within the modules of interest, for each of these three modules the top 50 genes selected by connectivity are depicted as networks in **Fig. 5**. TFs are usually more central in the network, examples include *Nr0b2* (**Fig. 5a**), *Isx* (**Fig. 5b**), and *Smad4* (**Fig. 5c**), where *Nr0b2* and *Isx* were also identified as DEGs.

Using the same WGCNA analysis on colon transcriptomes (with soft threshold power β of 26, **Fig. 4d**), 13 modules were identified and color-coded [referred by Colon(Color), **Fig. 4e** and **Supplementary Fig. 5**], with MS ranging from 87 to 5,799 genes (**Supplementary Table 6**). Four modules, Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) differed significantly with VA status, with *P* values of 5 x 10^{-4} , 0.002, 0.004, and 0.004, respectively (**Fig. 4f**). Among these four modules, Colon(Green) was enriched for "cytoplasmic translation," "ribosome biogenesis," and "ncRNA processing," while Colon(Red) was enriched for "anion transmembrane transporter activity" (**Supplementary Table 6**).

205 **Associations with Cell Markers**

206 To further query whether the expression pattern of each module was mainly
207 driven by a regulatory effect of VA on cell differentiation, a cell type marker enrichment
208 was performed. This test identified four SI modules that significantly overlap with the list
209 of SI cell markers derived from the CellMarker database [21] (**Fig. 6**). The SI(Yellow)
210 module, which as noted above is negatively correlated with VA (**Fig. 4c**), contains 24
211 Tuft cell markers, reaching a Bonferroni corrected *P* value (CP) of 4×10^{-7}. This
212 suggested a negative regulation of VA on the Tuft cell population. The SI(Green) module
213 also significantly overlapped with Tuft cell markers (CP=0.01). The SI(Turquoise)
214 module, the largest module with MS=5187 genes (**Supplementary Table 5**), was
215 significantly enriched for markers of goblet cells (CP=1×10^{-6}) and Paneth cells
216 (CP=0.001).

217 **Module preservation analysis suggests that gene networks controlled by VA are**
218 **poorly preserved between SI and colon**

219 To determine whether the modules identified as being significantly correlated
220 with VA in SI have a similar module structure in the colon, we performed module
221 preservation analysis using 14059 genes that were robustly expressed in both datasets
222 (**Fig. 2**). Network constructions were plotted for SI and colon separately. The modules
223 from this analysis are referred as mpSI(Color) and mpColon(Color). Module preservation
224 analysis was performed on mpSI and mpColon networks. For all three mpSI modules that
225 were significantly correlated with VA status (Red, Pink, and Yellow), we found only

226 “weak to moderate” evidence of preservation in the mpColon network according to
227 Zsummary (**Supplementary Table 7**). In addition, mpSI Red and Pink, two modules
228 associated with VAS status (**Fig. 4c**), overlapped merely with the mpColon modules
229 broken up into Black1 and Black2 (**Supplementary Fig. 6**) instead of an intact
230 mpColon(Black) module, which was originally also associated with VAS status (**Fig. 4f**).
231 mpSI(Yellow), another SI module significantly correlated with VA status (**Fig. 4c**), had a
232 lower preservation score than the mpSI (Red) and mpSI(Pink) modules (**Supplementary**
233 **Table 7**), and did not significantly overlap with any VA-related modules (Black, Green,
234 Purple, or Red) in the mpColon network (**Supplementary Fig. 6**).

235 Discussion

236 Differential expression analysis confirmed known targets under the control of VA in 237 SI and validated the nutritional model

238 The approach we have taken in this study was to use rapidly frozen intact tissue to
239 capture a “bird’s eye” comprehensive view of an entire organ, SI and colon individually,
240 in physiologically normal mice subjected to only dietary treatment as a means to
241 represent VA adequacy and deficiency in human populations. The use of intact tissue
242 eliminated the selective loss or enrichment of cells and genes within, such as could occur
243 during intestinal cell isolation protocols [22]. On the other hand, a limitation of our
244 approach is that signals from genes significantly regulated by VA but residing within
245 low-abundance cell types might have been overwhelmed and thereby missed. We first
246 found that of the 49 DEGs differing by VA status in SI there was a strong enrichment in

247 “retinoid metabolic process” driven by three DEGs: *Cyp26b1* (Fold Change=14.1), *Isx*
248 (Fold Change=3,231), and *Bco1* (Fold Change= -8.7) in VAS vs. VAD mice
249 (**Supplementary Table 1** and **2**). RA is known to induce the transcription of CYP26B1,
250 an enzyme that determines the magnitude of cellular exposure to RA by inactivating
251 intracellular RA, via nuclear receptor protein binding to multiple retinoic acid response
252 elements in the gene’s promoter region [23]; *Cyp26b1* has been shown to be dynamically
253 regulated in rodent liver [24], lung [25], and SI [26]. RA also induces the expression of ISX [27], a
254 transcriptional repressor for both BCO1, which catalyzes ß-carotene cleavage, and
255 SCARB1, a receptor for carotene uptake, in the intestine [28]. Thus, the sum of the
256 upregulation of *Cyp26b1* and *Isx*, as well as the downregulation of *Scarb1* and *Bco1* in
257 the VAS SI, confirms the well-established negative feedback regulation of dietary VA on
258 the intestinal RA concentration.
259 At the same time, a series of immunological genes previously reported to be
260 regulated by VA, albeit in a variety of tissues, were also identified in the SI dataset,
261 including *Mmp9*, *Cd207*, and enzymes produced by mast cells. The SI contains the
262 body’s largest population of immune cells [29], with many cell types present, and thus
263 finding genes in this category in our study was not unexpected. However, finding genes
264 in this category that are sensitive to VA status may be informative for designing future
265 studies. MMP9 is an extracellular matrix-degrading enzyme [30] that plays important
266 immunological roles in macrophages and DCs [31], wherein *Mmp9* expression is induced
267 by RA. In mouse myeloid DCs, chromatin immunoprecipitation assays showed that
268 *Mmp9* expression was enhanced by RA through a transcriptional mechanism involving
269 greater RAR-alpha promoter binding and chromatin remodeling [32]. CD207, also known

270 as langerin, is a marker of the langerin^{+} DCs, a cell type normally only abundant in skin
271 but not in the gut [33]. However, it has been reported that RA deficiency can change DC
272 subtypes in the intestine and dramatically elevate langerin^{+} DC numbers [34], which is in
273 concordance with our SI study, where *Cd207* expression was significantly higher in the
274 VAD group (**Fig. 3a** and **Supplementary Table 2**). Mast cell proteases (MCPT) 1, 2, 4,
275 and 6 (also known as TPSB2 or tryptase beta 2) were uniformly downregulated in the
276 VAS group, an observation in line with previous reports that retinol and RA inhibit the
277 growth and differentiation of human leukemic mast cells (HMC-1) [35], and the mast cells
278 derived from peripheral blood or cord blood samples [36]. In addition, peritoneal mast cell
279 numbers were also reported to be 1.5 times higher in VAD rats [37]. Overall, both
280 predictable observations based on previous analysis of the SI, and observations in the SI
281 that are congruent with the effects of VA and RA shown in other cell types, lend support
282 to the strong impact that VA nutrition has on retinoid homeostasis and immune-related
283 genes in the SI.

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

285 Our analysis also uncovered several genes that were significantly regulated by
286 VA status that may be targets for further study in the context of gut infections. One of
287 these is *Mbl2*, an active player in gut immunity, the expression of which may be
288 controlled by VA in SI (**Fig. 3a**). Mannose-binding lectin (MBL) is an important pattern-
289 recognition receptor in the innate immune system [38], and MBLs are protective against gut
290 infection by selectively recognizing and binding sugars presented on the pathogens [39].

This binding can neutralize the pathogen and inhibit infection by complement activation through opsonization or through the lectin pathway. Low levels of MBL have been related to increased susceptibility to several pathogens such as *Neisseria meningitidis, Staphylococcus aureus*, and *Streptococcus pneumoniae* [40]. In regards of the transcriptional regulation. SI is a predominant site of extrahepatic expression of MBL. Allelic variants of *Mbl2* have been associated with deficiencies in MBL protein production, which may increase the risk of mother-to-child HIV transmission [41]. However, interestingly, among infants with *Mbl2* variants, VA plus β-carotene supplementation partially counteracted the increased risk of transmission [42]. Currently, no mechanistic study has investigated how VA affects *Mbl2* gene expression; however, since *Mbl2* is an upregulated DEG ranked high in terms of both statistical significance and Fold Change (**Fig. 3a** and **Supplementary Table 1**), we postulate that VA may be an inducer of *Mbl2* gene expression. If so, studies of this gene may provide another line of evidence that VA participates in the regulation of the innate mucosal immunity.

Another candidate identified in our analysis is the *Nr0b2* (nuclear receptor subfamily 0, group B, member 2), encoding the TF small heterodimer partner (SHP). SHP is an atypical nuclear receptor that binds to a variety of nuclear receptors [43]. In liver, SHP plays an important role in controlling cholesterol, bile acid, and glucose metabolism. Hepatic *Nr0b2* is retinoid-responsive both *in vitro* and *in vivo* [44,45]. In the hepatocyte cell line AML12, *Nr0b2* was upregulated dose-dependently by RA [44,46]. However, in our SI study, *Nr0b2* was downregulated by VA (**Fig. 3a** and **Supplementary Table 2**). This may be due to organ specificity in the transcriptional regulation on *Nr0b2* gene, and/or may reflect the different roles played by liver and SI in cholesterol metabolism. *Nr0b2*

314 ranks relatively high regarding connectivity in our SI(Yellow) module (**Fig. 5a**), which is
315 in line with its annotation as a TF.

316 **VA as a potential regulator of the SI Tuft cell population**

317 An interesting observation that emerged from using the CellMarker database to
318 identify cell type-selective genes in the SI dataset was the significance enrichment of Tuft
319 cell markers in the SI(Yellow) module (**Fig. 6**). As this module is negatively correlated
320 with VAS status, this enrichment suggests that the SI Tuft cell population may be more
321 abundant under VAD status (**Fig.4c** and **Supplementary Fig. 4**). Tuft cells normally
322 represent only about 0.5% of the epithelial cells in the murine small and large intestines,
323 with a slightly more prevalent distribution in the distal SI [47]. As Tuft cells are a relatively
324 newly identified secretory cell type [48], they have not yet been well studied. We found no
325 literature on this cell type with respect to VA status. A speculation is that VAD status is
326 associated with higher numbers of ILC2 cells, which stimulate Tuft cell differentiation.
327 VA is known to regulate the balance between ILC2s and ILC3s in the intestinal mucosa [4].
328 VA deficiency has been shown to suppress ILC3 cell proliferation and function, and
329 reciprocally to promote ILC2s [4]. IL-13 is one of the ILC2 cytokines that can drive the
330 differentiation of epithelial precursors towards the Tuft lineage [47]. One of the advantages
331 of the co-expression network analysis used in our study is that cell type-specific
332 information may be derived through the analysis of whole intact tissues, without the
333 isolation of homogeneous populations of cells [22,49]. Given the small number of Tuft cells,
334 isolation may prove difficult but approaches such as single-cell RNAseq, laser capture

335 microdissection, and immunohistochemistry may be useful in the future. Although
336 markers of goblet and Paneth cell were significantly enriched in SI(Turquoise) module,
337 this module did not show significant correlation with VA status (**Fig. 4c**).

338 **VA differentially regulates SI and colon transcriptomes**

339 In general, DEGs in the SI were fewer in number but more highly regulated by
340 VA than those in the colon. The 49 DEGs in SI not only included cell markers
341 (**Supplementary Table 2**), but also genes in immunological and retinoid metabolic
342 pathways (**Fig. 3 c, d** and **Supplementary Fig. 3**). Compared with the SI, the colon was
343 enriched for fewer categories that differed by VA status. Of 94 colon DEGs, their GO
344 terms related to cell division, which suggests that the primary transcriptional role of VA
345 in colon is related to inhibition of cell proliferation and induction of epithelial
346 differentiation, two widely applicable functions of RA. Together with the upregulation of
347 goblet cell markers (*Tff3*, *Muc1*, *Clca1*, and *Clca4b*), our results are consistent with a
348 differentiation-promoting effect of VA in the colon (**Supplementary Table 3**).
349 Additionally, even though *Isx* was a DEG responsive to VA in both SI and colon, we did
350 not find *Bco1* and *Scarb1* to be differentially expressed in the colon, in line with the
351 processes of carotenoid digestion and absorption taking place primarily in SI. Other than
352 *Isx* and *Cxcl14*, a potent chemoattractant for immune cells [50], no other genes in our
353 studies were induced by VA in both organs.
354 Additionally, co-expression network analysis of the VA-correlated modules
355 revealed very distinctive module structures between the SI and colon. Although most

genes were commonly expressed in both organs (*n*=14,059, **Fig. 2**), the preservation and overlap of VA-correlated modules between the two organs were likely to be insignificant as judged from the Zsummary score (**Supplementary Table 7)** and the lack of counterpart modules (**Supplementary Fig. 6**). In sum, the networks under the influence of VA in SI appeared poorly preserved in colon. This is a further proof that although SI and colon are consecutive organs with similar developmental origins, their biological functions, especially their roles in VA metabolism, are very distinct from each other.

Conclusions

Taken together, the results of our analyses suggest the following: DEGs corresponding to VA effect are present in both the SI and colon. In SI, DEGs altered by VA are mainly involved in the retinoid metabolic pathway and immunity-related pathways. Our analysis uncovered novel target genes (e.g. *Mbl2*, *Cxcl14*, and *Nr0b2*) and moreover suggested that certain cell types—mast cells based on differential expression analysis and Tuft cells based on marker analysis—are under the regulation of VA, and therefore appear to be promising targets for further study. With regard to the colon, VA has fewer large effects but appears to play an overall suppressive role on cell division while promoting cell differentiation. Finally, the comparison of co-expression modules between SI and colon indicates distinct regulatory networks with few overlaps under the control of VA. Collectively, these data provide a new framework for further investigations of VA-regulated cell differentiation in the SI and colon, and for studies of the impact of infection on genes regulated by VA.

377 **Methods**

378 **Animals.** C57BL/6 mice, originally purchased from Jackson Laboratories (Bar
379 Harbor, ME, USA), were bred and maintained at the Pennsylvania State University
380 (University Park, PA, USA) for experiments. Mice were exposed to a 12 hour light, 12
381 hour dark cycle with *ad libitum* access to food and water. All animal experiments were
382 performed according to the Pennsylvania State University's Institutional Animal Care
383 and Use Committee (IACUC) guideline (IACUC # 43445). Vitamin A deficient (VAD)
384 and vitamin A sufficient (VAS) mice were generated as previously described [12,16,51,52].
385 Briefly, VAS or VAD diets were fed to pregnant mothers and the weanlings. VAS diet
386 contained 25 μg of retinyl acetate per day whereas VAD diet did not contain any VA. At
387 weaning, mice were maintained on their respective diets until the end of the experiments.
388 The SI study lasted 5 days following *C. rodentium* infection, whereas the colon study
389 ended during the peak of infection (post-infection day 10). Serum retinol concentrations
390 were measured by UPLC (Ultra Performance Liquid Chromatography) to verify the VA
391 status of experimental animals.

392 **UPLC serum retinol quantification.** Serum samples were saponified and
393 quantified for total retinol concentration by UPLC [53]. Briefly, serum aliquots (30–100 μl)
394 were incubated for 1 h in ethanol. Then 5% potassium hydroxide and 1% pyrogallol were
395 added. After saponification in a 55°C water bath, hexanes (containing 0.1% butylated
396 hydroxytoluene, BHT) and deionized water were added to each sample for phase
397 separation. The total lipid extract was transferred and mixed with a known amount of
398 internal standard, trimethylmethoxyphenyl-retinol (TMMP-retinol). Samples were dried
399 under nitrogen and re-constituted in methanol for UPLC analysis. The serum total retinol

concentrations were calculated based on the area under the curve relative to that of the internal standard.

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

Tissue collection and RNA extraction. Tissue were collected at the end of each study. During tissue collection, the dissected colons were first cut open to remove fecal contents. Both lower SI and colon tissues were snap frozen in liquid nitrogen immediately upon tissue harvest. In the SI study, lower SI tissue was used for total RNA extraction. In the colon study, the whole colon tissue free of colonic content was used to extract total RNA. For both tissues, total RNA extractions were performed using the Qiagen RNeasy Midi Kit according to the manufacturer's protocol. RNA concentrations were measured by Nanodrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). TURBO DNA-free kit (Ambion, Austin, TX, USA) were used to remove genomic DNA. RNA quality, assessed by RNA Integrity Number (RIN), were determined by Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). RNA samples of sufficient quality (RIN > 8) were used to construct RNA-seq libraries.

RNAseq library preparation, sequencing and mapping. Library preparation and sequencing were done in the Penn State Genomics Core Facility (University Park, PA, USA). 200ng of RNA per sample were used as the input of TruSeq Stranded mRNA Library Prep kit to make barcoded libraries according to the manufacturer's protocol (Illumina Inc., San Diego, CA, USA). The library concentration was determined by

423 qPCR using the Library Quantification Kit Illumina Platforms (Kapa Biosystems, Boston,
424 MA, USA). An equimolar pool of the barcoded libraries was made and the pool was
425 sequencing on a HiSeq 2500 system (Illumina) in Rapid Run mode using single ended
426 150 base-pair reads. 30-37 million reads were acquired for each library. Mapping were
427 done in the Penn State Bioinformatics Consulting Center. Quality trimming and adapter
428 removal was performed using trimmomatic [54]. The trimmed data were then mapped to the
429 mouse genome (mm10, NCBI) using hisat2 alignment program. Hisat2 alignment was
430 performed by specifying the 'rna-strandness' parameter. Default settings were used for all
431 the other parameters [55]. Coverages were obtained using bedtools [56]. Mapped data was
432 visualized and checked with IGV [57]. Reads mapped to each gene were counted using
433 featureCounts [58]. The two RNAseq datasets reported in this article are available in NCBI
434 Gene Expression Omnibus (GEO) under the accession number GSE143290.

435 **Data pre-processing.** The raw count of *n*=24,421 transcripts was obtained after
436 mapping in each of the two studies. Transcripts with low expression levels were first
437 removed, for the purpose of assuring that each analysis would be focused on the tissue-
438 specific mRNAs and would not be disrupted by the genes not expressed in that organ. In
439 the SI study, if the expression level of a gene was lower than 10 in more than eight of the
440 samples, or if the sum of the expression level in all 12 samples was lower than 200, this
441 transcript was removed from the later analysis. In the colon study, if the expression level
442 was lower than 10 in more than 10 of the samples, or if the sum of the expression level in
443 all 16 samples was lower than 220, the transcript was regarded as lowly expressed gene
444 and removed. As a result, 14,368 and 15,340 genes were remained in the SI and colon
445 datasets, respectively. Both datasets, following normalization according to DESeq2

446 package, were used for the following DE analysis and co-expression network analysis
447 (**Fig. 2**).
448 **Differential expression**. Differential expression (DE) analysis were performed
449 using the DESeq2 package [59]. Because of the two-by-two factorial design, the effect of
450 VA, *C. rodentium* infection and VA x *C. rodentium* Interaction were examined
451 separately. This article focused on the VA effect, with adjusted $P<0.05$ (to ensure
452 statistical significance) and |Fold change|> 2 (to ensure biological significance) as the
453 criteria for differentially expressed genes (DEGs). To visualize the result and prioritize
454 genes of interest, heatmaps and volcano plots were depicted using the R packages
455 pheatmap, ggplot, and ggrepel.
456 **WGCNA.** For consistency and comparability, the same data matrix in differential
457 expression were used to build signed co-expression networks using WGCNA package in
458 R [60]. In other words, the dataset went through the same screening and normalization
459 process as the differential expression analysis (**Fig. 2**). For each set of genes, a pair-wise
460 correlation matrix was computed, and an adjacency matrix was calculated by raising the
461 correlation matrix to a power. A parameter named Softpower was chosen using the scale-
462 free topology criterion [61]. An advantage of weighted correlation networks is the fact that
463 the results are highly robust with respect to the choice of the power parameter. For each
464 pairs of genes, a robust measure of network interconnectedness (topological overlap
465 measure) was calculated based on the adjacency matrix. The topological overlap (TO)
466 based dissimilarity was then used as input for average linkage hierarchical clustering.
467 Finally, modules were defined as branches of the resulting clustering tree. To cut the
468 branches, a hybrid dynamic tree-cutting algorithm was used [62]. To obtain moderately

469 large and distinct modules, we set the minimum module size to 30 genes and the
470 minimum height for merging modules at 0.25. As per the standard of WGCNA tool,
471 modules were referred to by their color labels henceforth and marked as SI(Color) or
472 Colon(Color). Genes that did not belong to any modules were assigned to the Grey
473 module. Color assignment together with hierarchical clustering dendrogram were
474 depicted. To visualize the expression pattern of each module, heatmap were drawn using
475 the pheatmap package in R. Eigengenes were computed and used as a representation of
476 individual modules. Eigengenes are the first principal component of each set of module
477 transcripts, describing most of the variance in the module gene expression. Eigengenes
478 provides a coarse-grained representation of the entire module as a single entity [20]. To
479 identify modules of interest, module-trait relationships were analyzed. The modules were
480 tested for their associations with the traits by correlating module eigengenes with trait
481 measurements, i.e., categorical traits (VA status, infection status, and gender) and
482 numerical traits (shedding, body weight, body weight change percentage, and RIN score).
483 The correlation coefficient and significance of each correlation between module
484 eigengenes and traits were visualized in a labeled heatmap, color-coded by blue (negative
485 correlation) and red (positive correlation) shades. Because the present paper has a focus
486 on the VA effect, the only trait reported was the VA status. The correlation results from
487 all other traits were beyond the scope of this article and will be discussed elsewhere.
488 **Functional enrichment analyses.** Gene ontology (GO) and KEGG enrichment
489 were assessed using clusterProfiler (v3.6.0) in R [63]. For DEG list and co-expression
490 modules, the background was set as the total list of genes expressed in the SI (14,368
491 genes) or colon (15,340 genes), respectively. The statistical significance threshold level

492 for all GO enrichment analyses was p.adj (Benjamini and Hochberg adjusted for multiple
493 comparisons) <0.05.

494 **Cell type marker enrichment.** The significance of cell type enrichment was
495 assessed for each module in a given network with a cell type marker list of the same
496 organ, using a one-sided hypergeometric test (as we were interested in testing for over-
497 representation). To account for multiple comparisons, a Bonferroni correction was
498 applied on the basis of the number of cell types in the organ-specific cell type marker list.
499 The criterion for a module to significantly overlap with a cell marker list was set as a
500 corrected hypergeometric *P* value < 0.05 [22,49,64]. The cell type marker lists of SI and colon
501 were obtained from CellMarker database (http://biocc.hrbmu.edu.cn/CellMarker/), a
502 manually curated resource of cell markers [21].

503 **Module visualization.** A package named igraph was used to visualize the co-
504 expression network of genes in each individual modules based on the adjacency matrix
505 computed by WGCNA. The connectivity of a gene is defined as the sum of connection
506 strength (adjacency) with the other module members. In this way, the connectivity
507 measures how correlated a gene is with all other genes in the same module [65]. The top 50
508 genes in each module ranked by connectivity were used for the visualization. In order to
509 further improve network visibility, adjacencies equal to 1 (adjacencies of a gene to itself)
510 or lower than 0.7 (less important co-expression relationships) were not shown in the
511 diagram. Each node represents a gene. The width of the lines between any two nodes was
512 scaled based on the adjacency of this gene pair. The size of the node was set proportional
513 to the connectivity of this gene. For the shape of the nodes, squares were used to
514 represent all the mouse transcription factors (TFs), based on AnimalTFDB 3.0 database

515 (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/) [66]. Circles represent all the non-TFs. In
516 terms of the color of the nodes, upregulated DEGs in the VAS condition were marked
517 with red color and downregulated DEGs in blue. For the font of the gene labels, red bold
518 italic was used to highlight genes which are cell type markers according to CellMarker
519 database (http://biocc.hrbmu.edu.cn/CellMarker/) [21].

520 **Module preservation.** Genes with low expression (based on the criteria stated in
521 Differential Expression session) in the SI or Colon datasets were excluded from this
522 analysis. In other words, only the probes that were robustly expressed in both organs
523 were used for module preservation analysis. Therefore, reconstruction of co-expression
524 networks were done in both datasets separately, with minimum module size = 30 and the
525 minimum height for merging modules = 0.25 [60-62]. All the modules from this analysis
526 were marked as mpSI(Color) or mpColon(Color).

527 Color assignment was remapped so that the corresponding modules in the SI
528 network and the re-constructed mpSI network sharing the similar expression pattern were
529 assigned with the same color. Similarly, the modules in the re-constructed mpColon
530 network were also assigned same colors according to the expression pattern resemblance
531 with their corresponding Colon modules. In order to quantify network preservation
532 between mpSI and mpColon networks at the module level, permutation statistics is
533 calculated in R function modulePreservation, which is included in the updated WGCNA
534 package [65]. Briefly, module labels in the test network (mpColon network) were randomly
535 permuted for 50 times and the corresponding preservation statistics were computed,
536 based on which Zsummary, a composite preservation statistic, were calculated.
537 Zsummary combines multiple preservation Z statistics into a single overall measure of

538 preservation [65]. The higher the value of Zsummary, the more preserved the module is
539 between the two datasets. As for threshold, if Zsummary >10, there is strong evidence
540 that the module is preserved. If Zsummary is >2 but <10, there is weak to moderate
541 evidence of preservation. If Zsummary <2, there is no evidence that the module is
542 preserved [65]. The Grey module contains uncharacterized genes while the Gold module
543 contains random genes. In general, those two modules should have lower Z-scores than
544 most of the other modules.

545 **Module overlap.** Module overlap is the number of common genes between one
546 module and a different module. The significance of module overlap can be measured
547 using the hypergeometric test. In order to quantify module comparison, in other words, to
548 test if any of the modules in the test network (mpColon network) significantly overlaps
549 with the modules in the reference network (mpSI network), the overlap between all
550 possible pairs of modules was calculated along with the probability of observing such
551 overlaps by chance. The significance of module overlap was assessed for each module in
552 a given network with all modules in the comparison network using a one-sided
553 hypergeometric test (as we were interested in testing for over-representation) [49]. To
554 account for multiple comparisons, a Bonferroni correction was applied on the basis of the
555 number of modules in the comparison network (mpColon network). Module pair with
556 corrected hypergeometric P value < 0.05 was considered a significant overlap.

557 **Acknowledgements**

558 We would like to acknowledge the sequencing service performed by C. Praul and
559 colleagues in the Genomic Core Facility at the Pennsylvania State University. The study
560 was supported by research funding (R56 AI114972 and R01 GM109453) and training
561 grant (T32GM108563) provided by National Institute of Health (NIH) as well as J. Lloyd
562 Huck Graduate Fellowship and Dorothy Foehr Huck Chair endowment provided by the
563 Huck Institutes of Life Sciences. The funders had no role in the study design, data
564 collection and analysis, decision to publish, or preparation the manuscript.

565

566 **Author Contributions**

567 A.C.R., M.T.C., and Z.C. conceived the study, designed the experiments, and made major
568 contributions to drafting the manuscript. A.C.R. and M.T.C. supervised the projects and
569 organized the coworkers. Z.C. performed most of the experiments and data analyses. Y.L.
570 and Q.L. provided guidance on the selection and application of bioinformatics tools.
571 C.W. measured the serum retinol level using UPLC and Q.C., L.M.S., and V.W. assisted
572 with the animal studies. A.S. and I.A. contributed on mapping, data pre-processing, and
573 bioinformatics consulting. All authors reviewed the text and approved the final
574 manuscript.

575

576 **Competing interests**

577 The authors declare no competing interests.

References

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

2 West, K. P., Jr. Extent of vitamin A deficiency among preschool children and women of reproductive age. *J. Nutr.* **132**, 2857-2866 (2002).

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

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

5 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. *J. Immunol.* **152**, 1515-1522 (1994).

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

7 Mora, J. R. & von Andrian, U. H. Role of retinoic acid in the imprinting of gut-homing IgA-secreting cells. *Semin. Immunol.* **21**, 28-35 (2009).

8 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).

9 Cha, H. R. *et al.* Downregulation of Th17 cells in the small intestine by disruption of gut flora in the absence of retinoic acid. *J. Immunol.* **184**, 6799-6806 (2010).

10 Osanai, M. *et al.* Cellular retinoic acid bioavailability determines epithelial integrity: role of retinoic acid receptor alpha agonists in colitis. *Mol. Pharmacol.* **71**, 250-258 (2007).

11 Collins, J. W. *et al.* Citrobacter rodentium: infection, inflammation and the microbiota. *Nat. Rev. Microbiol.* **12**, 612-623 (2014).

12 McDaniel, K. L. *et al.* Vitamin A-deficient hosts become nonsymptomatic reservoirs of Escherichia coli-like enteric infections. *Infect. Immun.* **83**, 2984-2991 (2015).

13 Wiles, S. *et al.* Organ specificity, colonization and clearance dynamics in vivo following oral challenges with the murine pathogen Citrobacter rodentium. *Cell. Microbiol.* **6**, 963-972 (2004).

14 Blomhoff, R., Green, M. H. & Norum, K. R. Vitamin A: physiological and biochemical processing. *Annu. Rev. Nutr.* **12**, 37-57 (1992).

15 Villablanca, E. J. *et al.* MyD88 and retinoic acid signaling pathways interact to modulate gastrointestinal activities of dendritic cells. *Gastroenterology.* **141**, 176-185 (2011).

16 Smith, S. M., Levy, N. S. & Hayes, C. E. Impaired immunity in vitamin A-deficient mice. *J. Nutr.* **117**, 857-865 (1987).

17 Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773-1808 (2002).

18 Wiede, A. *et al.* Localization of TFF3, a new mucus-associated peptide of the human respiratory tract. *Am. J. Respir. Crit. Care Med.* **159**, 1330-1335 (1999).

19 Nyström, E. E. *et al.* Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity. *EBioMedicine.* **33**, 134-143 (2018).

20 Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).

21 Shi, A. *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, 721-728 (2018).
22 Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* **474**, 380-384 (2011).
23 Zhang, Y., Zolfaghari, R. & Ross, A. C. Multiple retinoic acid response elements cooperate to enhance the inducibility of CYP26A1 gene expression in liver. *Gene.* **464**, 32-43 (2010).
24 Ross, A. C. Retinoid production and catabolism: role of diet in regulating retinol esterification and retinoic acid oxidation. *The Journal of nutrition.* **133**, 291-296 (2003).
25 Wu, L. & Ross, A. C. Acidic retinoids synergize with vitamin A to enhance retinol uptake and STRA6, LRAT, and CYP26B1 expression in neonatal lung. *J. Lipid Res.* **51**, 378-387 (2010).
26 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. *Arch. Biochem. Biophys.* **401**, 235-243 (2002).
27 Lobo, G. P. *et al.* Genetics and diet regulate vitamin A production via the homeobox transcription factor ISX. *J. Biol. Chem.* **288**, 9017-9027 (2013).
28 Lobo, G. P. *et al.* ISX is a retinoic acid-sensitive gatekeeper that controls intestinal β, β-carotene absorption and vitamin A production. *FASEB J.* **24**, 1656-1666 (2010).
29 Couter, C. J. & Surana, N. K. Isolation and flow cytometric characterization of murine small intestinal lymphocytes. *JoVE.* **111**, 111-116 (2016).
30 Hijova, E. Matrix metalloproteinases: their biological functions and clinical implications. *Bratisl. Lek. Listy.* **106**, 127-132 (2005).
31 MacLauchlan, S. *et al.* Macrophage fusion, giant cell formation, and the foreign body response require matrix metalloproteinase 9. *J. Leukoc. Biol.* **85**, 617-626 (2009).
32 Lackey, D. E. & Hoag, K. A. Vitamin A upregulates matrix metalloproteinase-9 activity by murine myeloid dendritic cells through a nonclassical transcriptional mechanism. *J. Nutr.* **140**, 1502-1508 (2010).
33 Merad, M., Ginhoux, F. & Collin, M. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. *Nat. Rev. Immunol.* **8**, 935-947 (2008).
34 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 (2010).
35 Alexandrakis, M. G. *et al.* Inhibitory effect of retinoic acid on proliferation, maturation and tryptase level in human leukemic mast cells (HMC-1). *Int. J. Immunopathol. Pharmacol.* **16**, 43-47 (2003).
36 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).
37 Wiedermann, U. *et al.* Vitamin A deficiency increases inflammatory responses. *Scand. J. Immunol.* **44**, 578-584 (1996).
38 Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. Phylogenetic perspectives in innate immunity. *Science.* **284**, 1313-1318 (1999).
39 Neth, O. *et al.* Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infect. Immun.* **68**, 688-693 (2000).
40 Shi, L. *et al.* Mannose-binding lectin-deficient mice are susceptible to infection with Staphylococcus aureus. *J. Exp. Med.* **199**, 1379-1390 (2004).
41 Turner, M. W. Mannose-binding lectin (MBL) in health and disease. *Immunobiology.* **199**, 327-339 (1998).

42 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. *Am. J. Clin. Nutr.* **84**, 610-615 (2006).
43 Lu, T. T. *et al.* Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Mol. Cell.* **6**, 507-515 (2000).
44 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).
45 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. *J. Nutr. Biochem.* **47**, 63-74 (2017).
46 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).
47 Banerjee, A., McKinley, E. T., von Moltke, J., Coffey, R. J. & Lau, K. S. Interpreting heterogeneity in intestinal Tuft cell structure and function. *J. Clin. Invest.* **128**, 1711-1719 (2018).
48 Gerbe, F. *et al.* Distinct ATOH1 and Neurog3 requirements define Tuft cells as a new secretory cell type in the intestinal epithelium. *J. Cell Biol.* **192**, 767-780 (2011).
49 Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271-1282 (2008).
50 Hara, T. & Tanegashima, K. Pleiotropic functions of the CXC-type chemokine CXCL14 in mammals. *J. Biochem.* **151**, 469-476 (2012).
51 Snyder, L. M. *et al.* Retinoic acid mediated clearance of Citrobacter rodentium in vitamin A deficient mice requires CD11b+ and T Cells. *Front. Immunol.* **9** (2019).
52 Carman, J., Smith, S. & Hayes, C. Characterization of a helper T lymphocyte defect in vitamin A-deficient mice. *J. Immunol.* **142**, 388-393 (1989).
53 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. *J. Nutr.* **146**, 1953-1960 (2016).
54 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**, 2114-2120 (2014).
55 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods.* **12**, 357-360 (2015).
56 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 841-842 (2010).
57 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26 (2011).
58 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).
59 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550-571 (2014).
60 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* **9**, 559-572 (2008).
61 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4** (2005).
62 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).
63 Yu, G., Wang, L., Han, Y. & He, Q. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* **16**, 284-287 (2012).

64 Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12698-12703 (2010).

65 Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7** (2011).

66 Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, 33-38 (2018).

728 **Figure legends**

729 **Fig. 1 VAD mice have lower serum retinol levels than VAS mice.** Serum retinol
730 concentrations were measured by UPLC on post-infection day 5. Values are the mean ±
731 SEM of n=3-6 mice per group. Two-way ANOVA test. *** *P*<0.001. **Abbreviation:**
732 vitamin A (VA), vitamin A deficient (VAD), vitamin A sufficient (VAS), Ultra
733 Performance Liquid Chromatography (UPLC), standard error of the mean (SEM),
734 analysis of variance (ANOVA).
735

736 **Fig. 2 Overview of the results of bioinformatics analyses.** Initial mapping identified
737 24,421 genes. After data pre-processing, 14,368 genes in the SI study and 15,340 genes in
738 the colon study were maintained in the two datasets. Starting from the 14,368 SI-
739 expressed genes, 49 DEGs corresponded to VA effect, and 9 co-expression modules were
740 identified. Similarly for the colon study, 94 DEGs and 13 modules were identified among
741 the 15,340 colon-expressed genes in the colon study. A total of 14,059 genes were shared
742 between the two datasets and were used to for the module preservation analysis.
743

744 **Fig. 3 Transcriptional changes in the intestines corresponding to VA status. a**, **b**,
745 Volcano plots of DEGs in SI (**a**) and in colon (**b**). Red: DEGs upregulated in VAS group
746 comparing with VAD group; Blue: DEGs downregulated in VAS group comparing with
747 VAD group; Black: non-DEGs. Criteria: |Fold Change|=2 is marked out by two black
748 vertical lines x=1 and x=-1. adjp<0.05 is marked out by a horizontal black line y= -
749 Log10(0.05). **c**, **d**, GO enrichment terms upregulated (**c**) or downregulated (**d**) by VA in
750 SI. Criteria: p.adj<0.05. **Abbreviations:** padj: adjusted *P* value according to DESeq2,
751 differentially expressed gene (DEG), SI (small intestine), VAS (vitamin A sufficient),
752 VAD (vitamin A deficient), Gene ontology (GO), p.adj (BH-adjusted *P* value from the
753 hypergeometric test conducted by ClusterProfiler).
754

755 **Fig. 4 Co-expression network construction in the intestines via WGCNA. a**, **b**, **c**,
756 Results from SI study. **d**, **e**, **f**, Results from colon study. **a**, **d**, Soft threshold power beta
757 were chosen for SI (**a**) and colon (**d**) networks according to the scale-free topology
758 criteria (horizontal red lines). **b**, **e**, Clustering dendrograms of co-expression modules in
759 SI (**b**) and colon (**e**). Transcripts are represented as vertical lines and are arranged in
760 branches according to topological overlap similarity using the hierarchical clustering. The
761 dynamic treecut algorithm was used to automatically detect the nine highly connected
762 modules, referred to by a color designation. Module color designation are shown below
763 the dendrogram. **c**, **f**, Correlation of module eigengenes with VA status in SI (**c**) and colon
764 (**f**) networks. Upper numbers in each cell: Correlation coefficient between the module and
765 the trait. Lower numbers in parentheses: *P* value of the correlation. Significant
766 correlations (*P* <0.01) were obtained between VA status and the SI(Yellow), SI(Pink),
767 SI(Red) (**c**), Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) (**f**) modules.

768 **Abbreviation:** weighted gene co-expression network analysis (WGCNA), small intestine
769 (SI), vitamin A (VA).
770

771 **Fig. 5 Visualization of the co-expression network in three modules of interest in SI**. **a**,
772 SI(Yellow), **b**, SI(Red), and **c**, SI(Pink). In each of the modules of interest, the top 50
773 genes ranked by connectivity are depicted using igraph in R. Each node represents a gene.
774 The size of the node is proportional to the connectivity of the gene. The shape of the
775 nodes is coded squares = TFs, circles = non-TFs. Color of the nodes denote: Blue =
776 upregulated DEGs in VAS condition, Red =downregulated DEGs. Font: red bold italic =
777 cell type markers. Thickness of lines between nodes: proportional to the adjacency value
778 of the gene pair. In order to improve network visibility, adjacencies equal to 1 or lower
779 than 0.7 were omitted. **Abbreviations**: SI (small intestine), TF (transcription factor), VAS
780 (vitamin A sufficient).
781

782 **Fig. 6 Overlap between module genes and SI cell markers**. The cell type marker lists
783 of SI were obtained from CellMarker, a manually curated database. Upper numbers in
784 each cell: Number of probes shared between the designated module (y axis) and the cell
785 marker list (x axis). Lower numbers in parentheses: Bonferroni corrected *P* value (CP) of
786 the one-sided hypergeometric test, which evaluates if the gene list of a module
787 significantly over-represents any cell type markers. Criteria: CP<0.05. Color shade of
788 each cell is according to –log10(CP). Four significant overlaps were highlighted with red
789 shades.
790

791 **Tables**

792 The only tables are included as supplementary material.

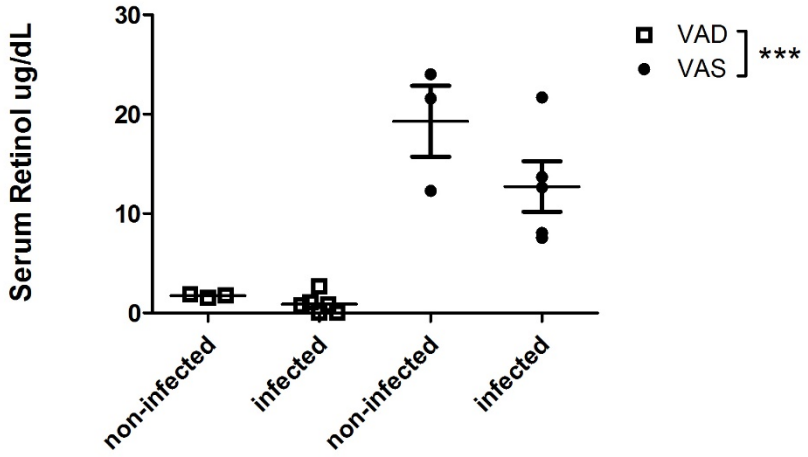
793

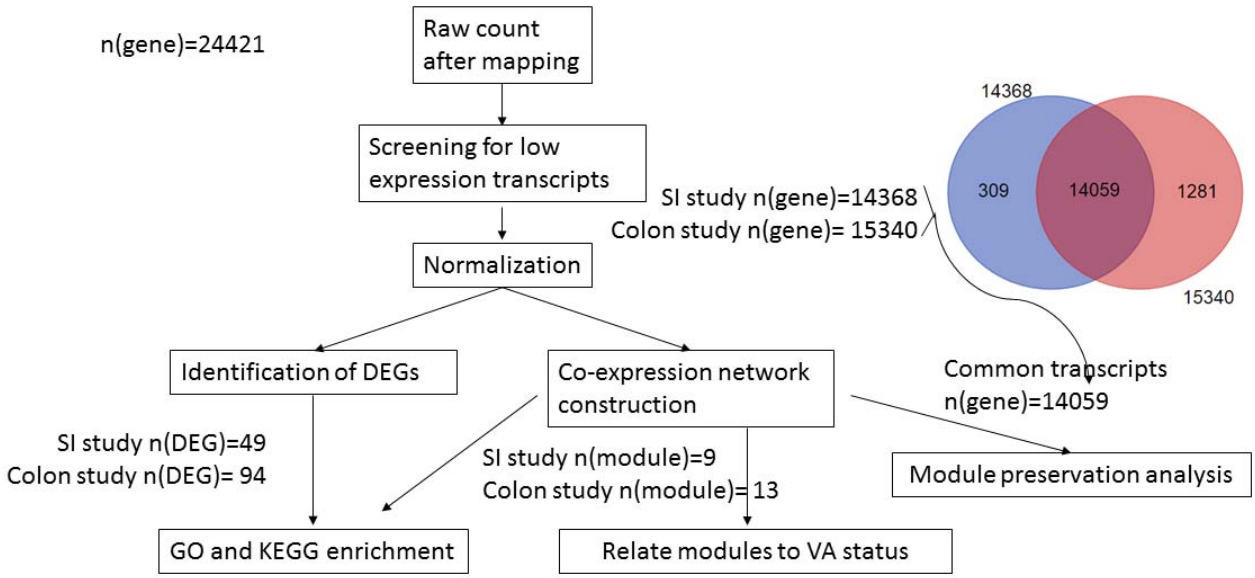
794 **Data availability**

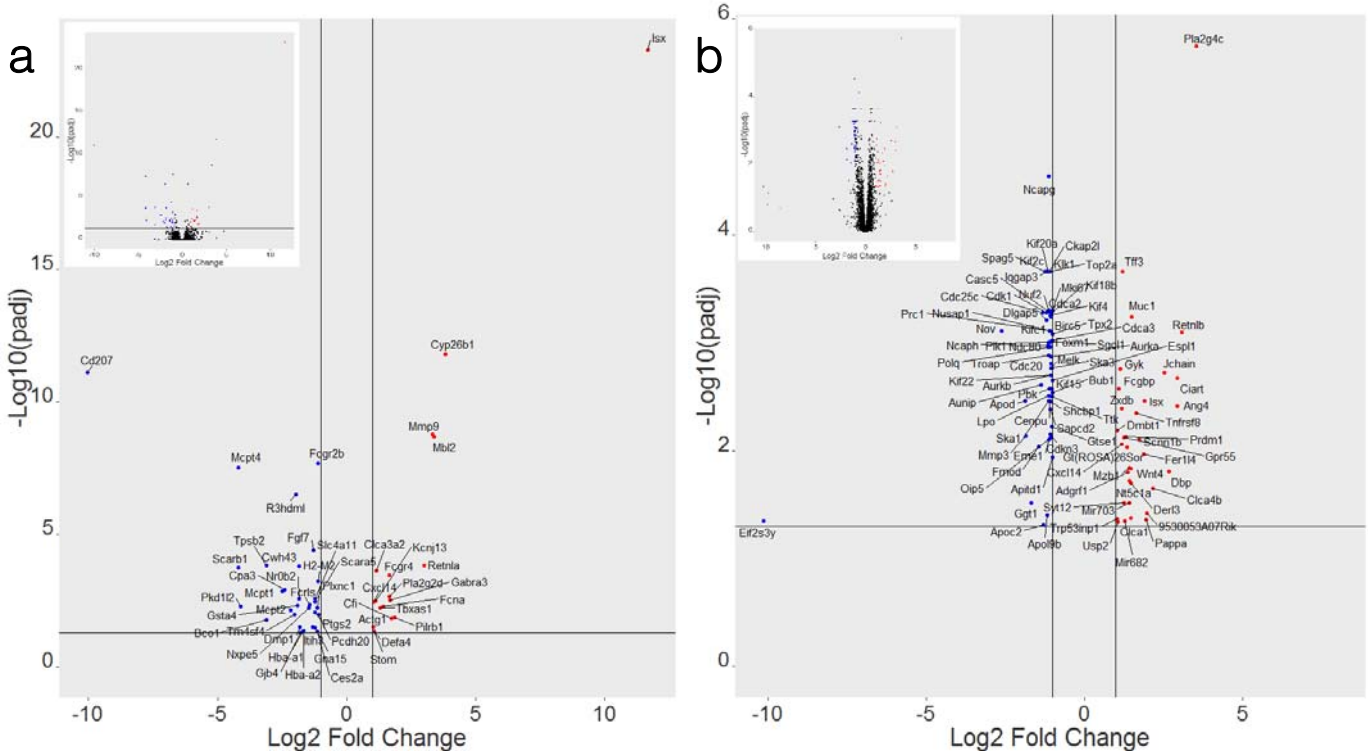
795 The RNAseq datasets generated for this study can be found in NCBI Gene Expression

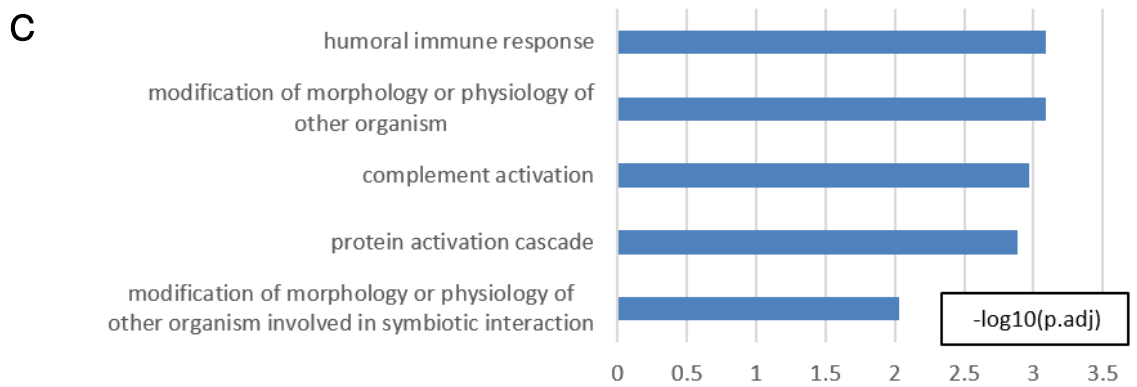
796 Omnibus (GEO) database with the accession number GSE143290 at

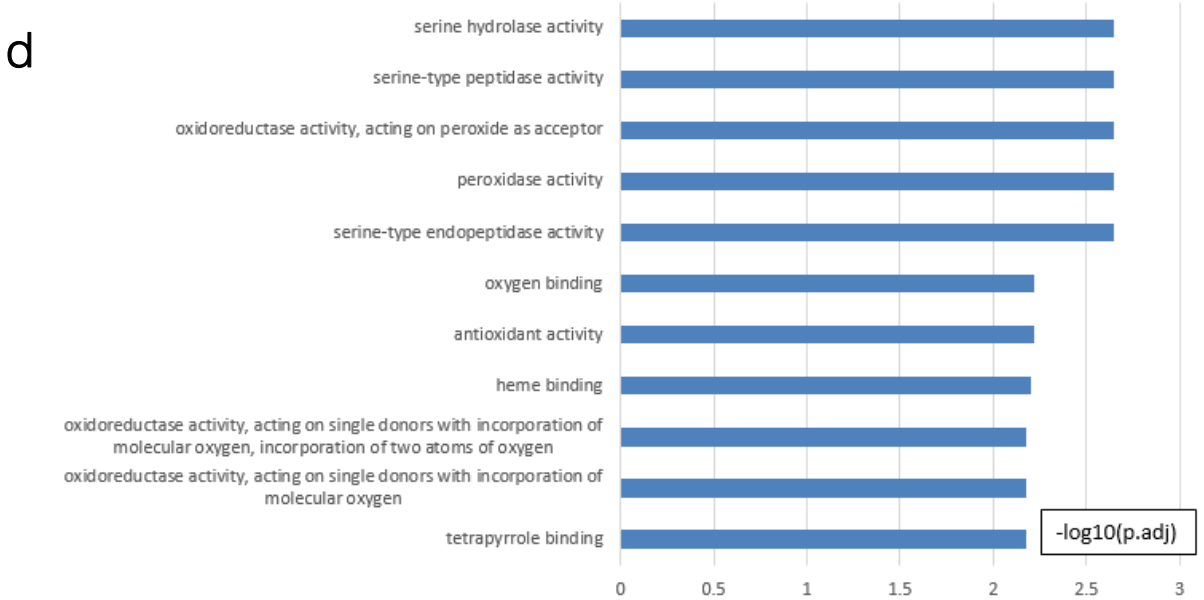
797 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143290.

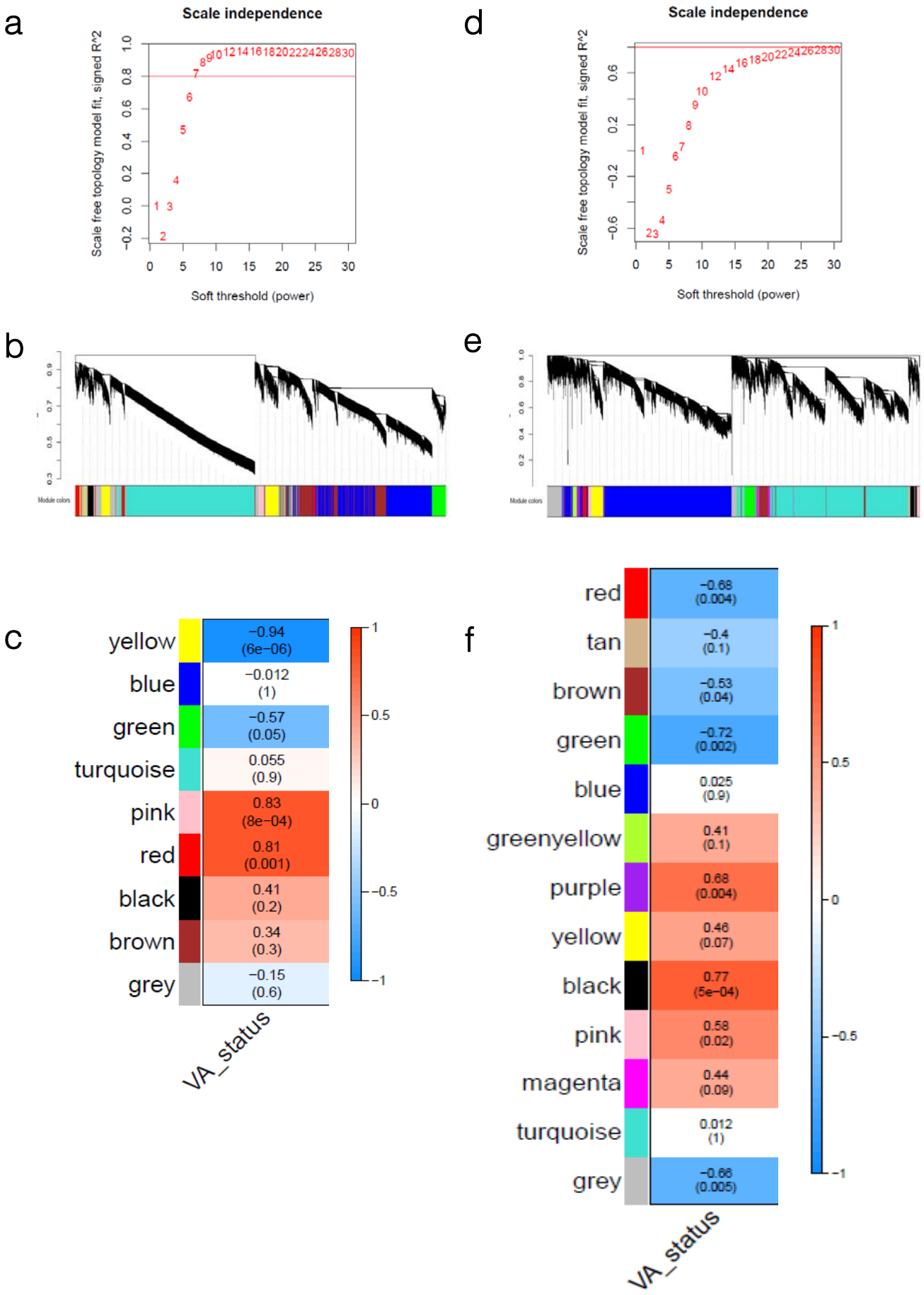

Serum Retinol ug/dL
30
20
10
0
non-infected
infected
non-infected
infected
VAD
VAS
***


n(gene)=24421
Raw count after mapping
Screening for low expression transcripts
Normalization
Identification of DEGs
Co-expression network construction
SI study n(DEG)=49
Colon study n(DEG)= 94
SI study n(module)=9
Colon study n(module)= 13
GO and KEGG enrichment
Relate modules to VA status
Module preservation analysis
SI study n(gene)=14368
Colon study n(gene)= 15340
14368
309
14059
1281
15340
Common transcripts
n(gene)=14059


a
b
-Log10(padj)
Log2 Fold Change
Isx
Cd207
Cyp26b1
Mmp9
Mbl2
Mcpt4
Fcgr2b
R3hdml
Pla2g4c
Ncapg
Eif2s3y

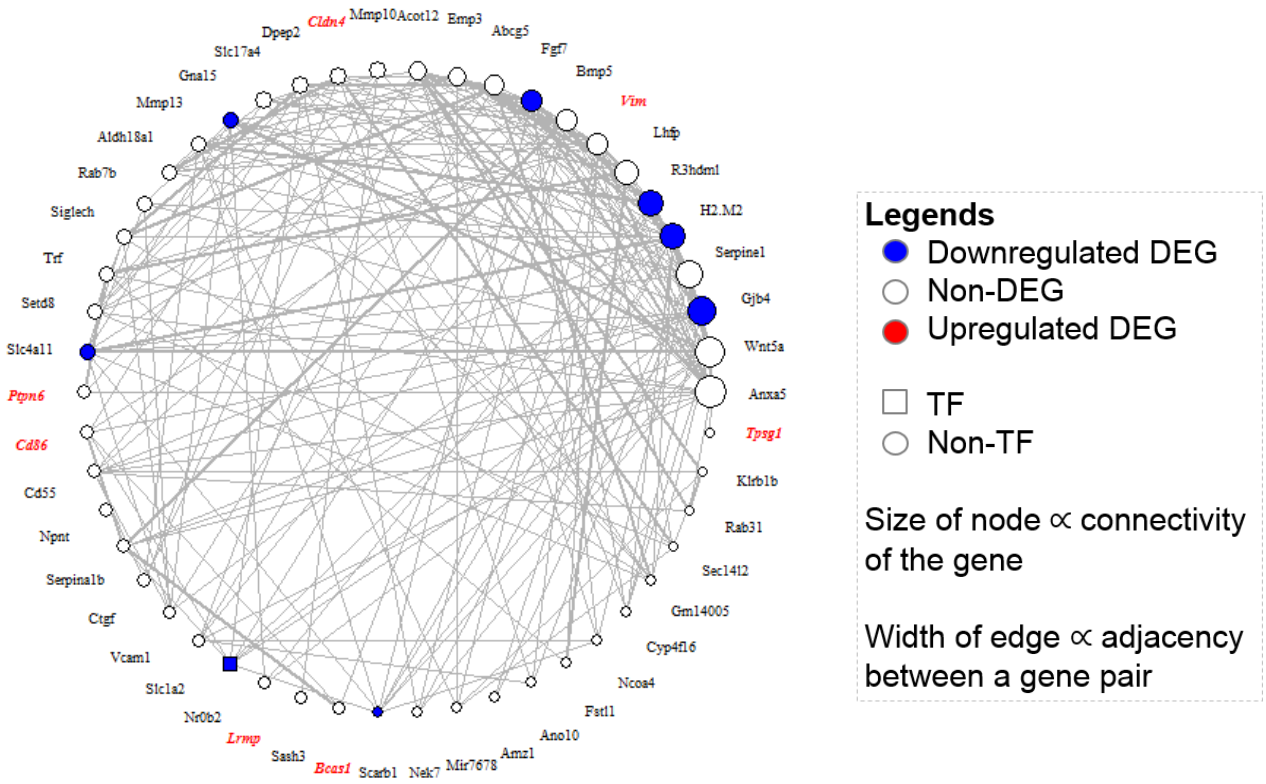

c
humoral immune response
modification of morphology or physiology of other organism
complement activation
protein activation cascade
modification of morphology or physiology of other organism involved in symbiotic interaction
-log10(p.adj)
0
0.5
1
1.5
2
2.5
3
3.5


d
serine hydrolase activity
serine-type peptidase activity
oxidoreductase activity, acting on peroxide as acceptor
peroxidase activity
serine-type endopeptidase activity
oxygen binding
antioxidant activity
heme binding
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen
tetrapyrrole binding
-log10(p.adj)
0
0.5
1
1.5
2
2.5
3


a
Scale independence
Scale free topology model fit, signed R^2
Soft threshold (power)
d
Scale independence
Scale free topology model fit, signed R^2
Soft threshold (power)
b
Module colors
e
Module colors
c
yellow −0.94 (6e−06)
blue −0.012 (1)
green −0.57 (0.05)
turquoise 0.055 (0.9)
pink 0.83 (8e−04)
red 0.81 (0.001)
black 0.41 (0.2)
brown 0.34 (0.3)
grey −0.15 (0.6)
VA_status
f
red −0.68 (0.004)
tan −0.4 (0.1)
brown −0.53 (0.04)
green −0.72 (0.002)
blue 0.025 (0.9)
greenyellow 0.41 (0.1)
purple 0.68 (0.004)
yellow 0.46 (0.07)
black 0.77 (5e−04)
pink 0.58 (0.02)
magenta 0.44 (0.09)
turquoise 0.012 (1)
grey −0.66 (0.005)
VA_status

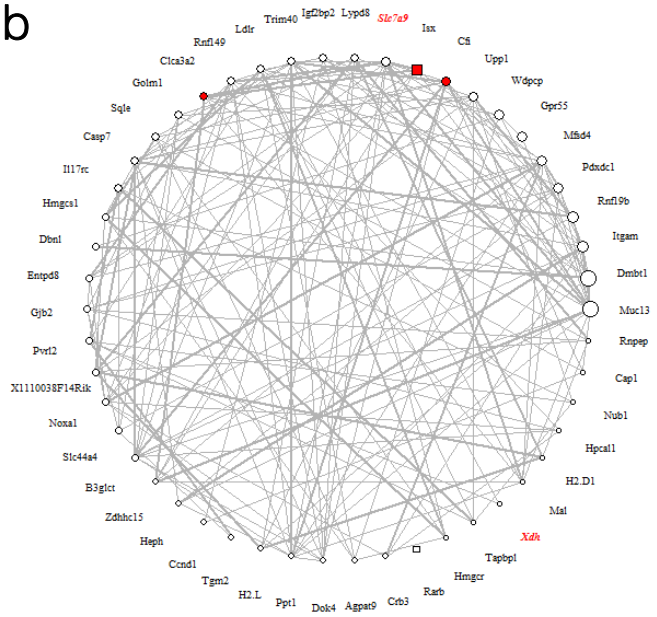
a

Co-expression Network of the Yellow Module



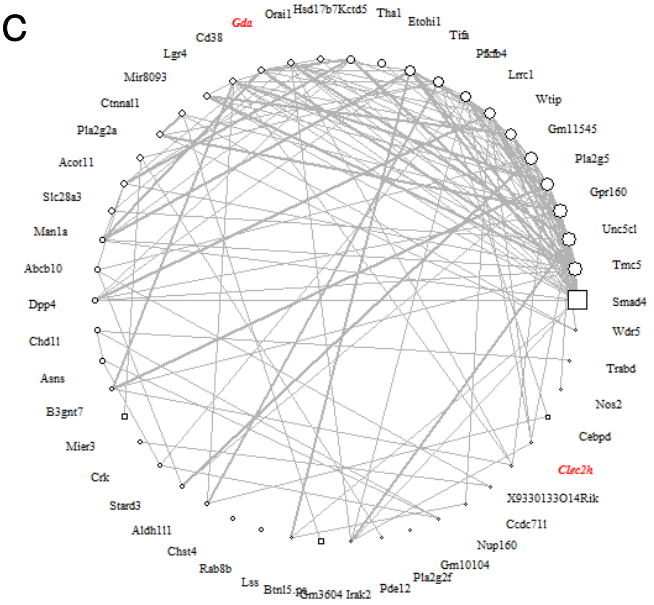
b

Co-expression Network of the Red Module



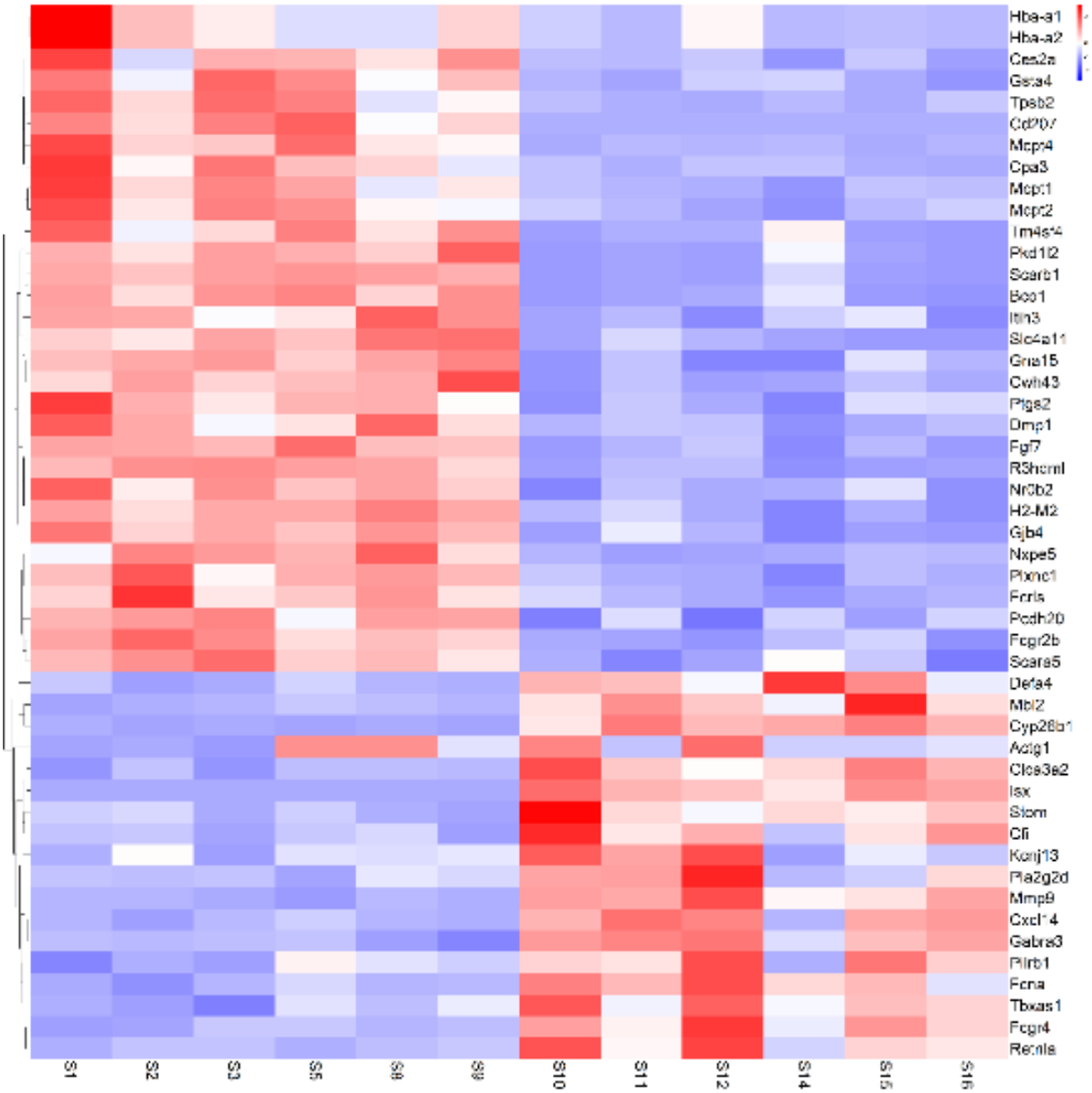
c

Co-expression Network of the Pink Module



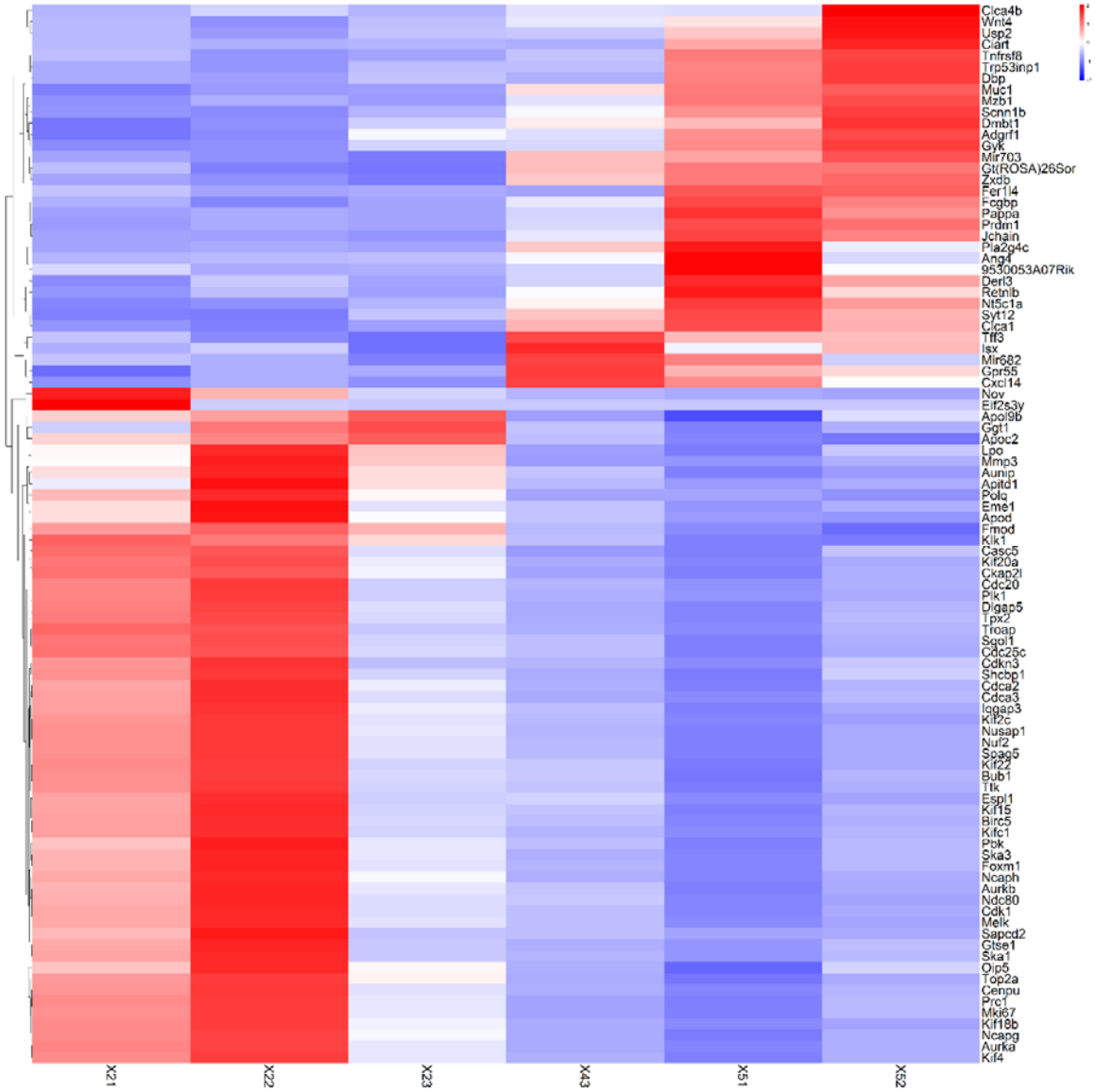
	Dendritic	Enterocyte	Enteroendocrine	Epithelial	Goblet	IgA lamina propria lymphocyte	Intestinal stem cell	Paneth	Quiescent SI stem	SI stem	Tuft_Brush
black	0 (1)	4 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	1 (1)
blue	0 (1)	9 (1)	11 (1)	0 (1)	4 (1)	1 (1)	0 (1)	0 (1)	0 (1)	0 (1)	8 (1)
brown	0 (1)	8 (1)	3 (1)	0 (1)	3 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	5 (1)
green	1 (1)	0 (1)	3 (1)	0 (1)	4 (1)	0 (1)	1 (1)	0 (1)	1 (1)	1 (1)	13 (0.01)
pink	0 (1)	4 (1)	0 (1)	0 (1)	1 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)
red	0 (1)	9 (0.1)	3 (1)	0 (1)	4 (1)	0 (1)	0 (1)	1 (1)	0 (1)	0 (1)	1 (1)
turquoise	1 (1)	41 (1)	11 (1)	0 (1)	56 (1e−06)	0 (1)	2 (1)	13 (0.001)	0 (1)	1 (1)	42 (1)
yellow	1 (1)	8 (1)	5 (1)	1 (1)	5 (1)	1 (1)	1 (1)	0 (1)	0 (1)	0 (1)	24 (4e−07)

Supplementary material



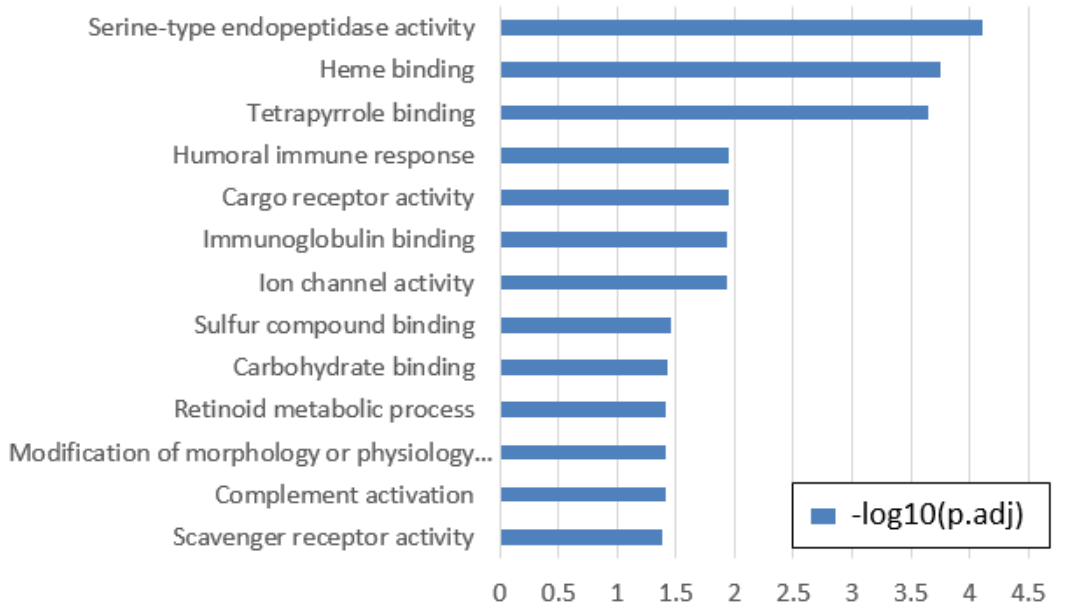
Supplementary Figure 1. Heat map for DEGs corresponding to VA effect in the SI study.

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).



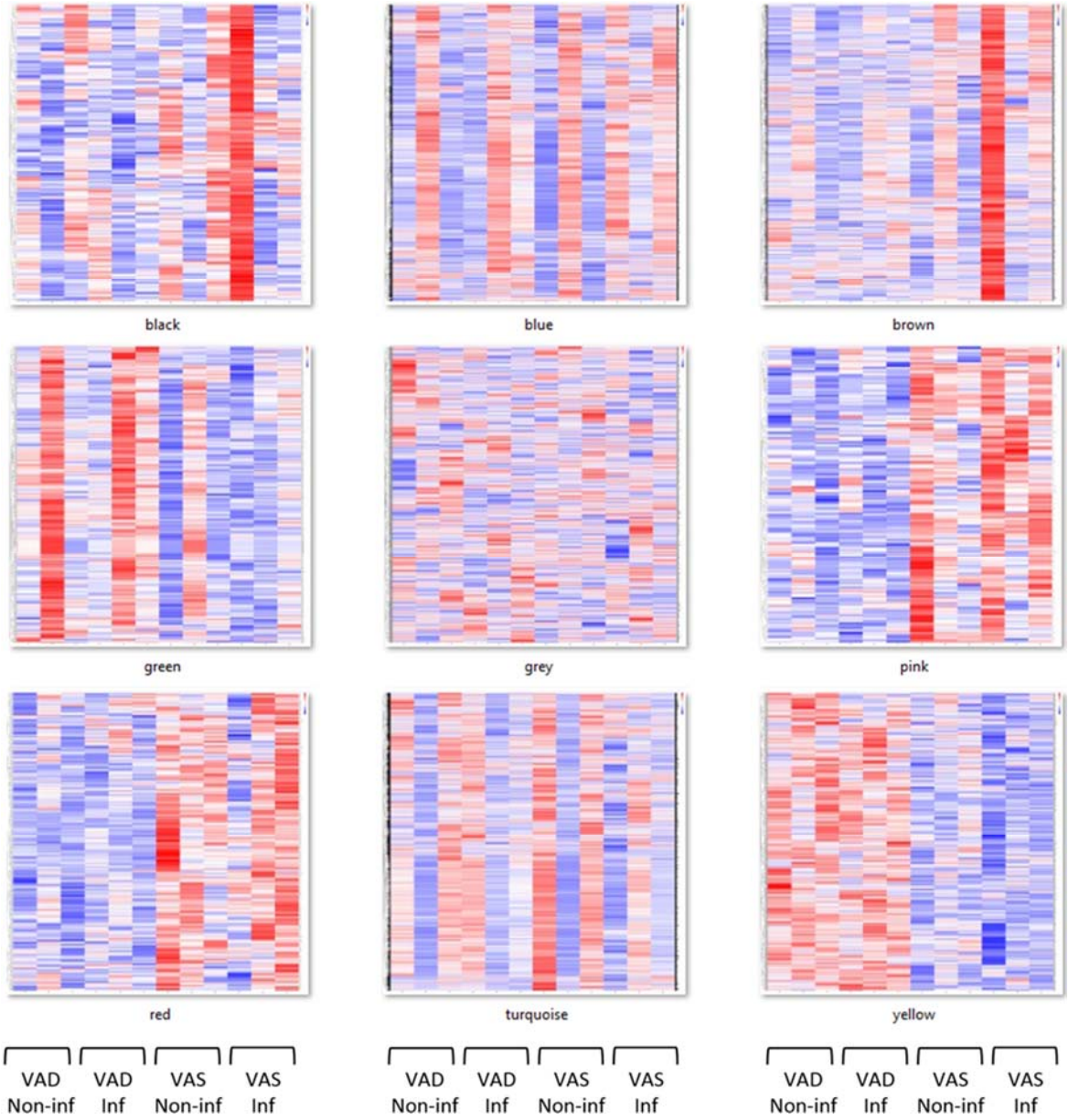
Supplementary Figure 2. Heat map for DEGs corresponding to VA effect in the colon study.

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).



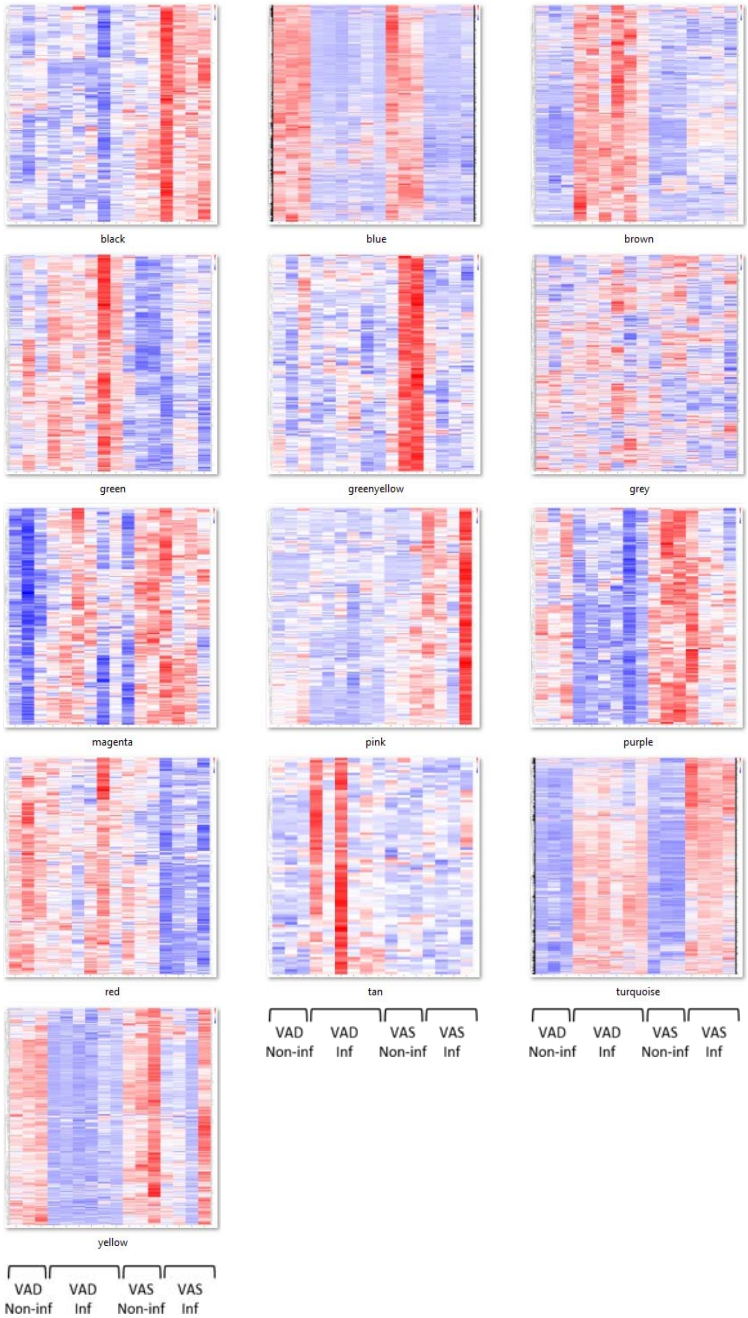
Supplementary Figure 3. GO enrichment terms of DEGs corresponding to VA effect in the SI study.

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)



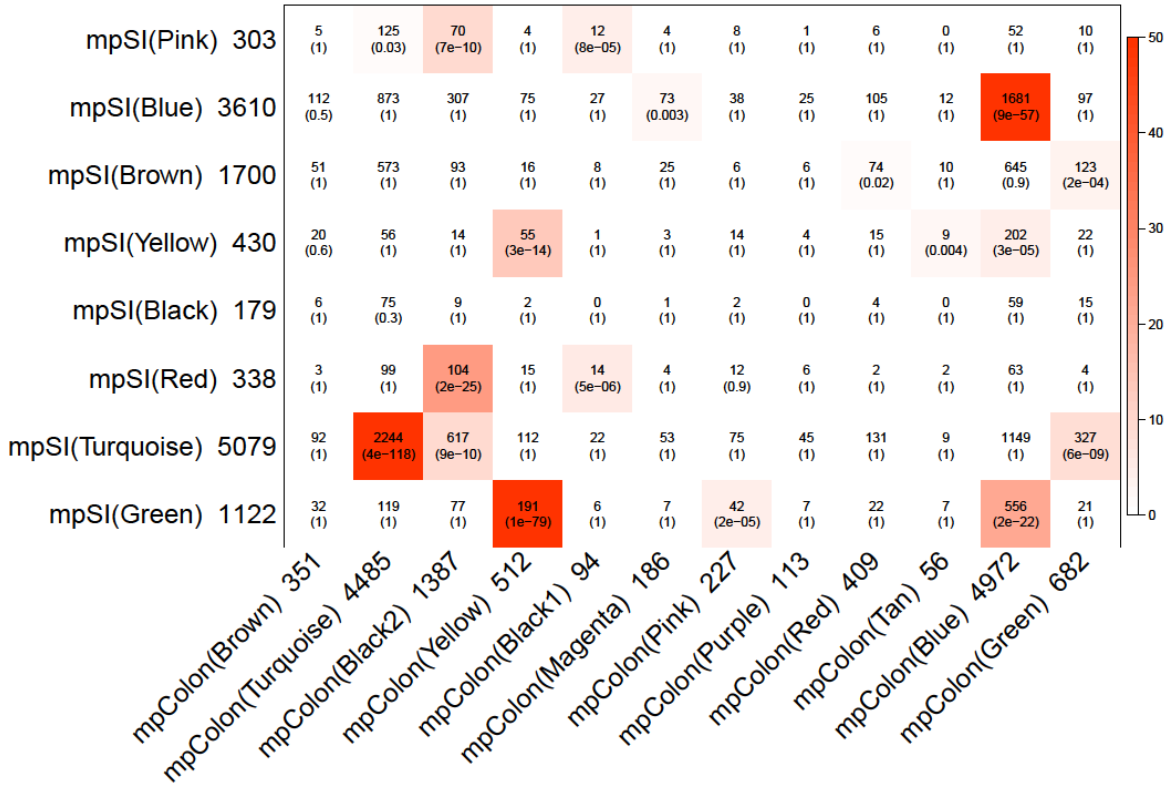
Supplementary Figure 4. Gene co-expression patterns identified by WGCNA in the SI study.

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 biogenesis, regulation	Pfas Ahctf1 Sfpq Camk1d Dnm1 Bms1 Shank3 Brpf1 Kifc2 Morc2a Timp3 Iars2 Nos3 Nxf1 Nckap5	1	0.05%	Cyp26b1

		of mRNA metabolic process, mRNA processing, circadian rhythm, active transmembrane transporter activity, neurotransmitter transporter activity	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 Rpl18 0610012G03Rik Uqcrq	6	0.12%	Fcgr4 Fcna Actg1 Pilrb1 Retnla Kcnj13

			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