HHS Public Access

Author manuscript

J Nutr Biochem. Author manuscript; available in PMC 2022 March 10.

Published in final edited form as:

J Nutr Biochem. 2021 December ; 98: 108814. doi:10.1016/j.jnutbio.2021.108814.

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

Zhi Chai[a,b,1], **Yafei Lyu**[c,2], **Qiuyan Chen**[b], **Cheng-Hsin Wei**[b,3], **Lindsay M. Snyder**[d,4], **Veronika Weaver**[d], **Aswathy Sebastian**[e], **István Albert**[f], **Qunhua Li**[g], **Margherita T. Cantorna**[d], **A. Catharine Ross**[b,*]

[a]Intercollege Graduate Degree Program in Physiology, The Pennsylvania State University, University Park, Pennsylvania, USA

[b]Department of Nutritional Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

[c]Intercollege Graduate Degree Program in Bioinformatics and Genomics, The Pennsylvania State University, University Park, Pennsylvania, USA

[d]Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

[e]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

[f]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania, USA

[g]Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, USA

Abstract

*Corresponding author at: 110 Chandlee Laboratory, University Park, Pennsylvania, USA. 16802. acr6@psu.edu (A.C. Ross).

[1]**Present addresses** : Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

[2]**Present addresses:** Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]**Present addresses:** Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

[4]**Present addresses:** Center for Evolutionary and Theoretical Immunology, The University of New Mexico, Albuquerque, New Mexico, USA

Author Contributions

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

Declaration of Competing Interest

The authors declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jnutbio.2021.108814.

Vitamin A (VA) deficiency remains prevalent in resource limited areas. Using *Citrobacter rodentium* infection in mice as a model for diarrheal diseases, previous reports showed reduced pathogen clearance and survival due to vitamin A deficient (VAD) status. To characterize the impact of preexisting VA deficiency on gene expression patterns in the intestines, and to discover novel target genes in VA-related biological pathways, VA deficiency in mice were induced by diet. Total mRNAs were extracted from small intestine (SI) and colon, and sequenced. Differentially Expressed Gene (DEG), Gene Ontology (GO) enrichment, and co-expression network analyses were performed. 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. 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, these data suggested that SI and colon have distinct networks under the regulation of dietary VA, and that preexisting VA deficiency could have a significant impact on the host response to a variety of disease conditions.

Keywords

vitamin A deficiency; small intestine; colon; gene expression profile; RNAseq; Weighted Gene Co-expression Network Analysis (WGCNA)

1. 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 preschoolage children has reduced diarrhea-related mortality by 33% and measles-related mortality by 50% [1].

Several lines of evidence indicate that VA is a pleiotropic regulator of gut immunity [3]. Retinoic acid (RA), the most bioactive metabolite of VA, functions as a pleiotropic hormone and ligand for nuclear retinoid receptors of the retinoic acid receptor (RAR) and retinoid X receptor (RXR) families. One RAR heterodimerizes with one RXR to form a dimeric complex that binds to retinoic acid response elements (RARE) in the promotor regions of numerous RA-responsive genes [4]. RA 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 [3]. 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 [13]. 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 [14], 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 [15]. 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 [16]. 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 [16]. 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.

2. Materials and methods

2.1. Animals

C57BL/6 mice, originally purchased from Jackson Laboratories (Bar Harbor, ME, USA), were bred and maintained at the Pennsylvania State University (University Park, PA, USA) for experiments. Mice were exposed to a 12-hour light, 12-hour dark cycle with *ad libitum* access to food and water. All animal experiments were performed according to the Pennsylvania State University's Institutional Animal Care and Use Committee (IACUC) guideline (IACUC # 43445). VAD and VAS mice were generated as previously described [12, 17–19]. Briefly, VAS or VAD diets were fed to pregnant mothers and the weanlings. VAS diet contained 25 μg of retinyl acetate per day whereas VAD diet did not contain any VA. At weaning, mice were maintained on their respective diets until the end of the experiments. Body weights were recorded prior to all studies. The SI study lasted 5 days following *C. rodentium* infection, whereas the colon study ended during the peak of infection (post-infection d 10). Those time points were chosen to ensure the presence of *C.*

rodentium pathogen in the target organs [12]. With respect to vitamin A status, the deficient and adequate states should remain the same, whether on d 5 or d 10 relative to infection, as the mice had been on their designated diets for at least 9 weeks. In the validation study, the VAD mice in the RA treatment group received a daily oral dose of 37.5 μg of RA in 10 μL corn oil for four consecutive days prior to the day of sample collection, whereas the VAD mice without repletion served as the control group which received 10 μL corn oil as placebo. Each of the experimental groups contained three mice (9 – 11 weeks, young adult) when tissue samples were collected. Serum retinol concentrations were measured by Ultra Performance Liquid Chromatography (UPLC) to verify the VA status of experimental animals.

2.2. Quantification of serum retinol by UPLC

Serum samples were saponified and quantified for total retinol concentration by UPLC [20]. Briefly, serum aliquots (30–100 μL) were incubated for 1 h in ethanol. Then 5% potassium hydroxide and 1% pyrogallol were added. After saponification in a 55°C water bath, hexanes (containing 0.1% butylated hydroxytoluene, BHT) and deionized water were added to each sample for phase separation. The total lipid extract was transferred and mixed with a known amount of internal standard, trimethylmethoxyphenyl-retinol (TMMP-retinol). Samples were dried 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.

2.3. 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. Mann-Whitney U tests were employed to compare the body weights. $P<.05$ was used as the cut off for a significant difference.

2.4. Tissue collection and RNA extraction

Tissues 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 separated into segments as diagramed in Fig. S1, stored in individual tubes, and snap frozen in liquid nitrogen immediately upon tissue harvest. In the SI and the validation studies, lower SI tissue was used for total RNA extraction. In the colon study, the distal 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.

2.5. 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 qPCR using the Library Quantification Kit Illumina Platforms (Kapa Biosystems, Boston, MA, USA). An equimolar pool of the barcoded libraries was made and the pool was sequencing on a HiSeq 2500 system (Illumina) in Rapid Run mode using single ended 150 base-pair reads. 30–37 million reads were acquired for each library. Mapping were done in the Penn State Bioinformatics Consulting Center. Quality trimming and adapter removal was performed using trimmomatic [21]. The trimmed data were then mapped to the mouse genome (mm10, NCBI) using hisat2 alignment program. Hisat2 alignment was performed by specifying the 'rna-strandness' parameter. Default settings were used for all the other parameters [22]. Coverages were obtained using bedtools [23]. Mapped data was visualized and checked with IGV [24]. Reads mapped to each gene were counted using featureCounts [25]. The two RNAseq datasets reported in this article are available in NCBI Gene Expression Omnibus (GEO) under the accession number GSE143290.

2.6. Data pre-processing

The raw count of n=24,421 transcripts was obtained after mapping in each of the two studies. Transcripts with low expression levels were first removed, for the purpose of assuring that each analysis would be focused on the tissue-specific mRNAs and would not be disrupted by the genes not expressed in that organ. In the SI study, if the expression level of a gene was lower than 10 in more than eight of the samples, or if the sum of the expression level in all 12 samples was lower than 200, this transcript was removed from the later analysis. In the colon study, if the expression level was lower than 10 in more than 10 of the samples, or if the sum of the expression level in all 16 samples was lower than 220, the transcript was regarded as lowly expressed gene and removed. As a result, 14,368 and 15,340 genes were remained in the SI and colon datasets, respectively. Both datasets, following normalization according to DESeq2 package, were used for the following DE analysis and co-expression network analysis (Fig. 2).

2.7. Differential expression

Differential expression (DE) analysis were performed using the DESeq2 package with the mulfi-factor designs and Wald test P values adjusted for multiple testing through the procedure of Benjamini and Hochberg [26]. Due to the two-by-two factorial design, the effect of VA, *C. rodentium* infection and VA x *C. rodentium* interaction were examined separately. This article focused on the VA effect (the comparison between non-infected VAS and the non-infected VAD groups, n=3 in each group), with adjusted P<.05 (to ensure statistical significance) and |Fold change|>2 (to ensure biological significance) as the criteria for differentially expressed genes (DEGs). To visualize the result and prioritize genes of interest, heatmaps and volcano plots were depicted using the R packages pheatmap, ggplot, and ggrepel. The validation dataset involves a pair comparison design via the DESeq2, where similar analysis and visualization were performed.

2.8. WGCNA

For consistency and comparability, the same data matrix in differential expression were used to build signed co-expression networks using WGCNA package in R [27]. In other words, the dataset went through the same screening and normalization process as the differential expression analysis (Fig. 2). As a result, all 12 and 16 samples were taken into consideration during the construction of co-expression network in SI and colon studies, respectively. For each set of genes, a pair-wise correlation matrix was computed, and an adjacency matrix was calculated by raising the correlation matrix to a power. A parameter named Softpower was chosen using the scale-free topology criterion [28]. An advantage of weighted correlation networks is the fact that the results are highly robust with respect to the choice of the power parameter. For each pairs of genes, a robust measure of network interconnectedness (topological overlap measure) was calculated based on the adjacency matrix. The topological overlap (TO) based dissimilarity was then used as input for average linkage hierarchical clustering. Finally, modules were defined as branches of the resulting clustering tree. To cut the branches, a hybrid dynamic tree-cutting algorithm was used [29]. To obtain moderately large and distinct modules, we set the minimum module size to 30 genes and the minimum height for merging modules at 0.25. As per the standard of WGCNA tool, modules were referred to by their color labels henceforth and marked as SI(Color) or Colon(Color). Genes that did not belong to any modules were assigned to the Grey module. Color assignment together with hierarchical clustering dendrogram were depicted. To visualize the expression pattern of each module, heatmap were drawn using the pheatmap package in R. Eigengenes were computed and used as a representation of individual modules. Eigengenes are the first principal component of each set of module transcripts, describing most of the variance in the module gene expression. Eigengenes provides a coarse-grained representation of the entire module as a single entity [30].
To identify modules of interest, module-trait relationships were analyzed. The modules were tested for their associations with the traits by correlating module eigengenes with trait measurements, *i.e.*, categorical traits (VA status, infection status, and gender) and numerical traits (shedding, body weight, body weight change percentage, and RIN score). The correlation coefficient and significance (cutoff set at $P<.01$) of each correlation between module eigengenes and traits were visualized in a labeled heatmap, color-coded by blue (negative correlation) and red (positive correlation) shades. Because the present paper has a focus on the VA effect, the only trait reported was the VA status. The correlation results from all other traits were beyond the scope of this article and will be discussed elsewhere.

2.9. Functional enrichment analyses

Gene ontology (GO) and KEGG enrichment were assessed using clusterProfiler (v3.6.0) in R [31]. For DEG list and co-expression modules, the background was set as the total list of genes expressed in the SI (14,368 genes) or colon (15,340 genes), respectively. The statistical significance threshold level for all GO enrichment analyses was p.adj (Benjamini and Hochberg adjusted for multiple comparisons) <0.05.

2.10. Cell type marker enrichment

The significance of cell type enrichment was assessed for each module in a given network with a cell type marker list of the same organ, using a one-sided hypergeometric test (as we were interested in testing for over-representation). To account for multiple comparisons, a Bonferroni correction was applied on the basis of the number of cell types in the organ-specific cell type marker list. The criterion for a module to significantly overlap with a cell marker list was set as a corrected hypergeometric P value < .05 [[32–34]. The cell type marker lists of SI and colon were obtained from CellMarker database (http://biocc.hrbmu.edu.cn/CellMarker/), a manually curated resource of cell markers [35].

2.11. Module visualization

A package named igraph was used to visualize the co-expression network of genes in each individual modules based on the adjacency matrix computed by WGCNA. The connectivity of a gene is defined as the sum of connection strength (adjacency) with the other module members. In this way, the connectivity measures how correlated a gene is with all other genes in the same module [36]. The top 50 genes in each module ranked by connectivity were used for the visualization. In order to further improve network visibility, adjacencies equal to 1 (adjacencies of a gene to itself) or lower than 0.7 (less important co-expression relationships) were not shown in the diagram. Each node represents a gene. The width of the lines between any two nodes was scaled based on the adjacency of this gene pair. The size of the node was set proportional to the connectivity of this gene. For the shape of the nodes, squares were used to represent all the mouse transcription factors (TFs), based on AnimalTFDB 3.0 database (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/) [37]. Circles represent all the non-TFs. In terms of the color of the nodes, upregulated DEGs in the VAS condition were marked with red color and downregulated DEGs in blue. For the font of the gene labels, red bold italic was used to highlight genes which are cell type markers according to CellMarker database (http://biocc.hrbmu.edu.cn/CellMarker/) [35].

2.12. Module preservation

Genes with low expression (based on the criteria stated in Differential Expression session) in the SI or Colon datasets were excluded from this analysis. In other words, only the probes that were robustly expressed in both organs were used for module preservation analysis. Therefore, reconstruction of co-expression networks were done in both datasets separately, with minimum module size = 30 and the minimum height for merging modules = 0.25 [[27–29]. All the modules from this analysis were marked as mpSI(Color) or mpColon(Color).

Color assignment was remapped so that the corresponding modules in the SI network and the re-constructed mpSI network sharing the similar expression pattern were assigned with the same color. Similarly, the modules in the re-constructed mpColon network were also assigned same colors according to the expression pattern resemblance with their corresponding Colon modules. In order to quantify network preservation between mpSI and mpColon networks at the module level, permutation statistics is calculated in R function modulePreservation, which is included in the updated WGCNA package [36]. Briefly, module labels in the test network (mpColon network) were randomly permuted for 50 times and the corresponding preservation statistics were computed, based on which Zsummary, a

composite preservation statistic, were calculated. Zsummary combines multiple preservation Z statistics into a single overall measure of preservation [36]. The higher the value of Zsummary, the more preserved the module is between the two datasets. As for threshold, if Zsummary >10, there is strong evidence that the module is preserved. If Zsummary is >2 but <10, there is weak to moderate evidence of preservation. If Zsummary <2, there is no evidence that the module is preserved [36]. The Grey module contains uncharacterized genes while the Gold module contains random genes. In general, those two modules should have lower Z-scores than most of the other modules.

2.13. Module overlap

Module overlap is the number of common genes between one module and a different module. The significance of module overlap can be measured using the hypergeometric test. In order to quantify module comparison, in other words, to test if any of the modules in the test network (mpColon network) significantly overlaps with the modules in the reference network (mpSI network), the overlap between all possible pairs of modules was calculated along with the probability of observing such overlaps by chance. The significance of module overlap was assessed for each module in a given network with all modules in the comparison network using a one-sided hypergeometric test (as we were interested in testing for over-representation) [33]. To account for multiple comparisons, a Bonferroni correction was applied on the basis of the number of modules in the comparison network (mpColon network). Module pair with corrected hypergeometric *P* value < .05 was considered a significant overlap.

2.14. RAR/RXR binding site prediction in promoter regions of genes of interest

CiiiDER is an integrated computational toolkit for transcription factor binding analysis [38]. It can be used to predict transcription factor binding sites (TFBSs) across the promoter regions. Promoter regions were defined as spanning −1500 bases to +500 bases relative to the transcription start site. A list of genes of interest, namely *Scarb1*, *Bco1*, *Isx*, *Cd207*, *Mbl2*, *Clca1*, *Mmp9*, and *Nr0b2*, were analyze via CiiiDER using Ensembl 94 annotations of the mouse GRCm38 genome and 2020 JASPAR core vertebrate matrices, with deficit threshold of 0.25.

3. Results

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

VAS and VAD mice were generated by feeding diets that were either adequate or deficient in VA to pregnant mice and to their offspring for eight to 10 weeks. The VA status of our animal model was validated by determining serum retinol concentrations at the end of the SI study [19]. Serum retinol levels were significantly lower in VAD *versus* VAS mice ($P<.001$) (Fig. 1), showing that VA deficiency was successfully induced in our model and enabling us to further analyze the effect of VA status on intestinal gene expression. The body weights of the VAD mice in the colon study were significantly lower than their VAS counterparts ($P=.02$, Fig. S1). However, this statistical difference was not preserved in the SI study ($P=.13$, Fig. S1).

3.2. Distribution of DEGs in SI and colon and their co-expression

We elected to use whole tissue RNAseq analysis to provide a comprehensive exploratory view of the expression profiles in both SI and colon [39]. Initial mapping for the SI revealed 24,421 genes, which were subsequently screened for low expression transcripts and normalized by DESeq2 (Fig. 2). Among the 14,368 SI-expressed genes, 49 DEGs differed with VA status, and 9 co-expression modules were identified. Similarly, among the 15,340 genes mapped in the colon dataset, 94 DEGs differed with VA status and 13 co-expression modules were identified. Of these, 14,059 genes were shared between the two datasets and were used for the module preservation analysis (Fig. 2).

3.3. Identification of DEGs

Of the 49 DEGs in the SI that differed significantly by VA status, 18 were upregulated (higher in VAS than VAD, Table S1) and 31 were downregulated (lower in VAS than VAD, Table S2). A heatmap of the DEGs in the SI study is shown in Fig. S2. 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 (Table S3) and 60 genes were of lower expression (Table S4). The heatmap for colon (Fig. S3) 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* [40], known as two goblet cell markers; and *Clca1* and *Clca4b*, known to regulate mucus production and neutrophil recruitment [41]. 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.

For the validation dataset, under the same criteria for DEG (adjusted $P<.05$ and |Fold change|> 2), four genes were found upregulated in the RA-treated VAD mice compared to the VAD mice without RA supplementation, including *Isx* and *Mbl2* (Fig. 7A and Table S8), whereas 13 genes were found downregulated in the RA-treated group, including *Bco1*, *Scarb1*, and *Cd207* (Fig. 7A and Table S9).

Full gene lists containing normalized counts, log2FoldChanges, and adjusted *P* values of the SI, colon, and the validation studies are provided in Table S10–12.

3.4. GO and KEGG enrichment of DEGs

We next performed enrichment analysis separately on all DEGs, upregulated DEGs, and downregulated DEGs to compare the VAS vs VAD condition in the two datasets (Fig. 2). Not surprisingly, the 49 SI DEGs were significantly enriched in “retinoid metabolic process” (Fig. S4), including in VAD mice a higher expression of *Bco1*, a beta-carotene cleavage enzyme, and in VAS mice a higher expression of *Cyp26b1*, a retinoic acid degrading enzyme, and *Isx*, a master regulator and repressor of *Bco1* and other retinoid-pathway

genes. The 18 upregulated SI DEGs were significantly enriched in five Biological Processes, whereas the 31 downregulated SI DEGs were significantly enriched in 11 categories of Molecular Function (Fig. 3C).

In the colon, downregulated DEGs (n=60) in VAS *versus* VAD mice were significantly enriched in “Cell Division” (GO term) and “Cell Cycle” (KEGG term, mmu04110), while upregulated colon DEGs (n=34) were not enriched in either GO or KEGG pathways.

3.5. WGCNA modules significantly correlated with VA status

We next applied WGCNA as a systems biology approach to understanding gene networks in contrast to individual genes (Fig. 2). In the SI, setting a soft threshold power β to 7 (Fig. 4A), WGCNA identified nine modules with module size (MS) ranging from 202 to 5,187 genes (Table S5). These modules in the SI network will be referred to henceforth by their color labels [e.g., SI(Color)], which is a standard approach wherein the colors are merely labels arbitrarily assigned to each module. Modules labeled by colors are depicted in the hierarchical clustering dendrogram in Figure. 4B. The Grey module was used to hold all genes that do not clearly belong to any other modules. The gene expression patterns of each module across all samples are demonstrated as heatmaps (Fig. S5).

To determine if any of the nine modules of co-expressed SI genes were associated with VA status, we tested the correlations between the module eigengenes and VA status (Fig. 4C). Eigengene is defined as the first principal component of a given module and may be considered as a representative of the gene expression profiles in that module [30]. Three modules were significantly correlated with VA status: the SI(Yellow), SI(Pink), and SI(Red) modules, with P values of 6×10^{-6}, 8×10^{-4} and .001, respectively.

Among these three modules, Yellow module is the largest (MS=924), with a correlation coefficient of −0.94 with VA status (Fig. 4C), thus, this module contains transcripts that were lower in the SI of VAS mice vs. their VAD counterparts (Fig. S5). According to ClusterProfiler, the SI(Yellow) module is functionally enriched in genes for adaptive immunity and cytokine activity (Table S5). The SI(Yellow) module contains known key players in RA metabolism (*Bco1*, *Scarb1*, *Gsta4*, *Aldh1a2*, and *Aldh1a3*), as well as potential RA targets (*Rdh1*, *Aldh3a2*, *Aldh7a1*, and *Aldh18a1*), along with several immune cell markers (*e.g.*, *Cd1d1*, *Cd28*, *Cd44*, *Cd86*, and *Cd207*).

The SI(Pink) module was positively correlated with VA status, however no 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 (Table S5). 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 Figure. 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 Fig. S6], with MS ranging from 87 to 5,799 genes (Table S6). Four modules, Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) differed significantly with VA status, with *P* values of 5×10^{-4}, .002, .004, and .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" (Table S6).

3.6. Associations with cell markers

To further query whether the expression pattern of each module was mainly driven by a regulatory effect of VA on cell differentiation, a cell type marker enrichment was performed. This test identified four SI modules that significantly overlap with the list of SI cell markers derived from the CellMarker database [35] (Fig. 6). The SI(Yellow) module, which as noted above is negatively correlated with VA (Fig. 4C), contains 24 Tuft cell markers, reaching a Bonferroni corrected *P* value (CP) of 4×10^{-7}. This suggested a negative regulation of VA on the Tuft cell population. The SI(Green) module also significantly overlapped with Tuft cell markers (CP=0.01). The SI(Turquoise) module, the largest module with MS=5,187 genes (Table S5), was significantly enriched for markers of goblet cells (CP = $1 \times 10^{-6)}$ and Paneth cells (CP=0.001).

3.7. Module preservation analysis suggests that gene networks controlled by VA are poorly preserved between SI and colon

To determine whether the modules identified as being significantly correlated with VA in SI have a similar module structure in the colon, we performed module preservation analysis using 14,059 genes that were robustly expressed in both datasets (Fig. 2). Network constructions were plotted for SI and colon separately. The modules from this analysis are referred as mpSI(Color) and mpColon(Color). Module preservation analysis was performed on mpSI and mpColon networks. For all three mpSI modules that were significantly correlated with VA status (Red, Pink, and Yellow), we found only "weak to moderate" evidence of preservation in the mpColon network according to Zsummary (Table S7). In addition, mpSI Red and Pink, two modules associated with VAS status (Fig. 4C), overlapped merely with the mpColon modules broken up into Black1 and Black2 (Fig. S7) instead of an intact mpColon(Black) module, which was originally also associated with VAS status (Fig. 4F). mpSI(Yellow), another SI module significantly correlated with VA status (Fig. 4C), had a lower preservation score than the mpSI (Red) and mpSI(Pink) modules (Table S7), and did not significantly overlap with any VA-related modules (Black, Green, Purple, or Red) in the mpColon network (Fig. S7).

3.8. RAR/RXR binding sites were predicted in the promoter regions of genes of interest

To provide *in silico* evidence that the genes of interest were potentially direct targets of RA, the promoter regions of *Scarb1*, *Bco1*, *Isx*, *Cd207*, *Mbl2*, *Clca1*, *Mmp9*, and *Nr0b2* were analyzed by CiiiDER software. Results related to RAR and/or RXR binding were visualized in Figure. 7B. All eight genes have multiple RAR/RXR binding sites predicted in their designated promoter regions (−1500 bases to +500 bases relative to the transcription start site).

4. Discussion

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

The approach we have taken in this study was to use rapidly frozen intact tissue to capture a “bird’s eye” comprehensive view of an entire organ, SI and colon individually, in physiologically normal mice subjected to only dietary treatment as a means to represent VA adequacy and deficiency in human populations. The use of intact tissue eliminated the selective loss or enrichment of cells and genes within, such as could occur during intestinal cell isolation protocols [34]. On the other hand, a limitation of our approach is that signals from genes significantly regulated by VA but residing within low-abundance cell types might have been overwhelmed and thereby missed. Additionally, the growth effect may constitute a confounding factor during the discovery interpretation, especially in the colon study, as the body weights were significantly lower in the VAD mice, although it was not shown in the SI study (Fig S1). We first found that, of the 49 DEGs differing by VA status in SI, there was a strong enrichment in “retinoid metabolic process” driven by three DEGs: *Cyp26b1* (Fold Change=14.1), *Isx* (Fold Change=3,231), and *Bco1* (Fold Change= −8.7) in VAS *verus* VAD mice (Table S1 and S2). RA is known to induce the transcription of CYP26B1, an enzyme that determines the magnitude of cellular exposure to RA by inactivating intracellular RA, via nuclear receptor protein binding to multiple retinoic acid response elements in the gene’s promoter region [42]; *Cyp26b1* has been shown to be dynamically regulated in rodent liver [43], lung [44], and SI [45]. RA also induces the expression of ISX [46], a transcriptional repressor for both BCO1, which catalyzes ß-carotene cleavage, and SCARB1, a receptor for carotene uptake, in the intestine [47]. Thus, the sum of the upregulation of *Cyp26b1* and *Isx*, as well as the downregulation of *Scarb1* and *Bco1* in the VAS SI, confirms the well-established negative feedback regulation of dietary VA on the intestinal RA concentration.

At the same time, a series of immunological genes previously reported to be regulated by VA, albeit in a variety of tissues, were also identified in the SI dataset, including *Mmp9*, *Cd207*, and enzymes produced by mast cells. The SI contains the body’s largest population of immune cells [48], with many cell types present, and thus finding genes in this category in our study was not unexpected. However, finding genes in this category that are sensitive to VA status may be informative for designing future studies. MMP9 is an extracellular matrix-degrading enzyme [49] that plays important immunological roles in macrophages and DCs [50], wherein *Mmp9* expression is induced by RA. In mouse myeloid DCs, chromatin immunoprecipitation assays showed that *Mmp9* expression was enhanced by RA through

a transcriptional mechanism involving greater RAR-alpha promoter binding and chromatin remodeling [51]. CD207, also known as langerin, is a marker of the langerin$^+$ DCs, a cell type normally only abundant in skin but not in the gut [52]. However, it has been reported that RA deficiency can change DC subtypes in the intestine and dramatically elevate langerin$^+$ DC numbers [53], which is in concordance with our SI study, where *Cd207* expression was significantly higher in the VAD group (Fig. 3A and Table S2). Mast cell proteases (MCPT) 1, 2, 4, and 6 (also known as TPSB2 or tryptase beta 2) were uniformly downregulated in the VAS group, an observation in line with previous reports that retinol and RA inhibit the growth and differentiation of human leukemic mast cells (HMC-1) [54], and the mast cells derived from peripheral blood or cord blood samples [55]. In addition, peritoneal mast cell numbers were also reported to be 1.5 times higher in VAD rats [56]. Overall, both predictable observations based on previous analysis of the SI, and observations in the SI that are congruent with the effects of VA and RA shown in other cell types, lend support to the strong impact that VA nutrition has on retinoid homeostasis and immune-related genes in the SI.

4.2. Potential new target genes regulated by VA in SI

Our analysis also uncovered several genes that were significantly regulated by VA status that may be targets for further study in the context of gut infections. One of these is *Mbl2*, an active player in gut immunity, the expression of which may be controlled by VA in SI (Fig. 3A). Mannose-binding lectin (MBL) is an important pattern-recognition receptor in the innate immune system [57], and MBLs are protective against gut infection by selectively recognizing and binding sugars presented on the pathogens [58]. 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* [59]. 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 [60]. However, interestingly, among infants with *Mbl2* variants, VA plus β-carotene supplementation partially counteracted the increased risk of transmission [61]. 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 Table S1), 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 [62]. 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* [63, 64]. In the hepatocyte cell line AML12, *Nr0b2* was upregulated dose-dependently by RA [63, 65]. However, in our SI study, *Nr0b2* was downregulated by VA (Fig. 3A and Table S2). 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* ranks relatively high regarding connectivity in our SI(Yellow) module (Fig. 5A), which is in line with its annotation as a TF.

4.3. RA treatment study and in silico prediction validated VA targets in SI

As an orthogonal verification, transcriptomes of lower SI of mice that were either VA deficient or VA deficient treated four times with an oral dose RA, the bioactive form of VA, were compared (Fig. 7A, Tables S8, S9, and S12). This validation dataset of oral RA supplementation focused on the acute transcriptional regulation by RA, which differs from the more accumulative state of VA deficiency that mimics the nutrient-deficient population living in resource-limited areas, a frequently studied topic in the field of nutritional science. With the stringent cutoff for DEG (adjusted $P<.05$ and |Fold change|> 2), the discoveries of the *Scarb1*, *Bco1*, *Isx*, *Cd207*, and *Mbl2* corresponding to the VA effect in our SI study were replicated (Fig. 7A, Tables S8 and S9). Furthermore, *Clca1*, *Mmp9*, *Nr0b2*, and *Cxcl14* were partially replicated because although they all had adjusted $P<.05$, their absolute fold changes were between 1.55 and 1.95 (Table S12), close to but not reaching the more stringent criteria of |Fold change|> 2, which we used in the analysis of our main study. The observed consistency between the genes discovered in the situation of acute RA treatment and with long-term VA status, taken together with the fact that RAR/RXR binding sites were predicted in the promoter regions of those genes (Fig. 7B), further suggests that those DEGs we have identified could be direct targets under the transcriptional regulation of RA in SI.

4.4. VA as a potential regulator of the SI Tuft cell population

An interesting observation that emerged from using the CellMarker database to identify cell type-selective genes in the SI dataset was the significance enrichment of Tuft cell markers in the SI(Yellow) module (Fig. 6). As this module is negatively correlated with VAS status, this enrichment suggests that the SI Tuft cell population may be more abundant under VAD status (Fig. 4C and Fig. S5). Tuft cells normally represent only about 0.5% of the epithelial cells in the murine small and large intestines, with a slightly more prevalent distribution in the distal SI [66]. As Tuft cells are a relatively newly identified secretory cell type [67], they have not yet been well studied. We found no literature on this cell type with respect to VA status. A speculation is that VAD status is associated with higher numbers of ILC2 cells, which stimulate Tuft cell differentiation. VA is known to regulate the balance between ILC2s and ILC3s in the intestinal mucosa [3]. VA deficiency has been shown to suppress ILC3 cell proliferation and function, and reciprocally to promote ILC2s [3]. IL-13 is one of the ILC2 cytokines that can drive the differentiation of epithelial precursors towards the Tuft lineage [66]. One of the advantages of the co-expression network analysis used in our study is that cell type-specific information may be derived through the analysis of whole intact tissues, without the isolation of homogeneous populations of cells [33, 34]. Given the small number of Tuft cells, isolation may prove difficult but approaches such as single-cell RNAseq, laser capture microdissection, and immunohistochemistry may be useful in the future. Although markers of goblet and Paneth cell were significantly enriched in SI(Turquoise) module, this module did not show significant correlation with VA status (Fig. 4C).

4.5. VA differentially regulates SI and colon transcriptomes

In general, DEGs in the SI were fewer in number but were condensed in more biological activities than those in the colon. The 49 DEGs in SI not only included cell markers (Table S2), but also genes in immunological and retinoid metabolic pathways (Fig. 3C and Fig. S4). Compared with the SI, the colon was enriched for fewer categories that differed by VA status. The 60 downregulated DEGs in colon under VAS condition were enriched for the GO term of cell division, which suggests that the primary transcriptional role of VA in colon is related to inhibition of cell proliferation and induction of epithelial differentiation, two widely applicable functions of RA. Together with the upregulation of goblet cell markers (*Tff3*, *Muc1*, *Clca1*, and *Clca4b*), our results are consistent with a differentiation-promoting effect of VA in the colon (Table S3). Nur et al. [13] reported vitamin A deficiency caused inflammatory changes in the rat colon that were similar to chemically induced colitis via DNA microarray. However, this was not replicated in the current colon RNAseq dataset, likely due to species and/or technical differences. Additionally, even though *Isx* was a DEG responsive to VA in both SI and colon, we did not find *Bco1* and *Scarb1* to be differentially expressed in the colon, in line with the processes of carotenoid digestion and absorption taking place primarily in SI. Other than *Isx* and *Cxcl14*, a potent chemoattractant for immune cells [68], no other genes in our studies were induced by VA in both organs.

Additionally, co-expression network analysis of the VA-correlated modules 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 (Table S7) and the lack of counterpart modules (Fig. S7). 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.

4.6. Limitations and future directions

Major limitations of the current work included: regarding the bulk RNAseq study design, relatively small sample sizes (n=3 in most groups), and whole tissue were used, leading to limited statistical power, and lack of direct evidence on the scale of specific cell types. The validation dataset focused on SI-only but lack the colon counterpart, and is potentially involved with a discrepancy between the effect of acute oral RA supplementation and the long-term accumulative effect of VA status. Finally, to strengthen the discoveries, functional verifications of immune players (*e.g. Mbl2*) under the regulation of VA, are highly desired.

In the future, larger sample sizes should be considered to increase power [69]. Single cell techniques may be employed to tease out the cell-type specific phenomena. For our validation study, we opted to study SI because when RA treatment is given orally, RA is readily absorbed in the SI and thus little RA reaches the colon; therefore, we focused on the region in which RA is absorbed. Validation of the discoveries in colon might require a different model system other than oral RA treatment (*e.g.*, transgenic mice that conditionally express dominant negative RAR in colon, or colonic instillation of RA). Furthermore, future

validation studies should also be expanded to the protein and functional levels. Our current work compared SI and colon on the scale of co-expression network. Alternative analysis focusing on the differential expression of individual genes between the two organs is also of future interest.

5. 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 appears to play an overall suppressive role on cell division while promoting cell differentiation, consistent with known functions of VA in other organ systems. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the sequencing service performed by C. Praul and colleagues in the Genomic Core Facility at the Pennsylvania State University.

Funding sources

This work was supported by National Institutes of Health research funding (R56 AI114972 and R01 GM109453 and 2 R01 HD066982) and training grant (T32GM108563), as well as J. Lloyd Huck Graduate Fellowship and Dorothy Foehr Huck Chair endowment provided by the Huck Institutes of Life Sciences. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation the manuscript.

Data and code availability

The RNAseq datasets generated for this study can be found in NCBI Gene Expression Omnibus (GEO) database with the accession number GSE143290. The R code can be found in Github (https://github.com/zhichai0120/RNAseqAnalysis.git).

References

[1]. Glasziou PP, Mackerras DE. Vitamin A supplementation in infectious diseases: a meta-analysis. BMJ 1993;306:366–70. [PubMed: 8461682]

[2]. West KP Jr. Extent of vitamin A deficiency among preschool children and women of reproductive age. J Nutr 2002;132:2857–66.

[3]. Sirisinha S. The pleiotropic role of vitamin A in regulating mucosal immunity. Asian Pac J Allergy Immunol 2015;33:71–89. [PubMed: 26141028]

[4]. Ross AC, Zolfaghari R, Weisz J. Vitamin A: recent advances in the biotransformation, transport, and metabolism of retinoids. Curr Opin Gastroenterol 2001;17:184–92. [PubMed: 11224677]

[5]. Cantorna MT, Nashold FE, Hayes CE. In vitamin A deficiency multiple mechanisms establish a regulatory T helper cell imbalance with excess Th1 and insufficient Th2 function. J Immunol 1994;152:1515–22. [PubMed: 8120366]

[6]. Iwata M, Hirakiyama A, Eshima Y, Kagechika H, Kato C, Song S-Y. Retinoic acid imprints gut-homing specificity on T cells. Immunity 2004;21:527–38. [PubMed: 15485630]

[7]. Mora JR, von Andrian UH. Role of retinoic acid in the imprinting of gut-homing IgA-secreting cells. Semin Immunol 2009;21:28–35. [PubMed: 18804386]

[8]. Choi JY, Cho KN, Yoon JH. All-trans retinoic acid induces mucociliary differentiation in a human cholesteatoma epithelial cell culture. Acta Otolaryngol 2004;124:30–5. [PubMed: 14977075]

[9]. Cha HR, Chang SY, Chang JH, Kim JO, Yang JY, Kim CH, et al. Downregulation of Th17 cells in the small intestine by disruption of gut flora in the absence of retinoic acid. J Immunol 2010;184:6799–806. [PubMed: 20488794]

[10]. Osanai M, Nishikiori N, Murata M, Chiba H, Kojima T, Sawada N. Cellular retinoic acid bioavailability determines epithelial integrity: role of retinoic acid receptor alpha agonists in colitis. Mol Pharmacol 2007;71:250–8. [PubMed: 17035595]

[11]. Collins JW, Keeney KM, Crepin VF, Rathinam VA, Fitzgerald KA, Finlay BB, et al. Citrobacter rodentium: infection, inflammation and the microbiota. Nat Rev Microbiol 2014;12:612–23. [PubMed: 25088150]

[12]. McDaniel KL, Restori KH, Dodds JW, Kennett MJ, Ross AC, Cantorna MT. Vitamin A-deficient hosts become nonsymptomatic reservoirs of Escherichia coli–like enteric infections. Infect Immun 2015;83:2984–91. [PubMed: 25964475]

[13]. Nur T, Peijnenburg AA, Noteborn HP, Baykus H, Reifen R. DNA microarray technology reveals similar gene expression patterns in rats with vitamin a deficiency and chemically induced colitis. J Nutr 2002;132:2131–6. [PubMed: 12163651]

[14]. Wiles S, Clare S, Harker J, Huett A, Young D, Dougan G, et al. Organ specificity, colonization and clearance dynamics in vivo following oral challenges with the murine pathogen Citrobacter rodentium. Cell Microbiol 2004;6:963–72. [PubMed: 15339271]

[15]. Blomhoff R, Green MH, Norum KR. Vitamin A: physiological and biochemical processing. Annu Rev Nutr 1992;12:37–57. [PubMed: 1503811]

[16]. Villablanca EJ, Wang S, de Calisto J, Gomes DC, Kane MA, Napoli JL, et al. MyD88 and retinoic acid signaling pathways interact to modulate gastrointestinal activities of dendritic cells. Gastroenterology 2011;141:176–85. [PubMed: 21596042]

[17]. Snyder LM, McDaniel KL, Tian Y, Wei C-H, Kennett MJ, Patterson AD, et al. Retinoic acid mediated clearance of Citrobacter rodentium in vitamin A deficient mice requires CD11b + and T Cells. Front Immunol 2019;9:3090. [PubMed: 30671060]

[18]. Carman J, Smith S, Hayes C. Characterization of a helper T lymphocyte defect in vitamin A-deficient mice. J Immunol 1989;142:388–93. [PubMed: 2463304]

[19]. Smith SM, Levy NS, Hayes CE. Impaired immunity in vitamin A-deficient mice. J Nutr 1987;117:857–65. [PubMed: 3495650]

[20]. Hodges JK, Tan L, Green MH, Ross AC. Vitamin A supplementation transiently increases retinol concentrations in extrahepatic organs of neonatal rats raised under vitamin A–marginal conditions. J Nutr 2016;146:1953–60. [PubMed: 27534819]

[21]. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. [PubMed: 24695404]

[22]. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60. [PubMed: 25751142]

[23]. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2. [PubMed: 20110278]

[24]. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6. [PubMed: 21221095]

[25]. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2013;30:923–30. [PubMed: 24227677]

[26]. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550–71. [PubMed: 25516281]

[27]. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559–72. [PubMed: 19114008]

[28]. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4:1–25.

[29]. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2007;24:719–20. [PubMed: 18024473]

[30]. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 2007;1:54. [PubMed: 18031580]

[31]. Yu G, Wang L, Han Y, He Q. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7. [PubMed: 22455463]

[32]. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 2010;107:12698–703. [PubMed: 20616000]

[33]. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. Nat Neurosci 2008;11:1271–82. [PubMed: 18849986]

[34]. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 2011;474:380–4. [PubMed: 21614001]

[35]. Shi A, Deng C, Zhao E, Quan F, Li F, Liao G, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 2018;47:721–8.

[36]. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLOS Comput Biol 2011;7:e31001057.

[37]. Hu H, Miao Y-R, Jia L-H, Yu Q-Y, Zhang Q, Guo A-Y. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res 2018;47:33–8.

[38]. Gearing LJ, Cumming HE, Chapman R, Finkel AM, Woodhouse IB, Luu K, et al. CiiiDER: A tool for predicting and analysing transcription factor binding sites. PLOS One 2019;14:e0215495. [PubMed: 31483836]

[39]. Balmer JE, Blomhoff R. Gene expression regulation by retinoic acid. J Lipid Res 2002;43:1773–808. [PubMed: 12401878]

[40]. Wiede A, Jagla W, Welte T, Kohnlein T, Busk H, Hoffmann W. Localization of TFF3, a new mucus-associated peptide of the human respiratory tract. Am J Respir Crit Care Med 1999;159:1330–5. [PubMed: 10194185]

[41]. Nyström EE, Birchenough GM, van der Post S, Arike L, Gruber AD, Hansson GC, et al. Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity. EBioMedicine 2018;33:134–43. [PubMed: 29885864]

[42]. Zhang Y, Zolfaghari R, Ross AC. Multiple retinoic acid response elements co-operate to enhance the inducibility of CYP26A1 gene expression in liver. Gene 2010;464:32–43. [PubMed: 20682464]

[43]. Ross AC. Retinoid production and catabolism: role of diet in regulating retinol esterification and retinoic acid oxidation. J Nutr 2003;133:291–6.

[44]. Wu L, Ross AC. Acidic retinoids synergize with vitamin A to enhance retinol uptake and STRA6, LRAT, and CYP26B1 expression in neonatal lung. J Lipid Res 2010;51:378–87. [PubMed: 19700416]

[45]. Wang Y, Zolfaghari R, Ross AC. Cloning of rat cytochrome P450RAI (CYP26) cDNA and regulation of its gene expression by all-trans-retinoic acid in vivo. Arch Biochem Biophys 2002;401:235–43. [PubMed: 12054474]

[46]. Lobo GP, Amengual J, Baus D, Shivdasani RA, Taylor D, von Lintig J. Genetics and diet regulate vitamin A production via the homeobox transcription factor ISX. J Biol Chem 2013;288:9017–27. [PubMed: 23393141]

[47]. Lobo GP, Hessel S, Eichinger A, Noy N, Moise AR, Wyss A, et al. ISX is a retinoic acid-sensitive gatekeeper that controls intestinal β, β-carotene absorption and vitamin A production. FASEB J 2010;24:1656–66. [PubMed: 20061533]

[48]. Couter CJ, Surana NK. Isolation and flow cytometric characterization of murine small intestinal lymphocytes. JoVE 2016;111:111–16.

[49]. Hijova E. Matrix metalloproteinases: their biological functions and clinical implications. Bratisl Lek Listy 2005;106:127–32. [PubMed: 16026148]

[50]. MacLauchlan S, Skokos EA, Meznarich N, Zhu DH, Raoof S, Shipley JM, et al. Macrophage fusion, giant cell formation, and the foreign body response require matrix metalloproteinase 9. J Leukoc Biol 2009;85:617–26. [PubMed: 19141565]

[51]. Lackey DE, Hoag KA. Vitamin A upregulates matrix metalloproteinase-9 activity by murine myeloid dendritic cells through a nonclassical transcriptional mechanism. J Nutr 2010;140:1502–8. [PubMed: 20534877]

[52]. Merad M, Ginhoux F, Collin M. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. Nat Rev Immunol 2008;8:935–47. [PubMed: 19029989]

[53]. Chang SY, Cha HR, Chang JH, Ko HJ, Yang H, Malissen B, et al. Lack of retinoic acid leads to increased langerin-expressing dendritic cells in gut-associated lymphoid tissues. Gastroenterology 2010;138:1468–78. [PubMed: 19914251]

[54]. Alexandrakis MG, Kyriakou DS, Seretakis D, Boucher W, Letourneau R, Kempuraj D, et al. Inhibitory effect of retinoic acid on proliferation, maturation and tryptase level in human leukemic mast cells (HMC-1). Int J Immunopathol Pharmacol 2003;16:43–7. [PubMed: 12578730]

[55]. Ishida S, Kinoshita T, Sugawara N, Yamashita T, Koike K. Serum inhibitors for human mast cell growth: possible role of retinol. Allergy 2003;58:1044–52. [PubMed: 14510724]

[56]. Wiedermann U, Chen XJ, Enerbäck L, Hanson L, Kahu H, Dahlgren U. Vitamin A deficiency increases inflammatory responses. Scand J Immunol 1996;44:578–84. [PubMed: 8972739]

[57]. Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz R. Phylogenetic perspectives in innate immunity. Science 1999;284:1313–18. [PubMed: 10334979]

[58]. Neth O, Jack DL, Dodds AW, Holzel H, Klein NJ, Turner MW. Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. Infect Immun 2000;68:688–93. [PubMed: 10639434]

[59]. Shi L, Takahashi K, Dundee J, Shahroor-Karni S, Thiel S, Jensenius JC, et al. Mannose-binding lectin-deficient mice are susceptible to infection with Staphylococcus aureus. J Exp Med 2004;199:1379–90. [PubMed: 15148336]

[60]. Turner MW. Mannose-binding lectin (MBL) in health and disease. Immunobiology 1998;199:327–39. [PubMed: 9777416]

[61]. Coutsoudis A, Trabattoni D, Archary D, Segat L, Clerici M, Crovella S, 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 2006;84:610–15. [PubMed: 16960176]

[62]. Lu TT, Makishima M, Repa JJ, Schoonjans K, Kerr TA, Auwerx J, et al. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. Mol Cell 2000;6:507–15. [PubMed: 11030331]

[63]. Mamoon A, Ventura-Holman T, Maher JF, Subauste JS. Retinoic acid responsive genes in the murine hepatocyte cell line AML 12. Gene 2008;408:95–103. [PubMed: 18006250]

[64]. Maguire M, Bushkofsky JR, Larsen MC, Foong YH, Tanumihardjo SA, Jefcoate CR. Diet-dependent retinoid effects on liver gene expression include stellate and inflammation markers and parallel effects of the nuclear repressor Shp. J Nutr Biochem 2017;47:63–74. [PubMed: 28570941]

[65]. Mamoon A, Subauste A, Subauste MC, Subauste J. Retinoic acid regulates several genes in bile acid and lipid metabolism via upregulation of small heterodimer partner in hepatocytes. Gene 2014;550:165–70. [PubMed: 25014134]

[66]. Banerjee A, McKinley ET, von Moltke J, Coffey RJ, Lau KS. Interpreting heterogeneity in intestinal Tuft cell structure and function. J Clin Invest 2018;128:1711–19. [PubMed: 29714721]

[67]. Gerbe F, van Es JH, Makrini L, Brulin B, Mellitzer G, Robine S, et al. Distinct ATOH1 and Neurog3 requirements define Tuft cells as a new secretory cell type in the intestinal epithelium. J Cell Biol 2011;192:767–80. [PubMed: 21383077]

[68]. Hara T, Tanegashima K. Pleiotropic functions of the CXC-type chemokine CXCL14 in mammals. J Biochem 2012;151:469–76. [PubMed: 22437940]

[69]. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA (New York, NY) 2016;22:839–51.

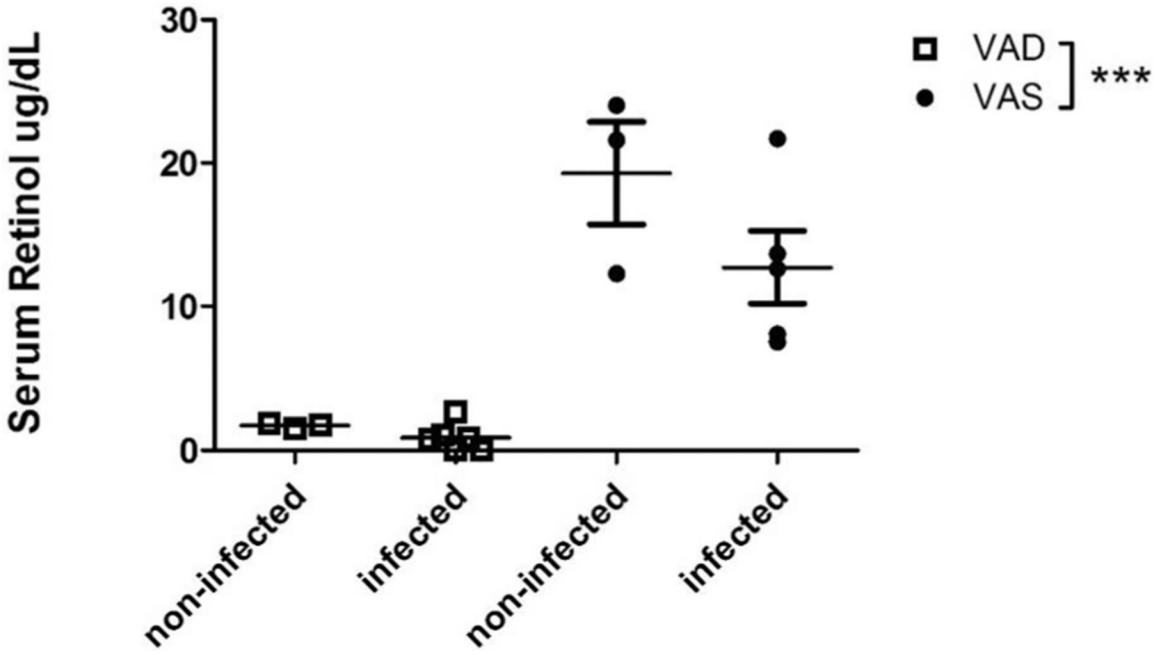

Fig. 1.

VAD mice have lower serum retinol levels than VAS mice. Serum retinol concentrations were measured by UPLC on post-infection day 5. Values are the mean ± SEM of n=3–6 mice per group. Two-way ANOVA test. *** $P<.001$. **Abbreviations:** vitamin A (VA), vitamin A deficient (VAD), vitamin A sufficient (VAS), Ultra Performance Liquid Chromatography (UPLC), standard error of the mean (SEM), analysis of variance (ANOVA).

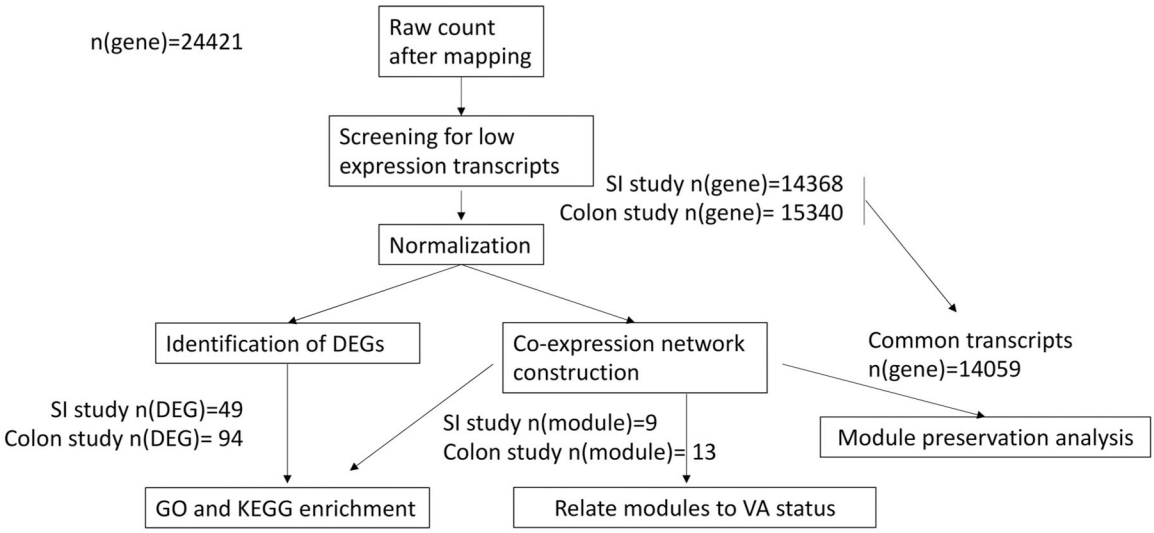

Fig. 2.

Overview of the results of bioinformatics analyses. Initial mapping identified 24,421 genes. After data pre-processing, 14,368 genes in the SI study and 15,340 genes in the colon study were maintained in the two datasets. Starting from the 14,368 SI-expressed genes, 49 DEGs corresponded to VA effect, and 9 co-expression modules were identified. Similarly, for the colon study, 94 DEGs and 13 modules were identified among the 15,340 colon-expressed genes in the colon study. A total of 14,059 genes were shared between the two datasets and were used to for the module preservation analysis. **Abbreviations:** differentially expressed gene (DEG), small intestine (SI), vitamin A (VA).

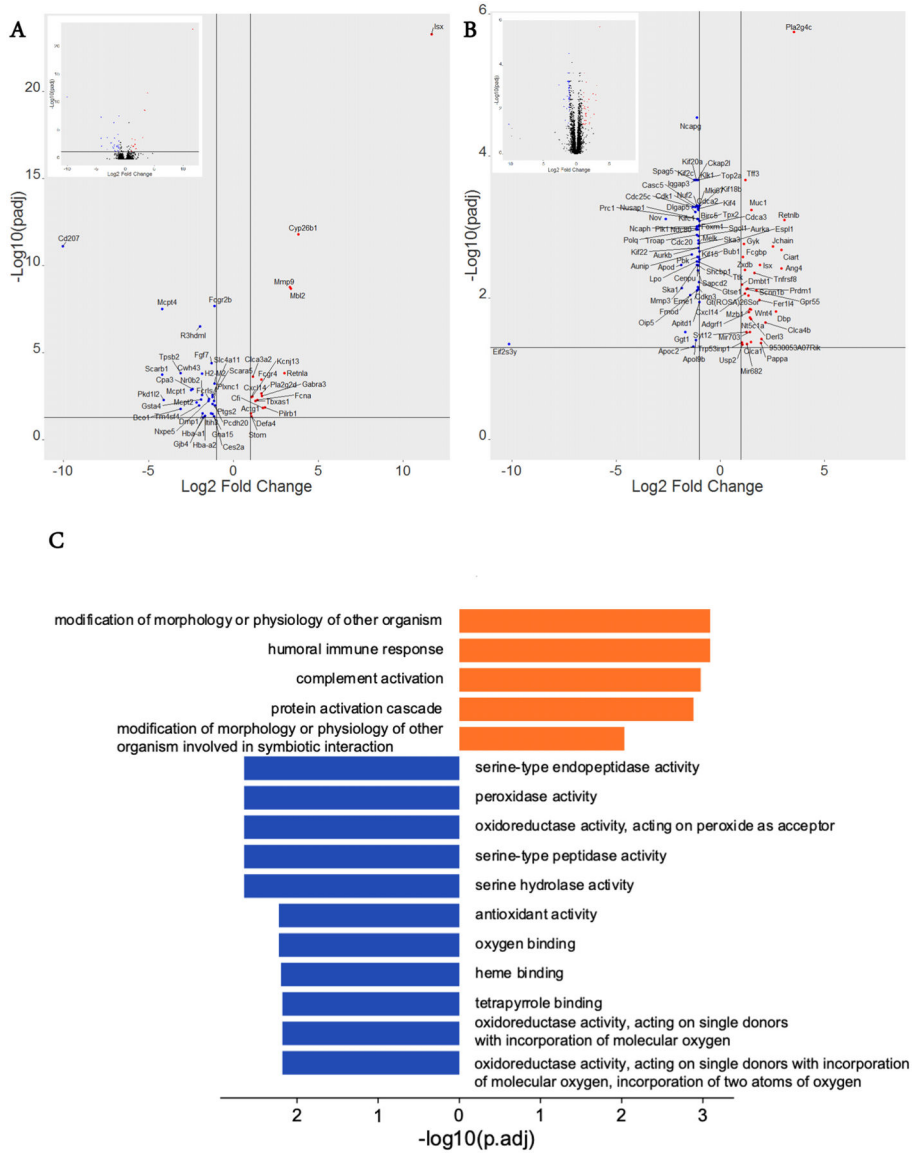

Fig. 3.

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

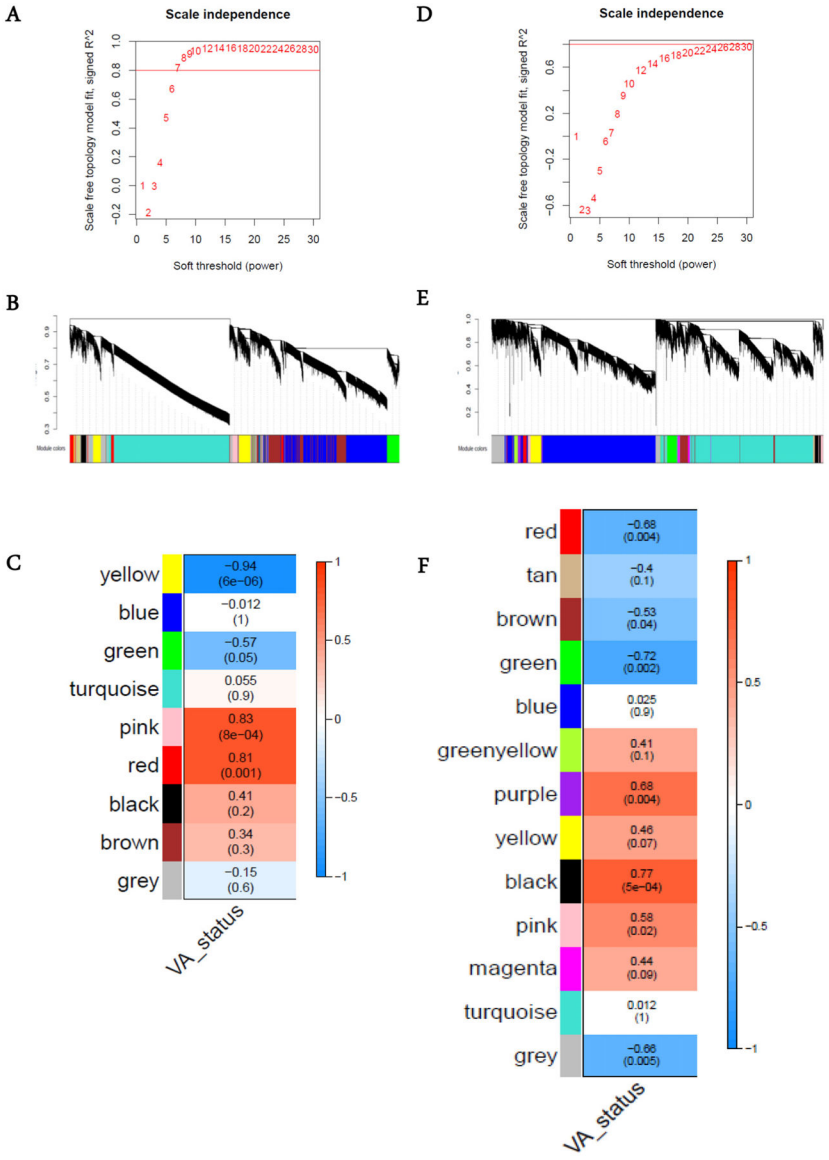

Fig. 4.

Co-expression network construction in the intestines via WGCNA. **A**, **B, C**, Results from SI study (12 samples involved). **D, E, F**, Results from colon study (16 samples involved). **A, D**, Soft threshold power beta were chosen for SI (**A**) and colon (**D**) networks according to the scale-free topology criteria (horizontal red lines). **B, E**, Clustering dendrograms of co-expression modules in SI (**B**) and colon (**E**). Transcripts are represented as vertical lines and are arranged in branches according to topological overlap similarity using the hierarchical clustering. The dynamic treecut algorithm was used to automatically detect a series of highly connected modules, referred to by a color designation. Module color designation are shown below the dendrogram. **C, F**, Correlation of module eigengenes with VA status in SI (**C**) and colon (**F**) networks. Upper numbers in each cell: Correlation coefficient between the module and the trait. Lower numbers in parentheses: *P* value of the correlation. Significant correlations ($P<0.01$) were obtained between VA status and the SI(Yellow), SI(Pink),

SI(Red) (**C**), Colon(Black), Colon(Green), Colon(Purple), and Colon(Red) (**F**) modules.
Abbreviation: weighted gene co-expression network analysis (WGCNA), small intestine (SI), vitamin A (VA).

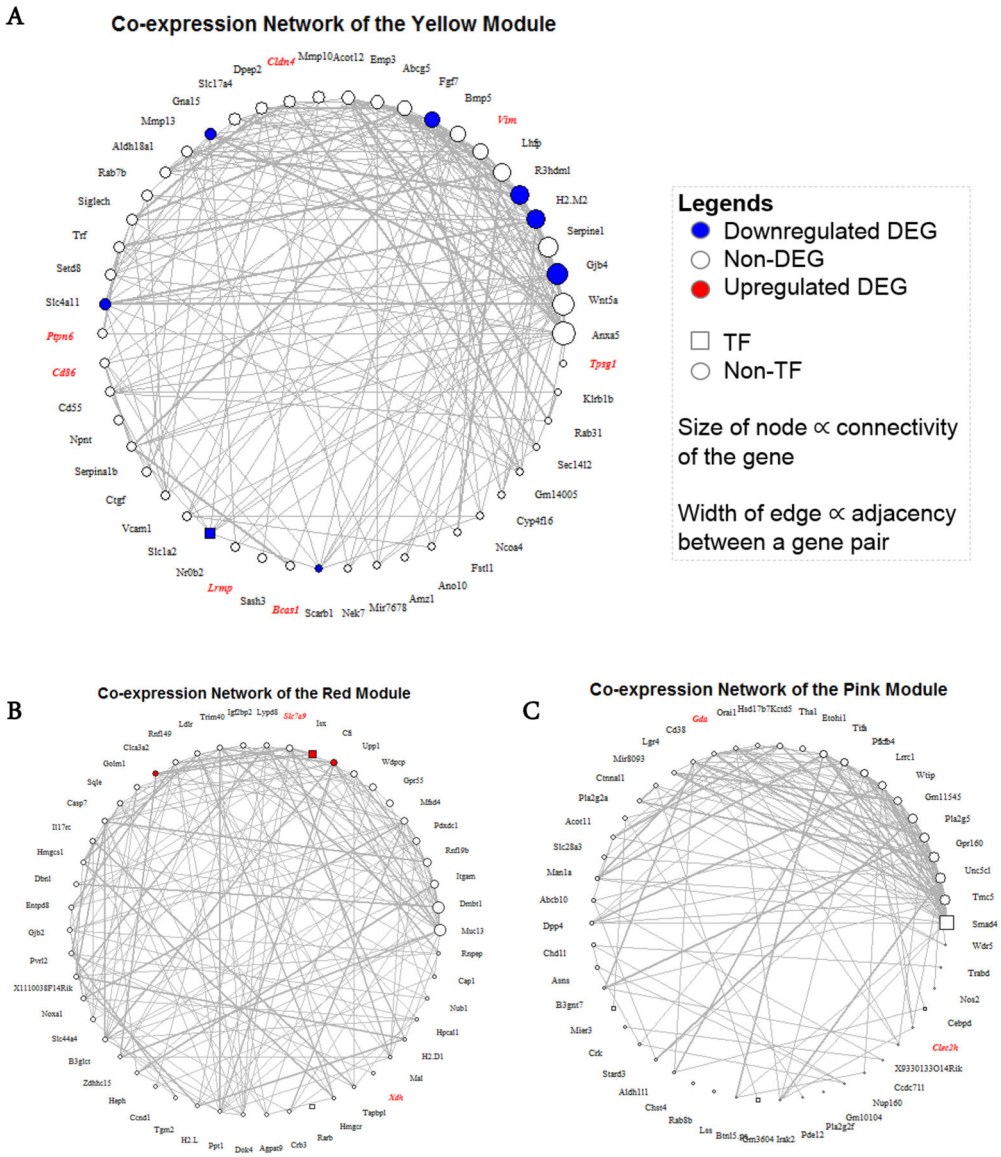

Fig. 5.

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

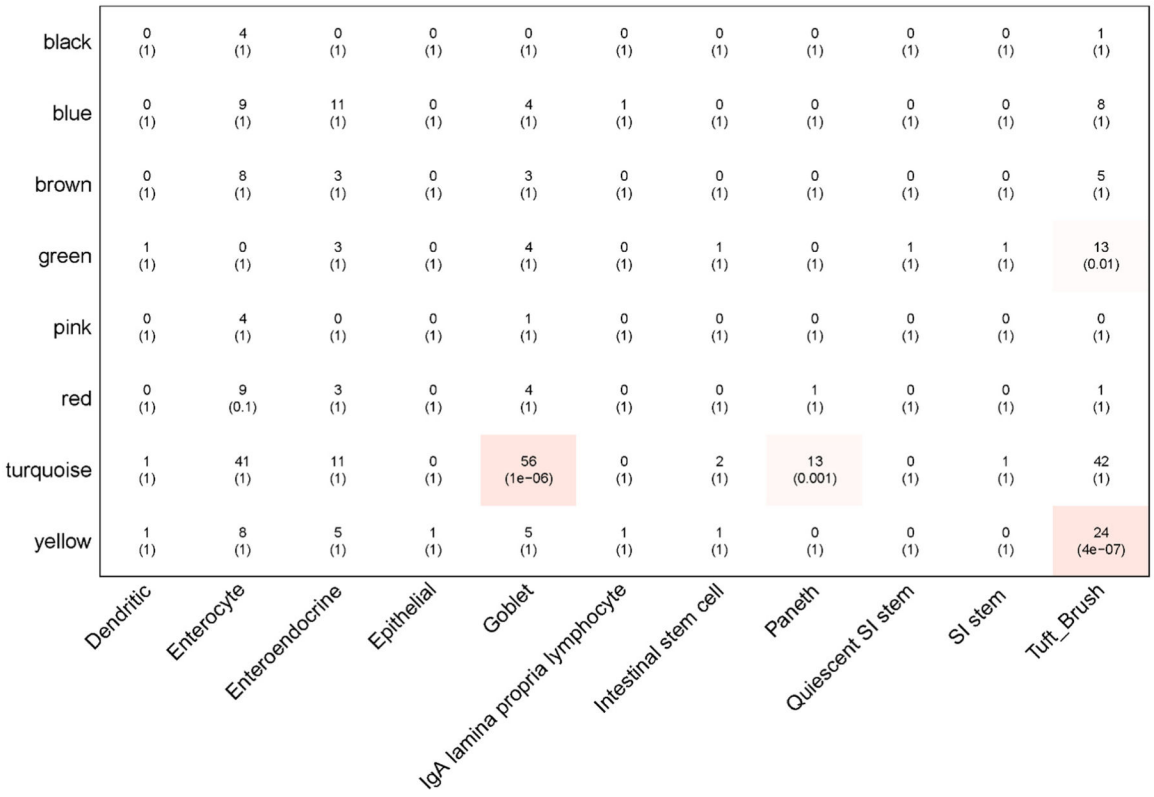

Fig. 6.

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

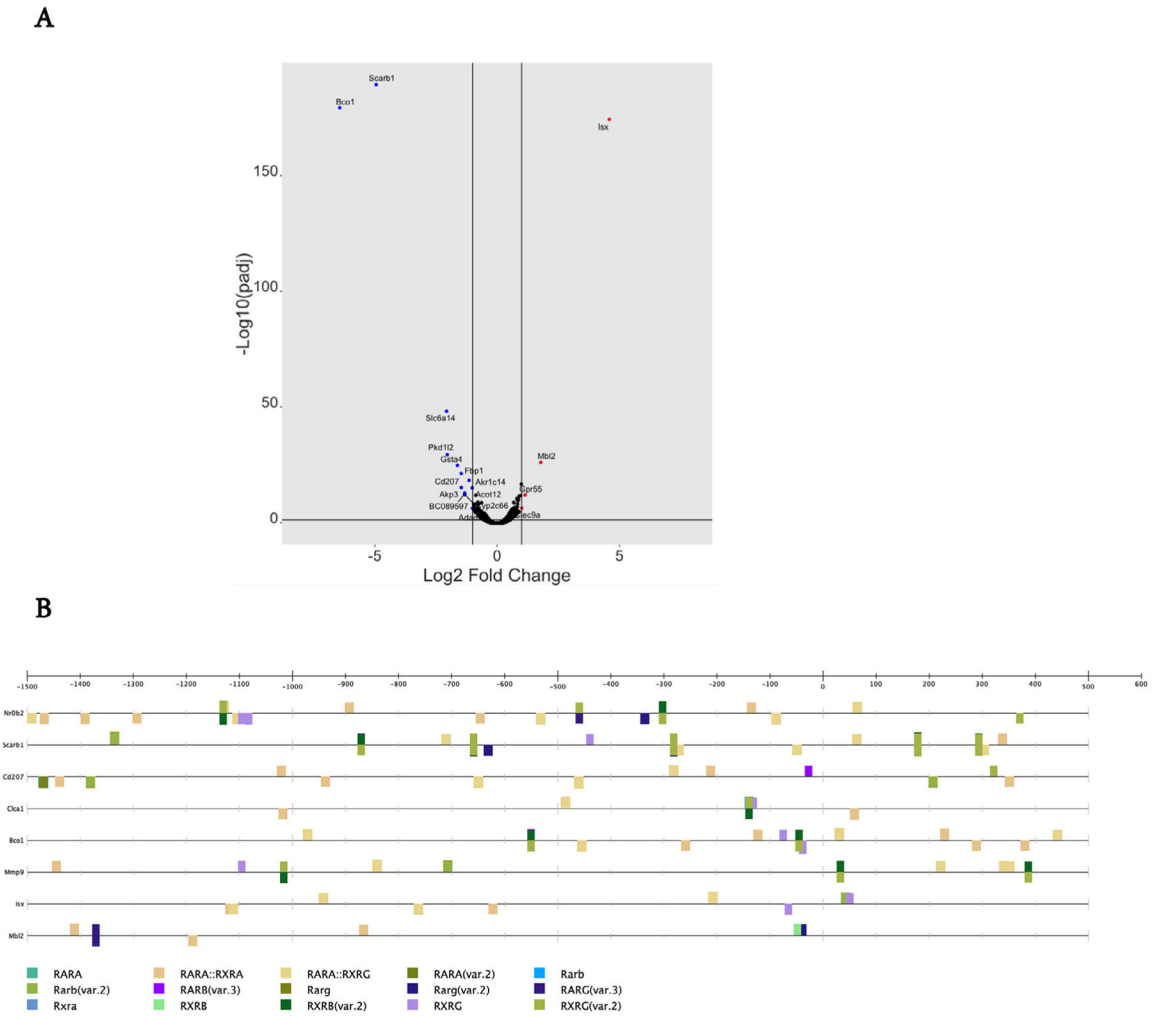

Fig. 7.

Experimental and bioinformatic validation study results. **A**, Volcano plot of DEGs in the validation dataset of SI. Red: DEGs upregulated in VAD mice orally treated with four times RA compared with VAD mice without oral RA supplementation; Blue: DEGs downregulated in VAD mice orally treated with four times RA compared with VAD mice without oral RA supplementation; Black: non-DEGs. Each experimental group contained three mice. Criteria: |Fold Change|=2 is marked out by two black vertical lines x=1 and x=−1. padj=0.05 is marked out by a horizontal black line y=−Log10(0.05). **B.** Visualization of RAR/RXR binding sites predicted in the promoter regions of *Nr0b2*, *Scarb1*, *Cd207*, *Clca1*, *Bco1*, *Mmp9*, *Isx*, and *Mbl2*. Promoter regions were defined as spanning −1500 bases to +500 bases relative to the transcription start site. CiiiDER software was employed, using Ensembl 94 annotations of the mouse GRCm38 genome and 2020 JASPAR core vertebrate matrices, with deficit threshold set at 0.25. Binding sites of other transcription factors were computed but not shown here. **Abbreviations:** padj (adjusted *P* value according to DESeq2), differentially expressed gene (DEG), small intestine (SI), vitamin A deficient (VAD), retinoic acid (RA), retinoic acid receptors (RAR), retinoid X receptor (RXR).