Transcriptome variation in response to Marek's disease virus early infection

Likit Preeyanon[1] C. Titus Brown[1,2] Hans H. Cheng[3,*]

1 Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA

2 Computer Science and Engineering, Michigan State University, East Lansing, MI, USA

3 USDA, ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, USA

∗ E-mail: Corresponding hans.cheng@ars.usda.gov

Abstract

Marek's disease (MD) is caused by highly oncogenic Marek's disease virus (MDV). Different MHC alleles have been associated with susceptibility and resistance to the disease. However, chickens line 6 and 7 show significant phenotypic differences when chanllenged with MDV even with the same MHC allele. Therefore, the major unanswered question is what genetic factors underlie the susceptibility and resistance of line 6 and 7. In this study, we identified differentially expressed genes and isoforms in chicken line 6 and 7 in response to Marek's disease infection. Results from pathway analysis and functional analysis show that changes in expression of genes and isoforms may contribute to resistance and susceptibility of the disease. We also identified many single nucleotide polymorphisms (SNPs) between line 6 and 7 that may cause alteration of alternative splicing, which in turn results in allele specific expression.

Author Summary

Introduction

Marek's disease is an economically significant chicken disease that affects a poultry industry worldwide. Total of $2 billion loss has been reported from outbreaks. The disease is caused by highly oncogenic Marek's disease virus (MDV) which causes T-cell lymphoma. Vaccination is effective in reducing incidence of tumor formation; however, current MD vaccines do not prevent infection or horizontal spread of the virus. Improper use of vaccines is a key factor that drives evolution of highly virulent strains in vaccinated flocks.

Need to add intro about breed selection against infection

Many studies have reported strong associations between MHC alleles and resistance and susceptibility of the disease. For example, chickens with MHC allele B^{21} is highly resistant in contrast to chickens with B^{16} allele. However, chickens line 6 and 7 , even with common allele (B^6), exhibit different phenotypic response to MD. The major unanswered question is what genetic factors contribute to susceptibility and resistance of the disease.

Need to add intro about gene and isoforms in immune system

In this study, we identified several annotated and unannotated genes and isoforms that are differentially expressed in response to MD infection. Pathway analysis and gene ontology suggest that these genes may have important roles in immune response to MD.

Results

Custom Gene Annotations

A large number of reads mapped to intergenic and intronic regions (table 1, supp.) regions of Ensembl annotations (Galgal4.72) indicates that many genes and isoforms are not included in the annotations.

To extensively study gene and isoform expression, we employed *de novo* transcriptome assembly and a reference-guided assembly to build custom gene annotations (see methods). The number of genes and isoforms is summarized in Table 1. Genes were identified by aligning against non-redundant protein database from NCBI. We found several genes that not annotated in either RefSeq or Ensembl, including chicken specific genes as well as genes found in other organisms (Table **??**).

Differential Gene Expression

Many previous studies have reported many differentially expressed (DE) genes between line 6 and line 7 from Microarray and RNA-Seq experiments from different stages of infection. Some known differentially expressed genes were identified in this study. For example, B6.1 (Bu-1) gene is known to be down-regulated approximately 2.5 fold in susceptible chicken with MHC allele B^{19} []. It was also found to be down-regulated about three fold in line 7.

Even though both lines shared a significant number of genes that were differentially expressed, approximately, twice as many genes in line 7 as in line 6 were differentially expressed (Table 2). Interestingly, some genes that were differentially expressed in both lines were regulated in opposite directions. As shown in table 3, LYG2, SFTPA1, LL and SERPINB10 were up-regulated in line 7 but down-regulated in line 6. In contrast, TAF1D and LOC424145 were up-regulated in line 6 but down-regulated in line 7. LYG2, for instance, was up-regulated by 2.57 folds in line 7 but down-regulated by 1.7 folds in line 6. This gene is involved in a defense against bacteria.

Gene network analysis

Pathway analysis shows that genes in line 7 are significantly enriched ($p < 0.05$) in pathways involved in immune system such as Toll-like receptor signaling, intestinal immune network for IgA production, and Phagosome pathway, which are not significantly enriched in line 6 (fig. 1). Figure 3 and 4 shows a diagram of phagosome and intestinal immune network for IgA production repectively. Genes from line 6 and line 7 are highlighted with yellow and red color. Common genes are highlighted with orange color. These two pathways show that both line 6 and line 7 regulated MHC class I gene in response to infection, but only line 7 regulated MHC class II gene. It has been shown that Marek's disease virus (MDV), unlike other herpes viruses, up-regulates MHC class II gene during infection. Moreover, regulation is believed to be mediated directly by MDV infection [1]. However, from these results, it appears that only MHC class II gene in line 7 was up-regulated.

Furthermore, differentially expressed genes in line 7 are enriched in biological processes involved in both adaptive and innate immune responses, whereas differentially expressed genes in line 6 are only enriched in innate immune responses (fig. 2). This might correspond to the fact that MHC class II gene was upregulated in line 7. High level of MHC class II molecules may help recruit more T cells, which in turn enhance activation of the adaptive immune response.

Differential Exon Usage

Functional Analysis

Polymorphisms in differentially expressed genes/exons

Discussion

Materials and Methods

0.1 Gene Models Construction

Due to lack of complete gene models for chickens, we employed two methods to construct gene models from RNA-Seq reads. First, short reads were quality trimmed with conditri/2.1 [] `(-cutfirst 10)` and assembled using Velvet/1.21 [] and Oases/0.2.06 [] to obtain long transcripts. Assembly was done with k value ranges from 21 to 31 for both local and global assembly (described in Gimme paper []). Poly-A tails were trimmed and low complexity transcripts were removed by Seqclean []. All transcripts were then aligned to chicken reference genome (galGal4, with unplaced and random chromosomes removed) with BLAT [] `(-t=dna -q=dna -noHead -out=psl -mask=lower -extendThroughN)`. Filtered alignments from BLAT were then used to produced gene models using Gimme [].

Second, short reads were aligned to reference genome using Tophat/2.0.5 [] and gene models were built by Cufflinks2 [] with default parameters. A combination of gene models from both set of gene models were used in this study.

0.2 Differential Gene Expression and Gene Ontology

To identify DE genes, we obtained read counts using multiBamCov command from BEDTools/12.13.1 [] for all datasets. Then DE gene were identified by DESeq/1.10.1 [2] from Bioconductor. Data from single- and paired-end datasets from the same line were treated as biological replicates. To identify enriched pathways and ontology terms, a list of DE genes was analysed by GOSeq/1.10.0 []. P-values were corrected by Benjamini-Hochberg multiple testing correction. Genes, patways and GO terms with corrected P-value < 0.05 were considered significant.

0.3 Differential Splicing Event

Gene models were converted to alternative splicing models using a Python script. In order to increase sensitivity, read counts from single- and paired-end samples were combined and treated as single-end reads for splicing event analysis with MISO/0.4.9 [3]. Splicing events with Bayes factor > 10 and $\Delta\Psi > 0.20$ were considered significant. Read coverages and Ψ distributions were plotted using Sashimi plot [4].

0.4 Variant Calling and *In Silico* Splicing Analysis

Variants were called using mpileup command from SAMTools/0.1.18 [] and BCFTools []. Only variants with quality score ≥ 20 were used for mutation analyses. Exon enhancers and suppressors were predicted using Human Splicing Enhancer web portal []. Human default parameter settings were used in all analyses. The following regulatory sequences were used to determined the effect of variants: HSF integrated matrices for serine/arginine-rich proteins (SRp40, SC35, SF2/ASF, SF2/ASF, IgM/BRCA1, and SRp55), exonic splicing enhancer (ESE), RESCUE-ESE hexamer (RESCUE-ESE), putative 8-mer ESEs (PESEs) and putative 8-mers exonic splicing silencers (PESSs), exon-identity and intron-identiry elements (EIEs and IIEs), hetronuclear ribonucleoprotein-binding motifs, and Fas exonic splicing silencers.

Acknowledgments

References

1. Niikura M, Kim T, Hunt HD, Burnside J, Morgan RW, et al. (2007) Marek's disease virus up-regulates major histocompatibility complex class II cell surface expression in infected cells. Virology 359: 212–219.

2. S A, W H (2010) Differential expression analysis for sequence count data. Genome Biology 11: R106–R106.

3. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature Publishing Group 7: 1009–1015.

4. Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, et al. (2013) Sashimi plots: Quantitative visualization of RNA sequencing read alignments. Audio and Electroacoustics Newsletter, IEEE : –.

Figure Legends

Tables

Table 1. Gene models summary

Method	Gene	Isoform
Assembly	25,290	54,044
Cufflinks	21,345	36,218
Combined	24,980	46,613

Number of genes and isoforms from gene models built from *de novo* assembly and Cufflinks. Combination of both methods decreased the number of genes and isoforms by merging fragmented transcripts to form more complete gene models.

Table 2. Differentially expressed genes

Sample	Up-regulated	Down-regulated	Total	Refseq	Ensembl
Line 6	1,043	595	1,188	984 (60.0%)	1,252 (76.4%)
Line 7	1,976	1,124	3,100	2,114 (68.2%)	2,574 (83.0%)
Both	681	236	917	664 (72.4)	771 (84.1%)

FDR < 0.1

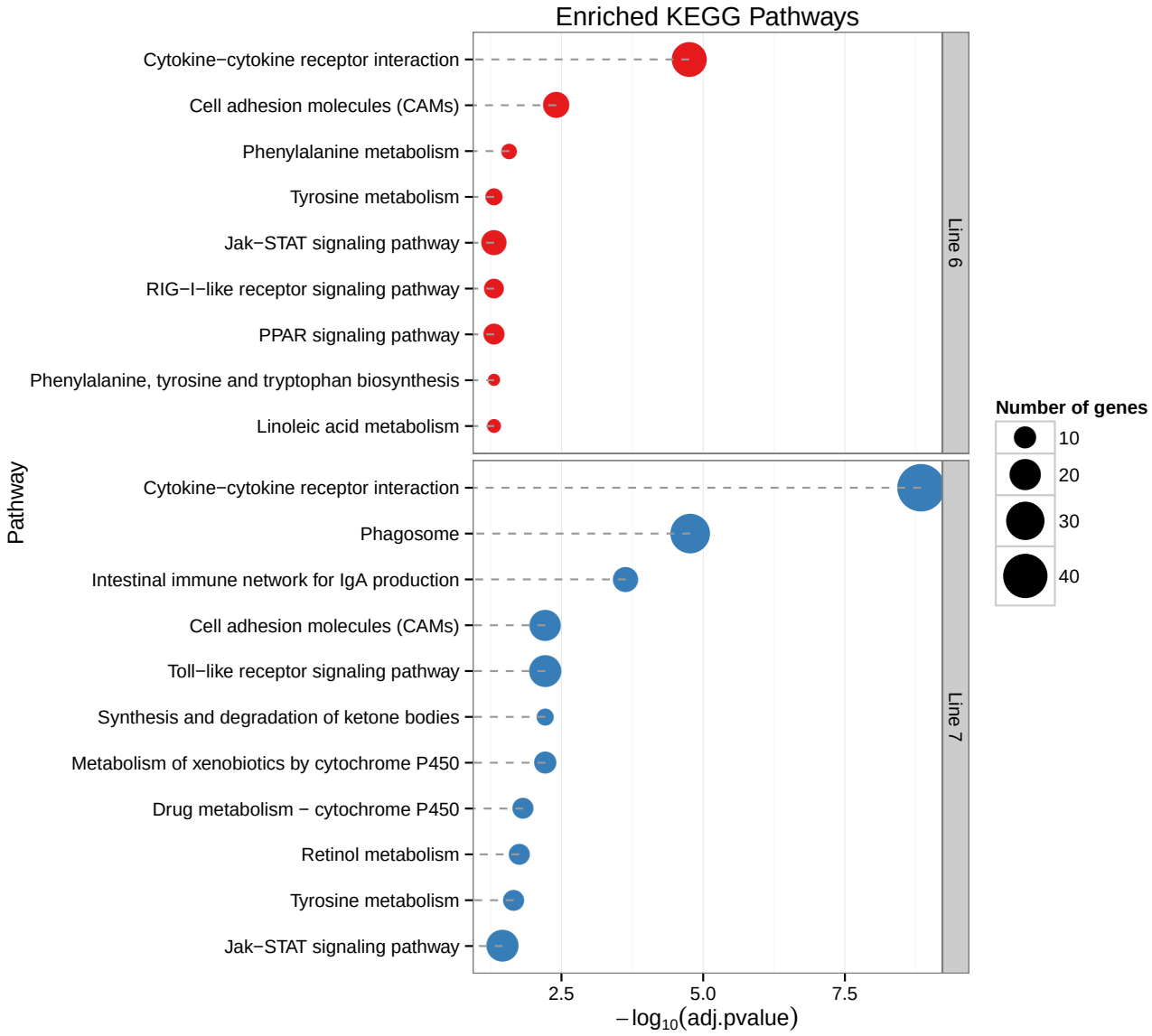

Figure 1. Enriched KEGG pathways. Rest of figure 2 caption. Caption should be left justified, as specified by the options to the caption package.

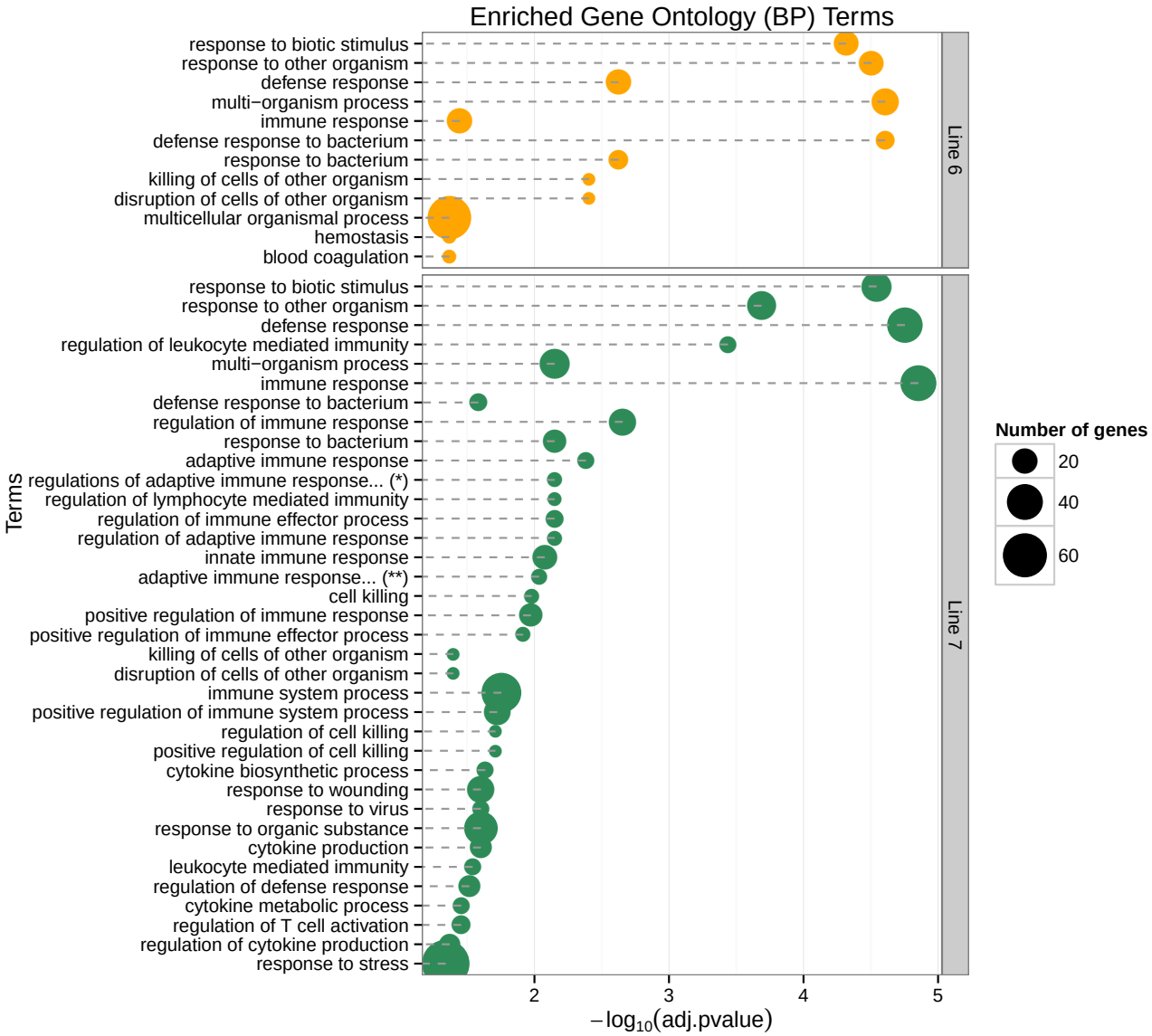

Figure 2. Enriched Gene Ontology (BP) Terms. Rest of figure 2 caption. Caption should be left justified, as specified by the options to the caption package.

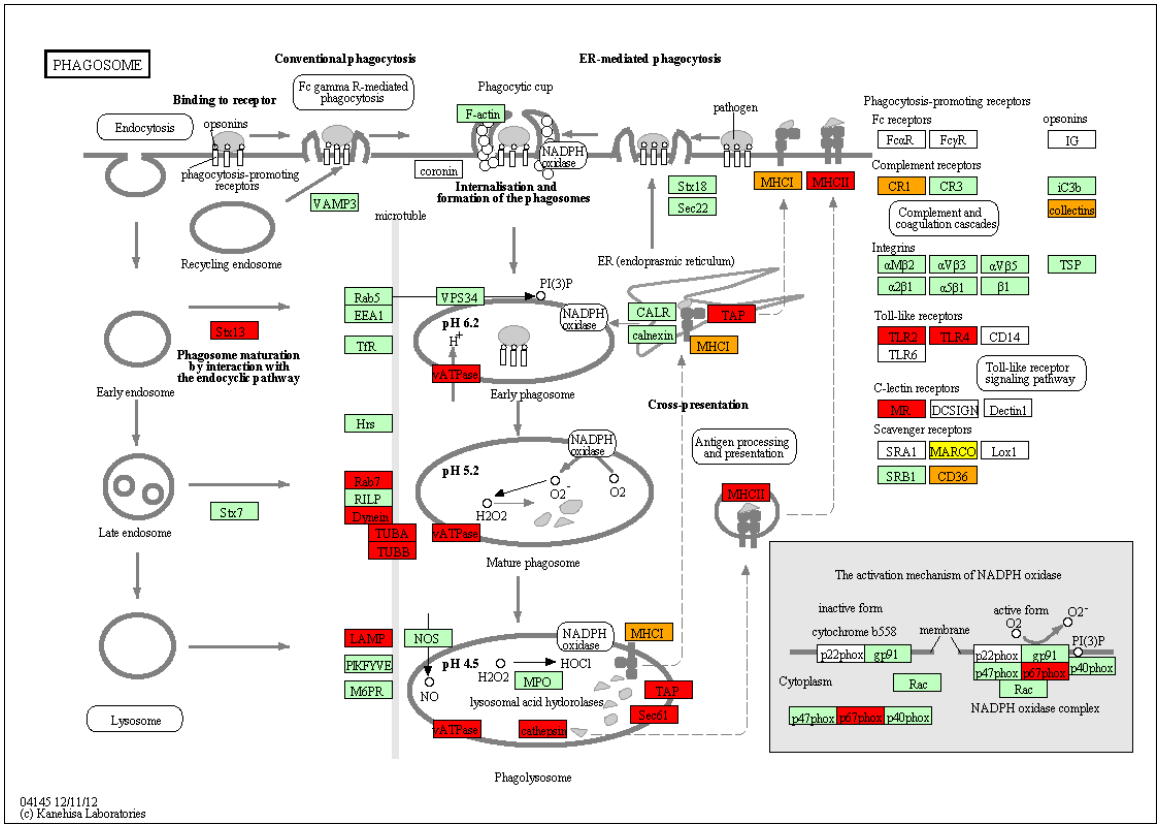

Figure 3. Enriched KEGG pathways in line 7. Rest of figure 2 caption. Caption should be left justified, as specified by the options to the caption package.

Table 3. Genes regulated in opposite direction

Gene	Entrez ID	Fold Change	
		Line 6	line 7
TAF1D	419002	+1.08	-1.13
LYG2	395708	-1.70	+2.57
SFTPA1	395308	-5.12	+3.51
LOC424145	424145	+1.57	-1.68
LL	423630	-3.30	+8.31
SERPINB10	395715	-1.12	+0.87

(-) down-regulated, (+) up-regulated

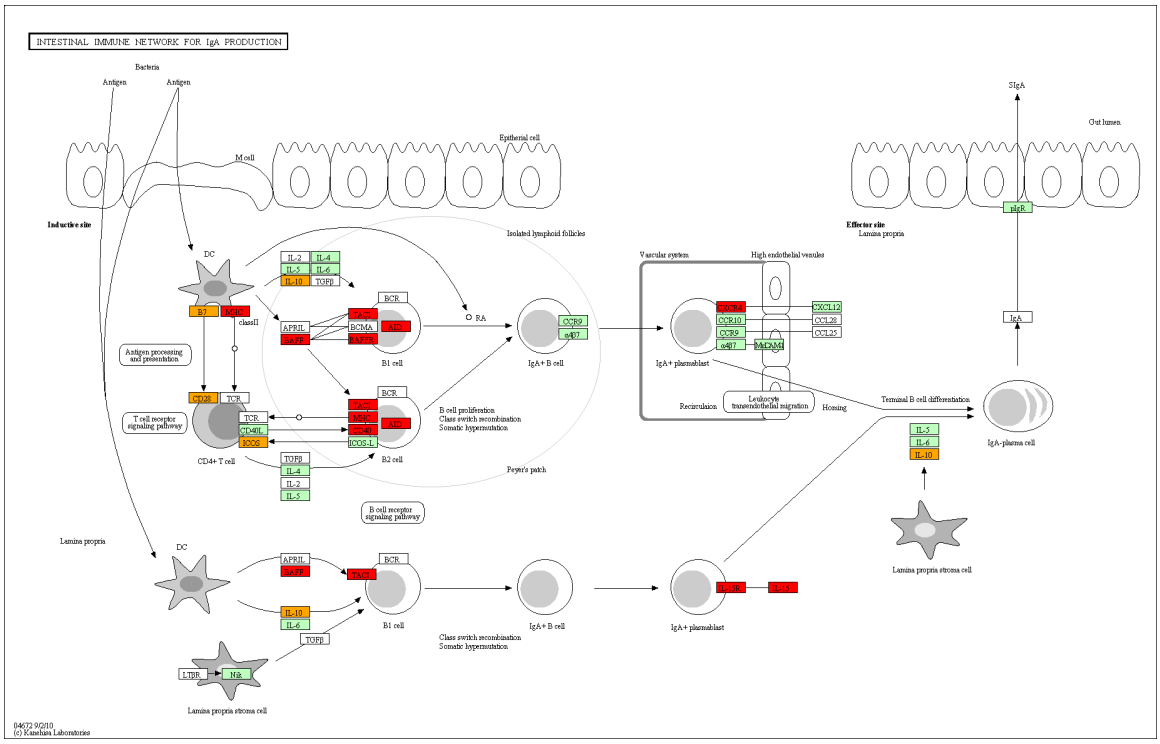

Figure 4. Intestinal network for IgA production pathway. Rest of figure 2 caption. Caption should be left justified, as specified by the options to the caption package.