Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A study on linking of clinical phenotype and gene expression profile in systemic lupus erythematosus

**Jian-Ruei Ciou**
China Medical University
**Pu-Wei Ho**
China Medical University Hospital
**Po-Chang Wu**
China Medical University Hospital
**Shu-I Chen**
Asia University Hospital
**Ching-Mao Chang**
Taipei Veterans General Hospital
**Hen-Hong Chang**
China Medical University
**Hsueh-Ting Chu** ( prof.htchu@gmail.com )
Asia University https://orcid.org/0000-0003-0618-5598

**Research article**

**Keywords:** Gene expression profile, systemic lupus erythematosus, malar rash, DAAM2, clinical phenotype

**Posted Date:** March 9th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-16323/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

# Abstract

Objectives:

Malar rash is one of clinical phenotypes seen in systemic lupus erythematosus (SLE). However, the pathogenesis of malar rash is not clear for each case of SLE patients. In this paper we endeavored to investigate the linking of clinical phenotype from the gene expression profiles between both patients with malar rash and without malar rash. Therefore we might perform better evaluation of the possible prognosis for different SLE patients in the future.

Methods: This study utilizes transcriptome sequencing (RNA-Seq) technologies to discover underlying gene expression profile for systemic lupus erythematosus patients. We performed transcriptome sequencing experiments and analyzed differentially expressed genes (DEGs) and associated pathways.

Results: From the analysis of gene expression profiling, we identified the gene DAAM2 is the most differentially expressed gene for patients with malar rash. Using a gene set enrichment analysis, we discuss the linkage between DAAM2 and the possible pathways for systemic lupus erythematosus with malar rash.

Conclusions: We identified DAAM2 as a candidate biomarker for the clinical phenotype of malar rash for systemic lupus erythematosus.


# Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease with multiple organs and systems damaging, including the kidneys, skin, cardiovascular system, central and peripheral nervous systems, and blood. SLE patients also have immunologic abnormalities, particularly the production of a number of antinuclear antibodies and anti-double stranded DNA antibody (dsDNA). It has previously been reported that ~ 7.4-159.4/100,000 people suffer from SLE worldwide, and women more affected than men, especially in childbearing years [1]. With complex interaction of immune dysfunction, genetic predisposition, and environmental factors, the pathogenesis of SLE isn't completely understood.

Genetic susceptibility was showed in SLE patients. Major histocompatibility complex (MHC), HLA-DR2 and HLA-DR3 alleles and homozygous C4a deficiency are associated with high risk of SLE. In present gene study, the microarray analysis of lncRNA target prediction indicated the presence of 474 matched lncRNA-mRNA pairs for 293 lncRNAs and 381 with differentially expressed (fold change, ≥ 3.0). However, no research show gene expression between SLE patterns and subtypes previously. So we want to investigate different gene expression in SLE patients to evaluate SLE subtypes for further diagnosis and treatment.

Next-generation sequencing (NGS) provides a powerful tool for identifying novel targets of epigenetics and their regulation pathway [2]. Using NGS platforms, RNA-Seq can detect more features of both genomic and transcriptomic targets such as single nucleotide variants, gene expression, transcript isoforms, gene fusions, etc. It was reported that 8,868 lncRNA and 6,876 mRNAs were highly differently expressed in SLE patients. In this study, we studied a RNA-Seq dataset from Gene Expression Omnibus (GEO) database. We investigated the signaling pathways and analyzed the differentially expressed genes (DEGs) for the dataset.

The Mucocutaneous lesions occurs in more than 80% patients. The most typical SLE pattern was malar rash, that "butterfly" rash over cheeks and bridge of nose. Frequency of clinical manifestations in prospective cohort of 1,000 patients with SLE showed arthritis in 48.1% patients, malar rash in 31.1% active nephropathy in 27.9%, neurologic involvement in 19.4%, Raynaud phenomenon in 16.3% patients. It was few report discussing the relationship between SLE pattern and RNA expression. In this study, we studied a RNA-Seq dataset from Gene Expression Omnibus (GEO) database and RNA expression for SLE patients. We investigated the signaling pathways and analyzed the differentially expressed genes (DEGs) for the dataset.

## Materials And Methods

# Patients recruitment

The peripheral blood samples were collected from a clinical trial of SLE patient to investigate Traditional Chinese medicine constitution Questionnaire and RNA-sEq. The clinical trial recruited patients who were diagnosed with SLE at the Rheumatology clinic in Chinese Medical University Hospital, Taiwan. RNA samples purified from 14 female patients were sequenced in this study in the period from 2016. All patients were checked autoantibodies and complement at baseline for immunologic response, including antinuclear antibody (ANA), anti-double stranded DNA antibody (dsDNA), anti-Ro antibody, anti-La antibody, C3, and C4.

# Ethics Statement

This study was approved by the Institutional Review Board and Ethics committee of Chinese Medical University Hospital. All the study participants gave informed consent and signed a partnership agreement.

# Rna-seq

The RNA-Seq samples were processed using Illumina TruSeq Stranded mRNA Library Preparation Kit. The contained mRNA molecules were enriched by poly(A) capture methods. Finally, the products were purified and enriched with PCR to create the cDNA library. The quality of the library is then assessed by the Bioanalyzer (Agilent). Pair-ended maximal 300-bp reads were generated by Illumina MiSeq sequencers.

# The Pipeline Of The Differentially Expressed Gene (deg) Analysis

Figure 1 shows the entire pipeline of the differentially expressed gene (DEG) analysis. In the first step, the raw reads in FASTQ format were aligned to the Ensembl[3] human genome GRCh38.84 using the STAR program[4] (version 2.5.2a). Then, the aligned reads of different genes were counted by the HTSeq program[5] (version 0.6.1). Finally, the differential expression (DE) of genes was estimated by DESeq2 (version 1.12.4)[6]. The results of DE analysis are listed with values of log2 fold changes and p values.

# Pathway Analysis

We performed the pathway analysis with R/Bioconductor packages including GAGE[7] and Pathview[8]. Generally Applicable Gene-set Enrichment (GAGE) method was used to perform a large variety of different enrichment group tests. GAGE uses the log2 fold changes for differentially expressed genes (DEGs) inferred by DeSeq2 and identifies KEGG pathways that are significantly enriched by a set of DEGs [9]. The package Pathview was used to visualize the changes on the pathway diagram from KEGG [8].

# Statistical analysis

The all statistical analysis was performed using R-Studio with different R packages, including DESeq2, IHW and etc. The DESeq2 uses shrinkage estimation for dispersions and fold changes to improve stability and interpretability of estimates for differential analysis of count data[6]. The method of independent hypothesis weighting (IHW) assigns weights using covariates independent of the P-values under the null hypothesis but informative of each test's power or prior probability of the null hypothesis [10].

## Results

# The demographics of study population

Table 1. demonstrated 14 patients include in this study. Six patients got malar rash, and other 8 patients don't have malar rash during disease course. The mean ± standard deviation (SD) of age in the patients with malar rash was 52.3 ± 11.2 years. And the mean ± standard deviation (SD) of age in the patients without malar rash were 43.0 ± 16.2. years.

Table 1 Demographics of study population

| Patient No. | Gender | Age |
|---|---|---|
| Malar rash case | | |
| 001 | F | 37 |
| 005 | F | 67 |
| 007 | F | 58 |
| 014 | F | 45 |
| 015 | F | 47 |
| 016 | F | 60 |
| Non- Malar rash case | | |
| 002 | F | 26 |
| 004 | F | 54 |
| 006 | M | 60 |
| 008 | F | 36 |
| 010 | F | 57 |
| 011 | F | 26 |
| 012 | F | 25 |
| 013 | F | 60 |

# Discovery Of Differentially Expressed Genes (degs)

Out of 31739 genes with nonzero total read count, there are 273 up-regulated DEGs (LogFoldChange > 0 and p-value < 0.1) and 472 down-regulated DEGs (LogFoldChange < 0 and p-value < 0.1).

Wmape list a set of 26 top up-regulated genes (Table 2) with significantly differential expression levels (log fold change > 1.5, p-value < 0.05).

Table 2 Top up-regulated genes for the Malar rash case (Log fold change ≥ 1.0, p-value < 0.05)

**Table 3. Enriched KEGG pathways from the gene-set enrichment analysis.**

| STAT | | | | Hiseq | | | |
|---|---|---|---|---|---|---|---|
| | baseMean | log2FoldChange | pvalue | | baseMean | log2FoldChange | pvalue |
| DAAM2 | 29.06676 | -1.94744 | 1.24E-09 | DAAM2 | 28.10381 | -1.99311 | 1.04E-09 |
| MAOA | 18.92453 | -1.47075 | 2.75E-06 | MAOA | 17.99152 | -1.47732 | 3.62E-06 |
| TPST1 | 87.71595 | -1.36307 | 1.41E-05 | TPST1 | 85.71138 | -1.37629 | 1.65E-05 |
| ZBTB16 | 61.11633 | -1.31519 | 3.75E-05 | ZBTB16 | 58.47011 | -1.33231 | 4.04E-05 |
| LRRC70 | 5.119989 | -1.19713 | 0.000141 | LRRC70 | 4.950254 | -1.22389 | 0.00015 |
| KCNH2 | 49.99896 | -1.1087 | 0.000467 | ESRG | 3.46006 | -1.21889 | 0.000104 |
| FLT3 | 100.1457 | -1.10808 | 0.000532 | AC012085.1 | 8.77048 | -1.16054 | 0.000348 |
| ESRG | 3.627885 | -1.03267 | 0.000597 | KCNH2 | 49.30827 | -1.1576 | 0.000329 |
| IL1R2 | 1674.154 | -1.03048 | 0.000959 | PTENP1 | 15.15322 | -1.14348 | 0.000338 |
| UBB | 4825.111 | -1.01323 | 0.001186 | LDHAP4 | 5.568518 | -1.13765 | 0.000378 |
| | | | | FLT3 | 99.66973 | -1.1312 | 0.000514 |
| | | | | AC116050.1 | 3.35729 | -1.06894 | 0.000554 |
| | | | | AC026877.1 | 3.394568 | -1.06276 | 0.001182 |
| | | | | UBB | 4328.675 | -1.05978 | 0.000868 |
| | | | | IL1R2 | 1620.838 | -1.05332 | 0.001004 |
| | | | | SFRP2 | 15.68336 | -1.04755 | 0.001306 |
| | | | | AL590609.2 | 10.02961 | -1.03551 | 0.0016 |
| | | | | IFITM9P | 4.27899 | -1.0301 | 0.001501 |
| | | | | ABCC13 | 96.07135 | -1.02802 | 0.001727 |
| | | | | CA1 | 1356.708 | -1.02659 | 0.001245 |
| | | | | AL109741.2 | 2.579441 | -1.02157 | 0.001633 |
| | | | | CAP1P2 | 11.50526 | -1.0076 | 0.001961 |
| | | | | | | | |
| TRAV27 | 10.64419 | 1.015294 | 0.001447 | CXCL10 | 36.66244 | 1.02917 | 0.001659 |
| CXCL10 | 37.16574 | 1.02493 | 0.001408 | KLRC4-KLRK1 | 6.186152 | 1.031459 | 0.001619 |
| KLRC4-KLRK1 | 6.77916 | 1.028773 | 0.001342 | AC124319.2 | 15.32082 | 1.060265 | 0.001233 |
| CTD-2047H16.4 | 15.88216 | 1.03887 | 0.00123 | TRAV27 | 9.564415 | 1.069475 | 0.001006 |

| KEGG entry | Pathway name | p-value (<0.05) |
|---|---|---|
| hsa03008 | Ribosome biogenesis in eukaryotes | 0.020306334 |
| hsa04612 | Antigen processing and presentation | 0.010615261 |
| hsa04622 | RIG-I-like receptor signaling pathway | 0.020244272 |
| hsa04623 | Cytosolic DNA-sensing pathway | 0.009087804 |
| hsa04672 | Intestinal immune network for IgA production | 0.040610364 |

# Differences In Performance Of Rna Under Stat And Htseq

Table 2. shows the difference in specific gene expression in SLE patients with malar rash. There are 26 genes of RNA that surpass twice the differential performance in patients without malar rash (Stat analysis 14 genes, HTseq analysis 26genes). 22 genes were lower than those of asymptomatic patients, and 4 genes were higher than those of asymptomatic patients. The DAAM2 gene, which has the largest difference in performance, was nearly four times lower than the asymptomatic group.

# The related KEGG pathways from gene set enrichment analysis of DEGs

From GAGE pathway analysis, we listed the top 5 statistically significant enriched pathways with FDR adjusted p value < 0.05 for pathways of common increased genes (Table 3). The pathways of the Ribosome biogenesis in eukaryotes, Antigen processing and presentation, RIG-I-like receptor signaling pathway, Cytosolic DNA-sensing pathway, and Intestinal immune network for IgA production were mapped for the sign of malar rash on SLE patients. Figures 2 to Fig. 6 depict the rendered graphs of KEGG pathways by Pathview. The red nodes in all of the graphs are up-regulated genes.

DAAM2 has been reported to be associated with glioma or psychosis. On the KEGG diagram, related to the WNT path. Figure 7 depict the DAAM2 on KEGG pathway.

## Discussion

The skin damage is the most common organ affected in patients with SLE, but the mechanisms involved in the pathogenesis of skin lesions and the formation of SLE skin manifestations are still unclear. Ultraviolet (UV), immune cells, cytokines and immunoglobulin deposition may play an important role in the development of skin inflammation and damage in SLE. In this study, we discover some genes differential expression in patients with malar rash. It could be a method study differences in phenotypes and genotypes with RNA-seq.

## Abbreviations

DE, differential expression; DEG, differentially expressed gene; FC, fold change; LFC, log fold change; OPD, outpatient department; ED, emergency department; PCR, Polymerase chain reaction; IHW, independent hypothesis weighting

## Declarations

### Ethics approval and consent to participate

This study was approved by the Institutional Review Board and Ethics committee of China Medical University Hospital. All the study participants gave informed consent and signed a partnership agreement.

### Consent for publication

Not applicable.

### Availability of data and materials

The results and data sets used in this study are availale at: https://github.com/htchu/DAAM2.

### Competing interests

The authors declare that they have no competing interests.


### Funding

This work was financially supported by China Medical University Hospital (DMR-107-166, DMR-108-8, ASIA-107-CMUH-12, ASIA-108-CMUH-22).


### Authors' Contributions

JRC and PWH did the analysis on the patient data. JRC, PWH, BJW, SIC, CMC, HHC and HTC discussed the project and jointly wrote the manuscript. HTC and HHC leaded the project. All authors read and approved the final manuscript.


### Corresponding Author

Correspondence to: Hsueh-Ting Chu htchu.taiwan@gmail.com, Hen-Hong Chang tcmchh55@gmail.com



### Acknowledgments

This work was financially supported by China Medical University Hospital (DMR-107-166, DMR-108-8, ASIA-107-CMUH-12, ASIA-108-CMUH-22).


## References


1. Mosca M, Tani C, Aringer M, Bombardieri S, Boumpas D, Brey R, Cervera R, Doria A, Jayne D, Khamashta MA *et al*: **European League Against Rheumatism recommendations for monitoring patients with systemic lupus erythematosus in clinical practice and in observational studies**. *Annals of the rheumatic diseases* 2010, **69**(7):1269-1274.
2. Shyr D, Liu Q: **Next generation sequencing in cancer research and clinical application**. *Biological Procedures Online* 2013, **15**:4-4.
3. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, Garcia Giron C, Hourlier T *et al*: **The Ensembl gene annotation system**. *Database : the journal of biological databases and curation* 2016, **2016**.
4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics (Oxford, England)* 2013, **29**(1):15-21.
5. Anders S, Pyl PT, Huber W: **HTSeq–a Python framework to work with high-throughput sequencing data**. *Bioinformatics (Oxford, England)* 2015, **31**(2):166-169.
6. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biology* 2014, **15**(12):550.
7. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ: **GAGE: generally applicable gene set enrichment for pathway analysis**. *BMC bioinformatics* 2009, **10**:161.
8. Luo W, Brouwer C: **Pathview: an R/Bioconductor package for pathway-based data integration and visualization**. *Bioinformatics* 2013, **29**(14):1830-1831.
9. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M: **New approach for understanding genome variations in KEGG**. *Nucleic Acids Res* 2019, **47**(D1):D590-d595.

10. Ignatiadis N, Klaus B, Zaugg JB, Huber W: **Data-driven hypothesis weighting increases detection power in genome-scale multiple testing**. 2016, **13**(7):577-580.

## Figures

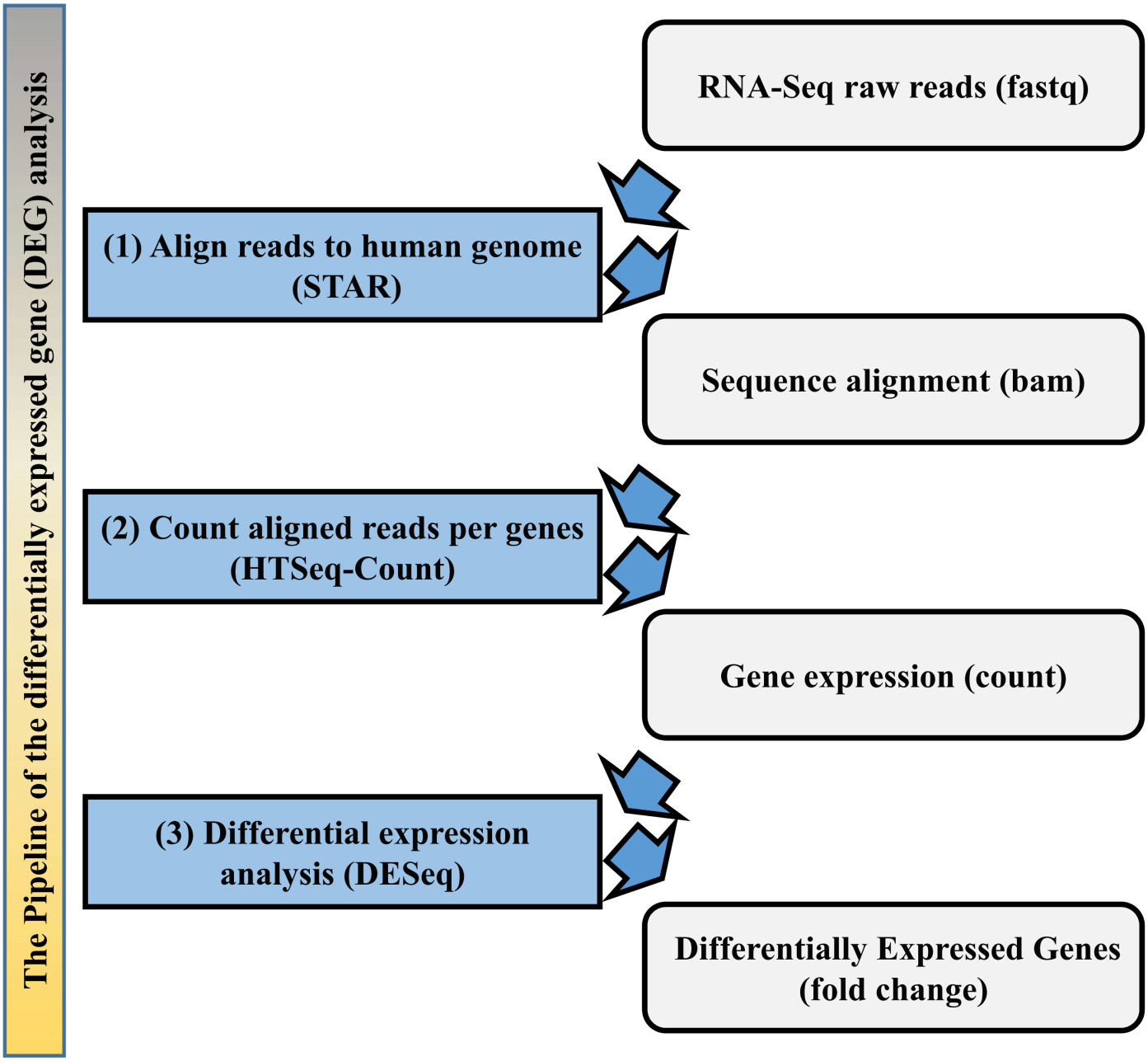
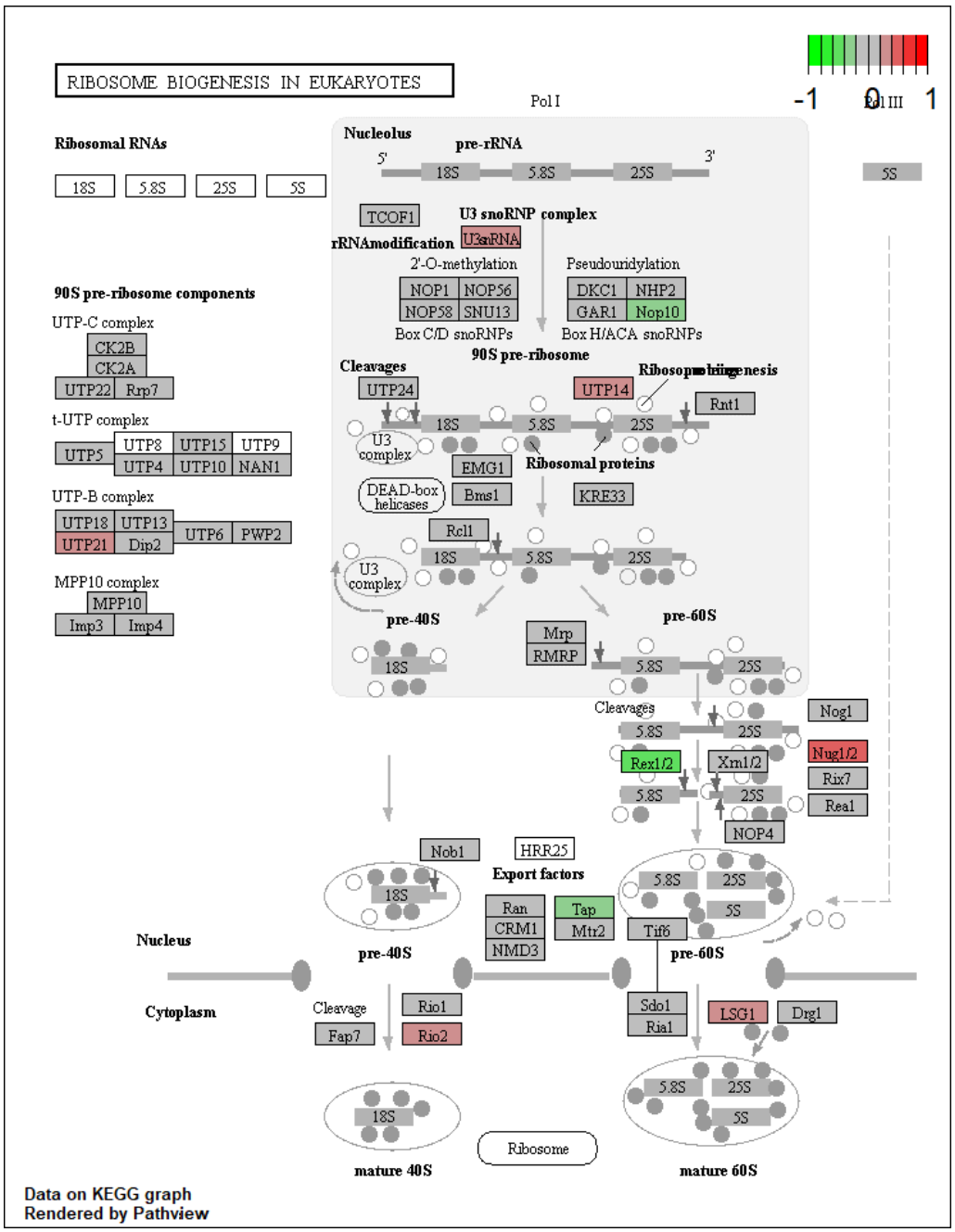


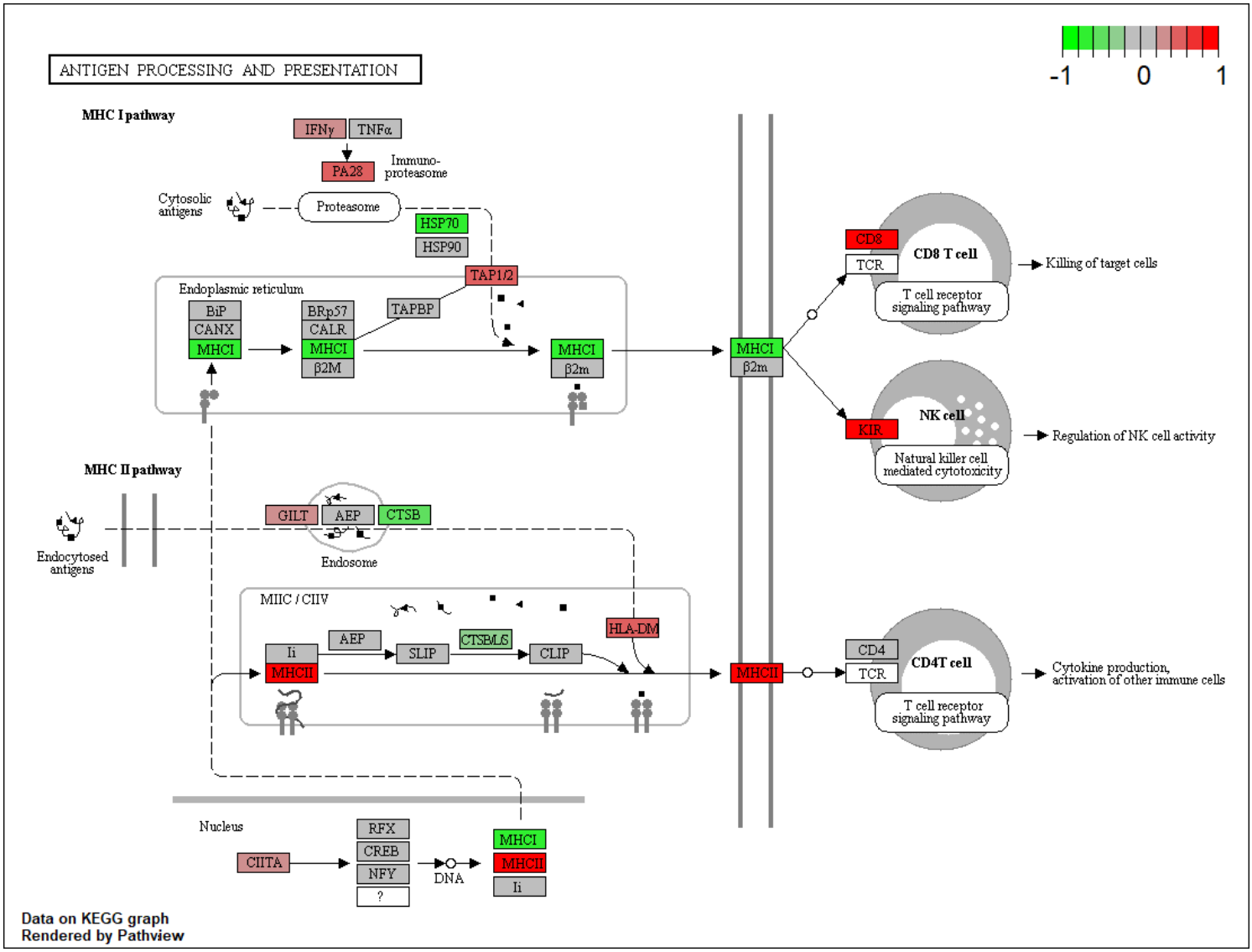

**Figure 1**

The Pipeline of the differentially expressed gene (DEG) analysis in this study.



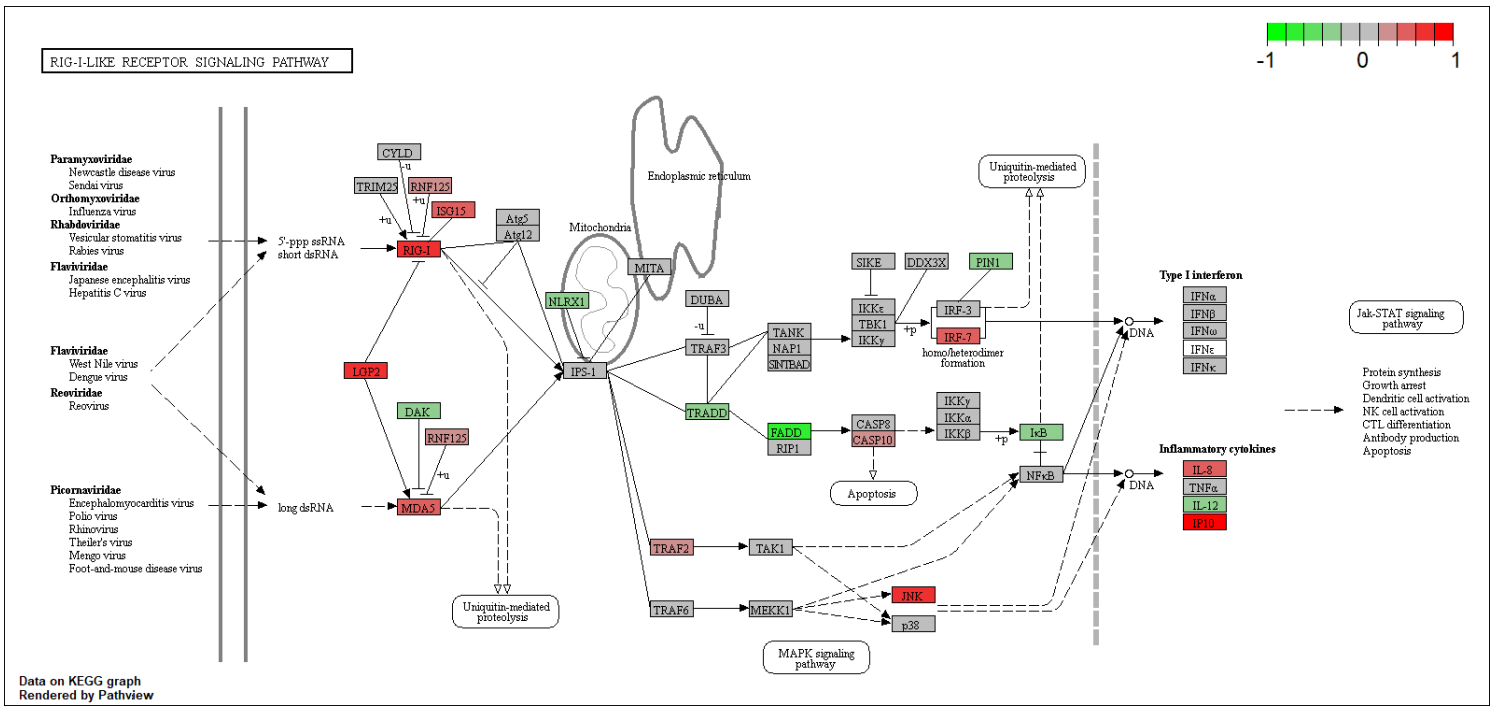

## Figure 2

Ribosome biogenesis in eukaryotesannotated with up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.



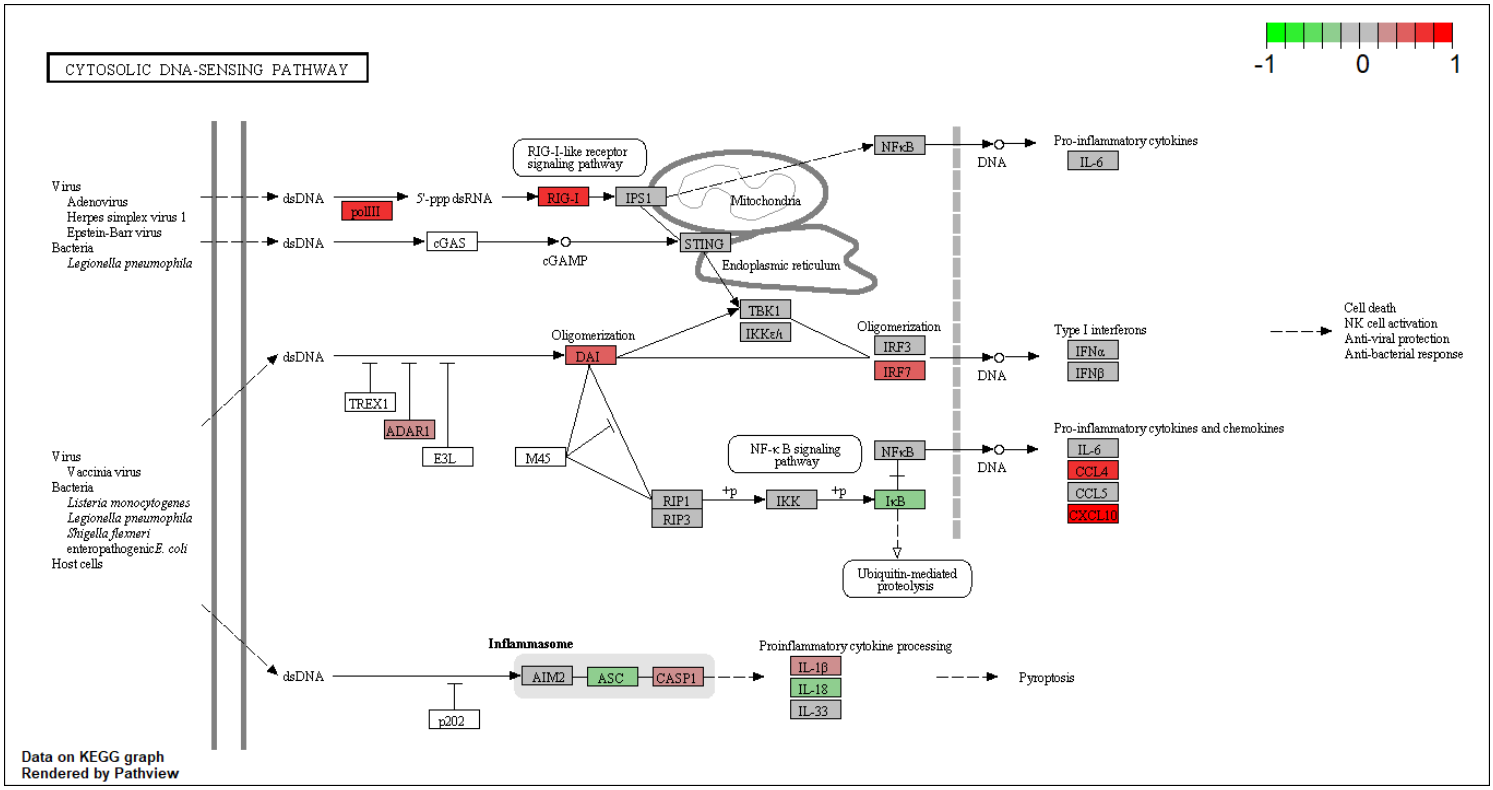

## Figure 3

Antigen processing and presentation with up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.



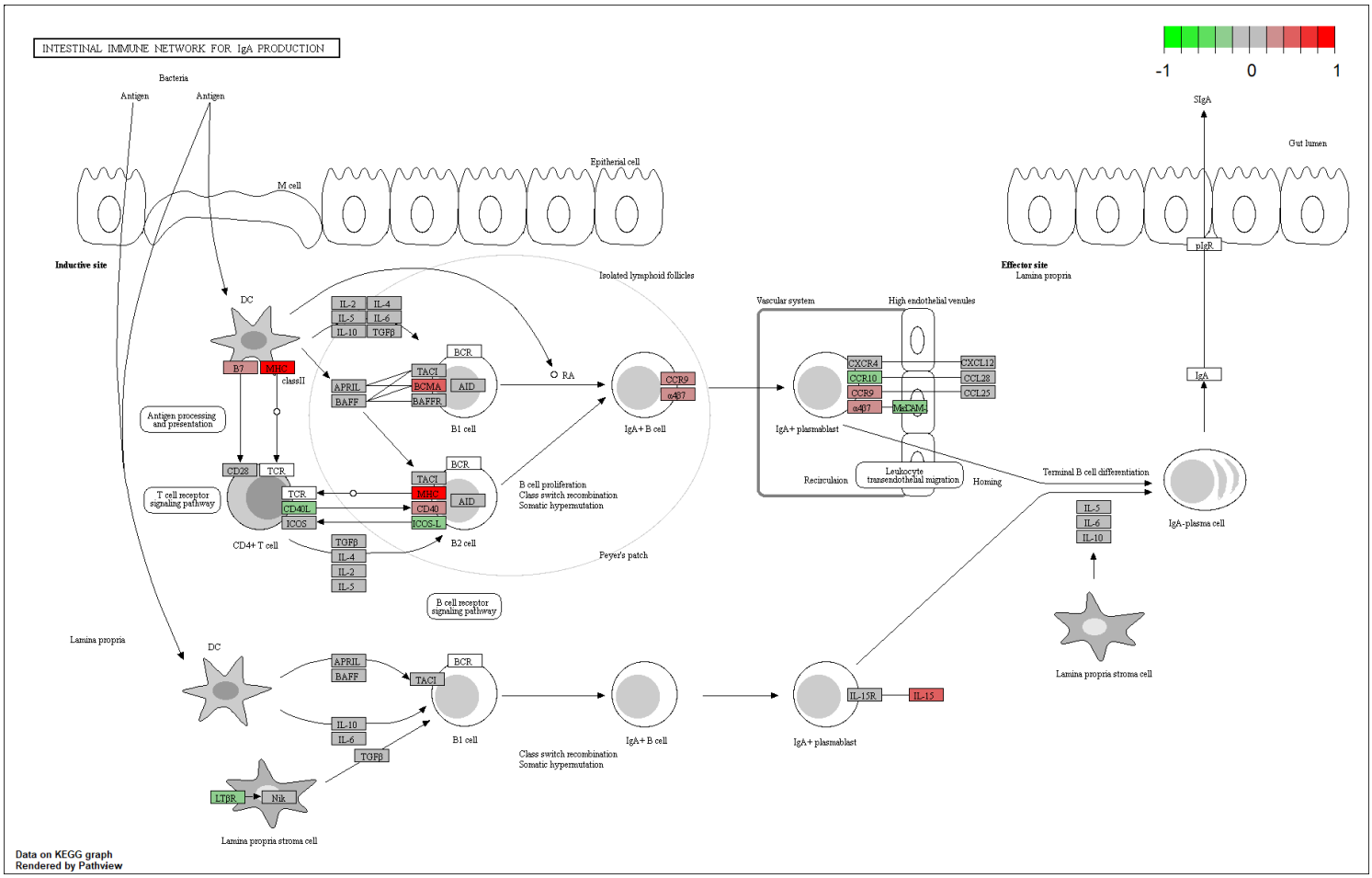

## Figure 4

RIG-I-like receptor signaling pathway with up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.



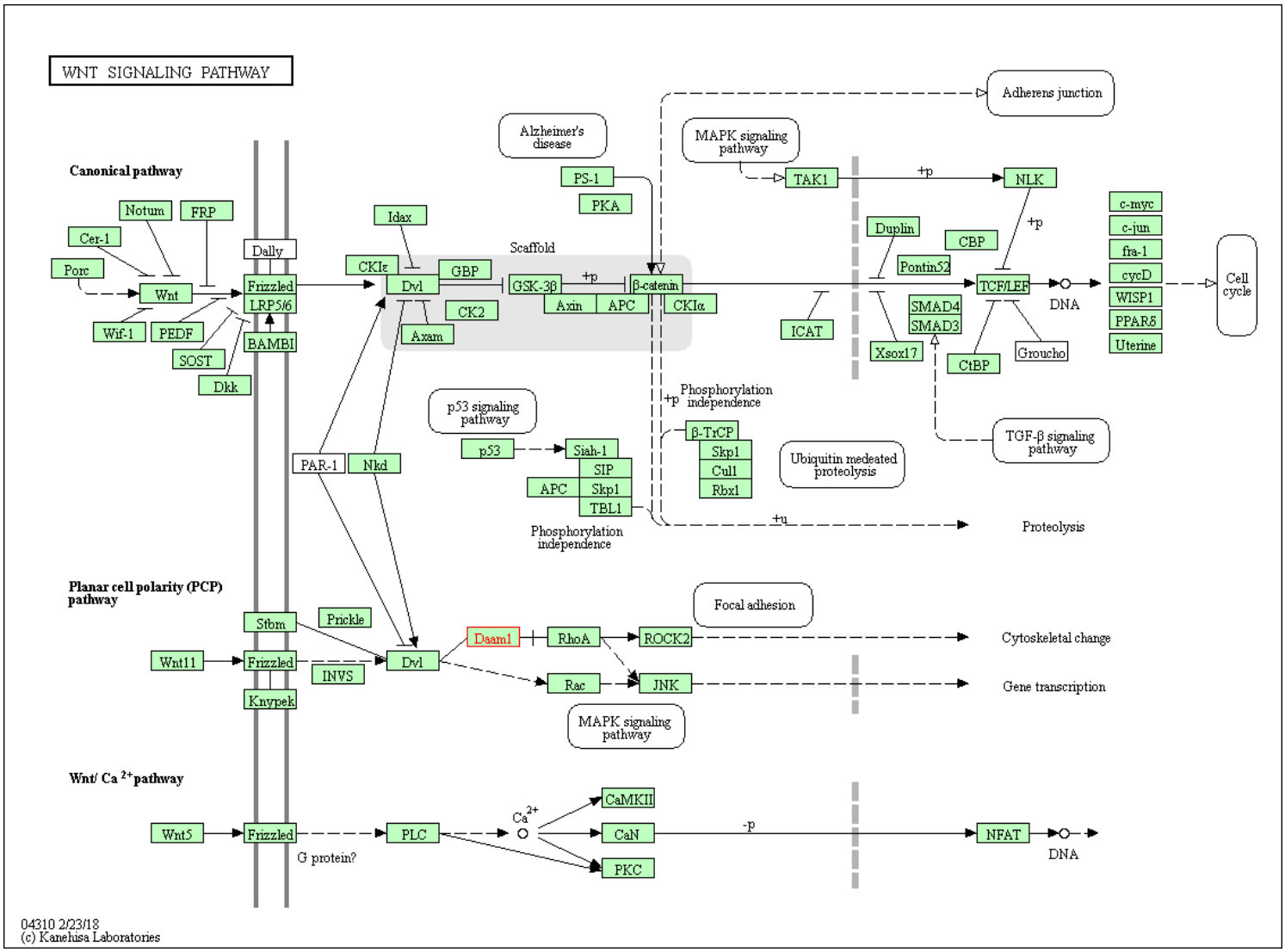

## Figure 5

Cytosolic DNA-sensing pathway with up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.




## Figure 6

Intestinal immune network for IgA production annotated with up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.




## Figure 7

Wnt signaling pathway annotated with DAAM2 up-regulated genes. Reprinted with permission from Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/kegg1.html.