**Genome-wide DNA methylation patterns in CD4+ T reveal significant contribution of DNA methylation to rheumatoid arthritis**

Shicheng Guo1, Ting Jiang1,2, Rongsheng Wang1,2, Yi Shen1,2, Xiao Zhu3, Fengmin Bai1,2, Qin Ding1,2

Guangjie Chen4×, Dongyi He1,2×

1Department of Rheumatology, ShanghaiGuanghuaHospital of Integrated Traditional and Western Medicine, Shanghai 200052, China.

2Arthritis Institute of integrated Traditional and Western medicine, Shanghai Chinese Medicine Research Institute, Shanghai 200052, China.

3Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Dongguan Scientific Research Center, GuangdongMedicalUniversity, Dongguan 523808, China

4Department of Immunology and Microbiology, ShanghaiJiaoTongUniversitySchool of Medicine, Shanghai 200025, China

Corresponding authors:

Dongyi He, Department of Rheumatology, Guanghua Integrative Medicine Hospital, Changning District, Shanghai 200052, China, Phone: +86-21-55664885, Fax: +86-21-55664885, E-mail: [dongyihe@medmail.com.cn](mailto:dongyihe@medmail.com.cn)

Guangjie Chen, Department of Immunology and Microbiology, ShanghaiJiaoTongUniversitySchool of Medicine, Shanghai 200025, China. Phone: +86-21-63846590,Email: [guangjie\_chen@163.com](mailto:guangjie_chen@163.com)

**Abstract**

Rheumatoid arthritis (RA) is an autoimmune disease that causes chronic inflammation of the joints. Evidence showed RA was involved with genetic and epigenetic aberrant. Recent evidence found more and more importance of the epigenetic contribution, especially the DNA methylation, to the pathogenesis of rheumatoid arthritis. To understand the extent and nature of dysregulated DNA methylation in rheumatoid arthritis T cells, we performed a genome-wide DNA methylation study in CD4+ T cells in 12 rheumatoid arthritis patients compared to 12 matched normal healthy controls. [Methods and Result] Cytosine methylation status was quantified with Illumina methylation 450K microarray (HM450K, 485512 CpG sites). We identified 810 hypomethylated and 392 hypermethylated CpG sites in RA CD4+ T cells compared to normal controls, representing 383 and 785 genes hypermethylated and hypomethylated in RA patients (P<3.4×10-7). Cluster analysis based on significantly differential methylated loci showed distinct separation between RA and normal controls. Gene ontology analysis showed alternative splicing (P=1.2×10-7, FDR) and phosphoprotein (1.7×10-2, FDR) were significantly aberrant in RA patients, indicating the abnormal of transcript alternative splicing and protein modification mediated by DNA methylation might play important role in the pathogenesis of rheumatoid arthritis. What’s more, the result showed human leukocyte antigen (HLA) region was frequently hypomethylated in RA patients, including HLA-DRB6, HLA-DQA1 and HLA-E, however, HLA-DQB1 showed different methylation profiles with significant hypermethylation in CpG island region and hypomethylation in CpG shelf region. Outside of the MHC region, the most hypermethylated genes in RA included HDAC4, NXN, TBCD and TMEM61 while the most significant hypomethylated genes included ITIH3, TCN2, PRDM16, SLC1A5 and GALNT9. [Conclusion] Genome-wide DNA methylation patterns revealed significant DNA methylation change in CD4+ T cells from patients with rheumatoid arthritis.

**Keywords:**

[DNA Methylation](http://www.tandfonline.com/action/doSearch?Keyword=DNA%20Methylation)，rheumatoid arthritis，CD4+ T cells， Genome-wide，Illumina methylation 450k microarray

**Introduction**

Rheumatoid arthritis (RA) is an autoimmune disease that mainly attacks the synovial tissues and therefore causes chronic inflammation of the joints. Autoimmune diseases cause the body’s immune system to mistake its own tissues for foreign invaders. The confused immune system develops antibodies to seek out and destroy the self-tissues in the synovium. About 0.5-1% of the population in the world were effected by RA[[1](#_ENREF_1)] and the corresponding economic burden was heavy all over the world.

In the past decades, the etiological factors of RA have been widely exploited. The most importance etiological source of RA comes from single nucleotide polymorphisms (SNP). Genome-wide association study based on SNPs have identified as many as SNPs which were believed to be significantly associated with RA[[2](#_ENREF_2), [3](#_ENREF_3)]. However, as demonstrated in our previous work, even for high familial risk disease as thyroid cancer, a few of the significant SNPs could just have limited prediction power [[4](#_ENREF_4)]. As the expectation, large number of copy number variations were demonstrated to be significantly associated with the [susceptibility to](http://www.ncbi.nlm.nih.gov/pubmed/23150419) RA[[5](#_ENREF_5), [6](#_ENREF_6)]. With the increasing knowledge of large number of genome-wide genetics association study, it is very interesting, it has been demonstrated that the acknowledged systemic autoimmune rheumatic diseases, such as systemic lupus erythematosus (SLE), gout, systemic Sclerosis (SSc), Sjögren’s syndrome (SjS) and osteoarthritis shared some common genetic risk and had some similar clinical characteristics.

The current estimated heritability of RA was about 20%-50% and it is significantly different between anti-citrullinated protein antibody positive RA and anti-citrullinated protein antibody negative RA[[7](#_ENREF_7)] which indicated that the epigenetic factors played more important roles in the etiology of RA. In addition, genome-wide DNA methylation of SLE[[8](#_ENREF_8), [9](#_ENREF_9)], SjS[[10](#_ENREF_10)] have showed that large number of DNA methylation were changed in the CD4+ or CD8+ T-cells in the patients. Therefore, it could be postulated that DNA methylation would also be significantly involved in the pathogenesis of RA. DNA methylation is one of most important epigenetic modifications. In our previous study, it has been found DNA methylation played important role in gene [[11](#_ENREF_11)]and microRNA [[12](#_ENREF_12)]expression regulation, cancer initiation/progress and therefore has been play important role in the cancer diagnosis [[13](#_ENREF_13), [14](#_ENREF_14)] and prognosis.

Based on above assumption, several studies has been conducted to identify DNA methylation changes in the RA genome and several important immune-related genes have been found to be aberrantly methylated in RA genome[[15-18](#_ENREF_15)]. However, the genome-wide DNA methylation profile of RA patient was still very limited, especially in Asian population and cannot be access publically. It provided big problem for the candidate gene based research to discover more RA related DNA methylation regions and to identify missing heritability of RA. In the present study, genome-wide DNA methylation profile was established in the CD4+ T-cells of 12 RA and matched healthy individuals and the contribution of the DNA methylation to RA were evaluated.

**Results**

We studied genome-wide CD4+ T-cell DNA methylation in rheumatoid arthritis patients and controls using a high-throughput method based on bead microarrays that allow simultaneous screening of 482,421 CpG sites spanning the promoter region of 21,231 genes (99% RefSeq genes). Twenty-four biological replicates (12 rheumatoid arthritis patients and 12 matched healthy controls) were collected and enrolled in the project. ([Table 1](http://www.tandfonline.com/action/showPopup?citid=citart1&id=T1&doi=10.4161/epi.6.5.15374)). 1067 probes were removed from the dataset filtered with detection p-value above 0.01. 443 probes were removed from the dataset with a beadcount <3 in at least 5% of samples. 29021 probes with SNPs were removed to decrease the bias of the analysis. 8510 and 11245 probes were removed since multiple alignment or located in chromosome X or chromosome Y. Eventually, 435226 probes were kept in the 24 samples.

In order to make sure the cells in our study were significantly derived from CD4+ T-cell, we projected our methylation signals into the PC1 and PC2 dimension of whole blood cell. The analysis showed our samples were clearly aggregated with CD4+ cells and was estranged with other kinds of cells, such as CD8+ T-cell, CD14+, CD19+ and so on which indicates the samples were successfully prepared in the process of sample collection. What’s more, to ensure that differential methylation patterns identified in our study were not influenced by potential differences in T-cell subset populations between RA patients and controls, we examined the methylation status of a number of genes known to be demethylated in particular T-cell subsets, such as  IL4, IL13 (Th2), IFNG (Th1) and IL17F (Treg). None of these CpG sites were found to be significantly different between case and control, suggesting no difference in T-cell subsets between RA patients and controls.

Meanwhile, PCA analysis was conducted to our methylation 450K, the result showed PC1 and PC2 explained the total variance of 29.9% and 14%, respectively while the top 13 principle component could explain as high as 80% variance. These results indicated the information of the data composed by limited informative clinical or demographical signals and our data were creditable for the further bioinformatic and biostatistical analysis (Supplementary Figure 1).

**Genome-wide DNA methylation profile of rheumatoid arthritis**

We identified 810 hypomethylated and 392 hypermethylated CpG sites in RA CD4+ T cells compared to normal controls, representing 785 and 383 genes which were hypomethylated and hypermethylated in RA patients with P<3.4×10-7(Paired t-test, FDR<0.05, Supplementary Table 1). More hypomethylated CpG site than hypermethylated CpG site indicated genome-wide hypo-methylation in the genome of rheumatoid arthritis. Interaction maps of differentially-methylated genes was constructed based on string 10.0 and the result showed these DMGs were highly interacted rather than functionally isolated (Figure 1C). Gene ontology analysis showed alternative splicing (P=1.2×10-7, FDR) and phosphoprotein (P=1.7×10-2, FDR) were significantly aberrant in RA patients (Table 2), indicating the abnormal of transcript alternative splicing and protein modification mediated by DNA methylation might play important role in the pathogenesis of rheumatoid arthritis.

What’s more, the result showed human leukocyte antigen (HLA) region was frequently hypomethylated in RA patients, including HLA-DRB6 (P=6.61×10-10), HLA-DQA1 (P=7.09×10-9) and HLA-E (P=3.24×10-7), however, HLA-DQB1 showed different methylation profiles with significant hypermethylation in CpG island region and hypomethylation in CpG shelf region(Table 3). Outside of the MHC region, the most hypermethylated genes in RA included HDAC4 (P=1.47×10-7), NXN(P=5.5×10-9), TBCD (P=4.48×10-8) and TMEM61(P=1.7×10-7) while the most significant hypomethylated genes included ITIH3 (P=1.16×10-7), TCN2 (P=1.57×10-8), PRDM16 (P=3.1×10-9), SLC1A5 (P=2.94×10-7)and GALNT9(P=8.26×10-9).

**Correlation between DNA methylation and disease characteristics.**

As table 1 showed, large number of clinical characteristics were records for our patients. Identifying clinical relevant DNA methylation loci would provide important insight to the pathological mechanism of RA and valuable clinical application. Association analysis between clinical characteristics and identified differential methylated loci were conducted. We found the methylation level of OR5A2 (cg02981094, P=2.6×10-4), ALDH9A1 (cg03984859, P=2.8×10-4) and C5orf32 (cg02070114, P=2.2×10-4) were significantly associated with the disease course of RA. In addition, the methylation level of ZC3H11A (cg02337583) was significantly associated with the level of [rheumatoid factor](http://www.webmd.com/rheumatoid-arthritis/guide/rheumatoid-factor-test) in RA patients (P=8.9×10-4). What’s more, the methylation level of OAS2 (cg00085448) was significantly associated with the HZPG in RA patients (P=4.1×10-4). In addition, C16orf71 (cg04705084), LOC100129716 (cg00598143) and miR-762 (cg02558026) were found to be significantly associated with DAS 28 with P-value of 5.8×10-3, 5.2×10-3, and 7.5×10-3, respectively. Five loci which included SLC38A8 (cg01740650, P=3.0×10-3), C18orf19 (cg00448482, P=3.0×10-3), COL18A1 (cg04760448, P=1.9×10-3), BAT3 (cg05649229, P=4.9×10-3) and PLD3 (cg07071106, P=4.4×10-3) were significantly associated with ESR. In the end, we found HSPA12A (cg06942850) was significantly associated with tender joint count (TJC) with P-value of 3.2×10-3.

**Discussion**

In summary, we built the genome-wide DNA methylation profile of the CD4+ T-cell in the 12 RA and matched 12 health individuals from Chinese Han population with Illumina methylation 450K microarray. The preparation of the CD4+ T-cell and the methylation measurement were demonstrate to be of high creditability. We found 1202 CpG loci were significantly different in the CD4+ T-cells between RA and controls. Gene ontology and interaction analysis showed these genes were functional related and interacted with each other, indicating these genes might play some role in the pathogenesis of RA in the form of pathway or functional system. More hypomthylated regions were identified compared with hypermethylated region indicating the genome-wide hypomethylation status of the RA genome and it was consistent with previous report [[19](#_ENREF_19)].

Compared with DNA methylation change of cancer which usually contains thousand differential methylated loci, systemic autoimmune rheumatic diseases seems only have few differential methylation regions. PCA analysis to our methylation dataset also revealed there is no significant separation between RA and control indicating there would be not so many different methylation regions in CD4+ T-cells between RA and normal individual. Kazuhisa and colleagues conducted a genome-wide DNA methylation comparison between fibroblast-like synoviocytes (FLS) and RA with Human Methylation 450K microarray and they identified 1859 differential methylation loci[[20](#_ENREF_20)]. Matlock and colleagues identified 341 differential methylation loci in CD4+ between SLE and health control[[21](#_ENREF_21)]. Nezam and colleagues identified 753 differential methylation loci in CD4+ between SjS and health control[[10](#_ENREF_10)]. Although multiple test correction has been conducted in these papers, we believe, there are still large number of differential methylated loci would be false positive. Therefore, the candidate differential methylated loci underlying the pathogenesis of RA would be limited. Compared with Kazuhisa’s study, there are 81 shared differential methylated CpG loci which might be very important in the pathogenesis of RA. More genome-wide DNA methylation profile from different population such as Caucasian, Japanese as well as African would provide powerful source to identify more RA related methylation regions.

**Methods**

**Patients and controls.**

Twelve RA patients and 12 matched controls were studied ([**Table 1**](http://www.tandfonline.com/action/showPopup?citid=citart1&id=T1&doi=10.4161/epi.6.5.15374)). The mean age was 42.83 in patients and 43.75 in controls without significant difference (p = 0.95). All patients fulfilled the American College of Rheumatology classification criteria for RA. The research was approved by academic advisory board of Guanghua Hospital. Clinical data were collected at the time of sample collection.

**Isolation of PBMCs and CD4+ T cells**

(PBMCs) Mononuclear cells were prepared from blood specimens of RA and normal samples by Ficoll-Hypaque centrifugation(Amersham Biosciences) using the standard protocol and immediately processed for cell culture[[22](#_ENREF_22)]. CD4+ T cells were prepared from freshly isolated PBMCs by depleting cells expressing CD8, CD14, CD16, CD19, CD36, CD56, CD123,γ/δ T cell receptors, and glycophorin A using No-Touch T cellisolation kits (Miltenyi Biotec). The purity of the CD4+ T cells was 95–98%, as determined by flow cytometry using specific antibodies. Whole genomic DNA was then prepared using a Qiagen DNEasy kit (Qiagen, Germantown, MD) and bisulfite-treated using a Zymo EZ DNA Methylation Kit (Zymo, Orange, CA).

**Illumina Infinium Human Methylation 450.**

Bisulfite-converted patient and control DNA samples were prepared and quantified using a NanoDrop scanning spectrophotometer (Thermo, Wilmington, DE). For each sample, 500 ng of whole-genome bisulfite-converted DNA was denatured, fragmented, amplified and methylation signals were detected by [Infinium HumanMethylation450K BeadChip](http://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit.html) (Illumina, San Diego, CA, USA).

Standard DNA methylation 450K analysis pipeline (SMAP) was implemented to conduct the methylation microarray analysis. GenomeStudio (Illumina) was used to generate signal intensities and detection p values with internal control normalization (ICN) and background subtraction (BS). Quality control and normalization were conducted with R package of "lumi”. Probes with SNPs or located in ChrX and ChrY were removed before further analysis. In addition, probes with a detection p value > 0.01 exceeding 5% of the samples were also filtered out while other probes less than 5% of samples were labeled as missing value (NA) to avoid further bias in the following statistic and bioinformatic analysis. And then the overall signal intensity, the distribution of M-values and the number of significantly detected sites were used to measure the quality of the beadchip. Significant outlier samples/beadchip were removed before the differential methylation loci identification. Color-bias adjustment and quantile normalization (QN) were performed on signal intensities with the package of "lumi". Finally, beta-mixture quantile normalization (BMIQ) to the β-values were conducted to adjust the bias caused by different types of probes (type I and type II).

**Statistical analysis.**

Differentially methylated loci were identified by paired t-test based on the beta value of the normalization data. The raw P-values were adjusted by false discovery rate (FDR) = 0.05 for the multiple test correction. Association between clinical characteristics and differential methylation loci were conducted with linear regression with the significant threshold of 0.005. Gene ontology analysis was conducted with DAVID bioinformatics resources[[23](#_ENREF_23)]. The interactions among the differential methylated genes were inferred by String (version 10.0) [[24](#_ENREF_24)]. Human references of GRCh37/hg19 were used in the bioinformatics analysis and result representation. The data were deposited in the Gene Expression Omnibus (GEO accession: please waiting my response). All methods and analyses were performed in R (version 3.2.1).

**Acknowledgements**

This work was funded by the National Natural Science Funds of China (81273979), Shanghai clinical base construction of traditional Chinese medicine (ZY-LCPT-1), Shanghai intensive entity construction of integrated traditional and western medicine[rheumatoid arthritis](http://www.baidu.com/link?url=iT4ATL3sR6yg06cPHgVkbWxtLah4KbbGMOTKi-SQ9uMAjpCCVUsmrIFNHQtlWbRrIUgRFAMxUJ7_lQZh6RW0w9FcG5RyLZskoFBvt8TFc7_)(zxbz2012-05) and Shanghai clinical intensive subject construction of traditional Chinese medicine-traditional Chinese [rheumatology](http://www.baidu.com/link?url=4Qku9Zb2wBIua7s1DyxfDyF0XZzIFL0m6nyz4LW66UiC3TyBDc31URXd5h_2afOtUNo74DSjURDzbTz6pNjZfkyaRMFKEfnfwCYRjU9oLDu&wd=&eqid=e102a46600012e760000000355827dac)(ZYXK2012012)

**References**

1. Firestein, G.S., *Evolving concepts of rheumatoid arthritis.* Nature, 2003. **423**(6937): p. 356-61.

2. Acosta-Colman, I., et al., *GWAS replication study confirms the association of PDE3A-SLCO1C1 with anti-TNF therapy response in rheumatoid arthritis.* Pharmacogenomics, 2013. **14**(7): p. 727-34.

3. Eleftherohorinou, H., et al., *Pathway-driven gene stability selection of two rheumatoid arthritis GWAS identifies and validates new susceptibility genes in receptor mediated signalling pathways.* Hum Mol Genet, 2011. **20**(17): p. 3494-506.

4. Guo, S., Wang, Y.-L., Li, Y., Jin, L., Xiong, M., Ji, Q.-H. and Wang, J, *Significant SNPs have limited prediction ability for thyroid cancer.* CANCER MEDICINE, 2014. **doi: 10.1002/cam4.211**.

5. Song, X., et al., *Association between HLA-DQA1 gene copy number polymorphisms and susceptibility to rheumatoid arthritis in Chinese Han population.* J Genet, 2014. **93**(1): p. 215-8.

6. Chen, J.Y., et al., *Association of FCGR3A and FCGR3B copy number variations with systemic lupus erythematosus and rheumatoid arthritis in Taiwanese patients.* Arthritis Rheumatol, 2014. **66**(11): p. 3113-21.

7. Frisell, T., et al., *Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age.* Arthritis Rheum, 2013. **65**(11): p. 2773-82.

8. Javierre, B.M., et al., *Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.* Genome Res, 2010. **20**(2): p. 170-9.

9. Richardson, B., et al., *Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis.* Arthritis Rheum, 1990. **33**(11): p. 1665-73.

10. Altorok, N., et al., *Genome-wide DNA methylation patterns in naive CD4+ T cells from patients with primary Sjogren's syndrome.* Arthritis Rheumatol, 2014. **66**(3): p. 731-9.

11. Wang, X., et al., *Hypermethylation reduces expression of tumor-suppressor PLZF and regulates proliferation and apoptosis in non-small-cell lung cancers.* FASEB J, 2013. **27**(10): p. 4194-203.

12. He, Y., et al., *Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma.* Neoplasia, 2011. **13**(9): p. 841-53.

13. Guo, S., et al., *Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC).* Clin Epigenetics, 2015. **7**(1): p. 3.

14. Guo, S., et al., *Quantitative assessment of the diagnostic role of APC promoter methylation in non-small cell lung cancer.* Clin Epigenetics, 2014. **6**(1): p. 5.

15. Glossop, J.R., et al., *DNA methylation profiling of synovial fluid FLS in rheumatoid arthritis reveals changes common with tissue-derived FLS.* Epigenomics, 2015. **7**(4): p. 539-51.

16. Liebling, M.R., *Methylation of the CTLA-4 promoter and Treg cell dysfunction in rheumatoid arthritis: comment on the article by Cribbs et al.* Arthritis Rheumatol, 2015. **67**(5): p. 1406.

17. van Steenbergen, H.W., et al., *Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study.* Rheumatology (Oxford), 2014. **53**(12): p. 2317-8.

18. Glossop, J.R., et al., *Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations.* Epigenetics, 2014. **9**(9): p. 1228-37.

19. Karouzakis, E., et al., *Increased recycling of polyamines is associated with global DNA hypomethylation in rheumatoid arthritis synovial fibroblasts.* Arthritis Rheum, 2012. **64**(6): p. 1809-17.

20. Nakano, K., et al., *DNA methylome signature in rheumatoid arthritis.* Ann Rheum Dis, 2013. **72**(1): p. 110-7.

21. Jeffries, M.A., et al., *Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus.* Epigenetics, 2011. **6**(5): p. 593-601.

22. Chen, G., et al., *Role of osteopontin in synovial Th17 differentiation in rheumatoid arthritis.* Arthritis Rheum, 2010. **62**(10): p. 2900-8.

23. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

24. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.

**Figure Legend**

Figure 1. genome-wide DNA methylation profile of RA and bioinformatics of differential methylation loci. A, PCA analysis showed the CD4+ cells from our study were aggregated with CD4+ T-cells from GSE35069, indicating the samples were well prepared. B, hierarchical cluster analysis based on differential methylated loci separated samples into RA and control group. C. Gene-gene interaction analysis to differential methylated genes inferred by protein-protein interaction database of String 10.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1. Characteristic of the enrolled RA and control samples | | | | | | | | | | | | | |
| SSID | Age | Gender | COD  (Year) | RF  (IU/mL) | Anti-CCP  (RU/ml) | ESR  (mm/h) | SJC | TJC | PGA | DAS28 | Control | Age | Gender |
| RA0001 | 47 | Male | 6 | <20 | 158.21 | 29 | 4 | 6 | 70 | 5.27 | HP0001 | 47 | Male |
| RA0010 | 37 | Female | 0.8 | 22.2 | 792.81 | 14 | 0 | 0 | 10 | 1.99 | HP0010 | 35 | Female |
| RA0011 | 52 | Male | 0.2 | 2260 | 104.86 | 46 | 10 | 12 | 60 | 6.35 | HP0011 | 52 | Male |
| RA0012 | 56 | Male | 1.5 | 1000 | 1600 | 140 | 10 | 11 | 85 | 7.39 | HP0012 | 56 | Male |
| RA0002 | 40 | Female | 3 | 198 | 789.54 | 36 | 1 | 1 | 30 | 3.77 | HP0002 | 39 | Female |
| RA0003 | 47 | Female | 3 | 152 | 270.38 | 140 | 18 | 19 | 90 | 8.35 | HP0003 | 46 | Female |
| RA0004 | 23 | Female | 1.5 | <20 | 71.42 | 24 | 3 | 4 | 45 | 4.46 | HP0004 | 23 | Female |
| RA0005 | 28 | Female | 11 | 127 | 306.09 | 54 | 18 | 18 | 80 | 7.48 | HP0005 | 27 | Female |
| RA0006 | 39 | Female | 20 | <20 | <25 | 66 | 3 | 3 | 50 | 5.09 | HP0006 | 40 | Female |
| RA0007 | 25 | Female | 4 | 1020 | 1540.83 | 23 | 18 | 19 | 55 | 6.59 | HP0007 | 27 | Female |
| RA0008 | 57 | Female | 10 | <20 | <25 | 78 | 3 | 3 | 60 | 5.34 | HP0008 | 57 | Female |
| RA0009 | 39 | Female | 15 | 1200 | 178.88 | 75 | 22 | 26 | 70 | 8.17 | HP0009 | 38 | Female |

COD:  course of a disease, SJC: swollen joint count; TJC: tender joint count; PGA: patient global assessment; DAS 28: disease activity score in 28 Joints

Table 2. Gene ontology analysis to differential methylation genes in RA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Term | Count | Freq(%) | P-Value | Fold Enrichment | Benjamini |
| Alternative splicing | 415 | 47.64638347 | 2.5244E-10 | 1.269098017 | 1.23191E-07 |
| Splice variant | 417 | 47.87600459 | 1.38522E-10 | 1.27222302 | 3.11674E-07 |
| Alternative products | 398 | 45.6946039 | 3.72753E-08 | 1.233381223 | 9.31881E-07 |
| Pituitary | 158 | 18.14006889 | 4.1045E-07 | 1.449208795 | 3.20146E-05 |
| Adrenal gland | 191 | 21.92881745 | 2.6667E-05 | 1.300196345 | 0.001039486 |
| Olfactory Bulb | 133 | 15.26980482 | 6.70298E-05 | 1.37534001 | 0.001306272 |
| Kidney | 287 | 32.95063146 | 6.46643E-05 | 1.195313738 | 0.001679913 |
| Colorectal Adenocarcinoma | 167 | 19.17336395 | 0.000135274 | 1.296805361 | 0.002108194 |
| Cell | 680 | 78.07118255 | 0.000396987 | 1.031933696 | 0.002181479 |
| Appendix | 208 | 23.88059701 | 0.00032176 | 1.229351318 | 0.003132752 |
| Thymus | 156 | 17.91044776 | 0.000318951 | 1.29067259 | 0.003548285 |
| Cardiac Myocytes | 372 | 42.70952928 | 0.000475 | 1.124007997 | 0.004109177 |
| Cell part | 680 | 78.07118255 | 0.000388116 | 1.032003299 | 0.004261 |
| Testis | 124 | 14.23650976 | 0.00078898 | 1.316874949 | 0.00558116 |
| Dorsal root ganglia | 101 | 11.59586682 | 0.000720311 | 1.372389316 | 0.005604686 |
| Chronic myelogenous leukemia | 99 | 11.36624569 | 0.000865408 | 1.371117131 | 0.005611783 |
| Occipital Lobe | 138 | 15.84385763 | 0.00261201 | 1.253090466 | 0.013508155 |
| Signaling by Rho GTPases | 16 | 1.836969001 | 0.000304882 | 2.888995164 | 0.013928772 |
| White blood cells monocyte | 72 | 8.266360505 | 0.00165455 | 1.43988182 | 0.028321431 |
| Testis Interstitial | 183 | 21.01033295 | 0.008720455 | 1.170122869 | 0.037243129 |

Table 3. Differential methylation status of HLA genes in RA

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gene | CpG site | P-Value | Delta-Beta | Ratio-Beta | Case | Control | Chrosome | Start | End | CpG Shore | Type |
| HLA-DOA | cg00540941 | 4.70E-08 | -0.18 | 0.65 | 0.34 | 0.52 | chr6 | 32974843 | 32974844 | N\_Shore | II |
| HLA-DQA1 | cg02919082 | 7.09E-09 | -0.38 | 0.44 | 0.30 | 0.67 | chr6 | 32605694 | 32605695 | NA | II |
| HLA-DQA1 | cg04054303 | 9.83E-12 | -0.47 | 0.40 | 0.31 | 0.78 | chr6 | 32606445 | 32606446 | NA | II |
| HLA-DQB1 | cg02902672 | 1.62E-08 | -0.63 | 0.18 | 0.14 | 0.76 | chr6 | 32635360 | 32635361 | S\_Shelf | II |
| HLA-DQB1 | cg04777551 | 1.44E-08 | -0.46 | 0.47 | 0.41 | 0.87 | chr6 | 32628953 | 32628954 | N\_Shelf | II |
| HLA-DQB2 | cg02964065 | 1.99E-07 | 0.08 | 1.11 | 0.77 | 0.69 | chr6 | 32729545 | 32729546 | Island | I |
| HLA-DRB1 | cg00211215 | 1.60E-07 | 0.66 | 3.39 | 0.94 | 0.28 | chr6 | 32552246 | 32552247 | Island | I |
| HLA-DRB1 | cg04026937 | 1.60E-10 | -0.62 | 0.08 | 0.05 | 0.66 | chr6 | 32549361 | 32549362 | N\_Shelf | II |
| HLA-DRB1 | cg06032479 | 3.67E-08 | 0.20 | 1.27 | 0.92 | 0.73 | chr6 | 32552026 | 32552027 | Island | I |
| HLA-DRB1 | cg06204447 | 4.05E-08 | -0.18 | 0.76 | 0.57 | 0.74 | chr6 | 32546665 | 32546666 | NA | II |
| HLA-DRB6 | cg00103771 | 6.61E-10 | -0.64 | 0.24 | 0.20 | 0.84 | chr6 | 32525805 | 32525806 | NA | II |
| HLA-DRB6 | cg04688450 | 5.38E-09 | -0.34 | 0.44 | 0.27 | 0.62 | chr6 | 32526366 | 32526367 | NA | II |
| HLA-DRB6 | cg06559318 | 1.17E-08 | -0.82 | 0.05 | 0.04 | 0.85 | chr6 | 32526260 | 32526261 | NA | I |
| HLA-E | cg02678305 | 4.74E-08 | -0.05 | 0.89 | 0.39 | 0.43 | chr6 | 30460322 | 30460323 | S\_Shelf | II |
| HLA-E | cg03725115 | 3.24E-07 | -0.02 | 0.72 | 0.05 | 0.07 | chr6 | 30458102 | 30458103 | Island | I |

Deta-beta represent the difference between RA and controls. Ratio-beta represent the fold-change of the average methylation level in RA compared with controls.