A Comparison of Co-methylation Relationships Between Rheumatoid Arthritis and Parkinson’s Disease

Guoping Tang†, Hongzhi Pan†, Liangde Xu†, Rennan Feng4, Yongshuai Jiang3*, Fanwu Kong5* and Simeng Hu3*

† The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Jinhua, China, 2 Collaborate Research Center, Shanghai University of Medicine and Health Sciences, Shanghai, China, 3 College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, 4 Department of Nutrition and Food Hygiene, Public Health College, Harbin Medical University, Harbin, China, 5 Department of Nephrology, The Second Affiliated Hospital, Harbin Medical University, Harbin, China

Rheumatoid arthritis (RA) is a complex autoimmune disease. Recent studies have identified the DNA methylation loci associated with RA and found that DNA methylation was a potential mediator of genetic risk. Parkinson’s disease (PD) is a common neurodegenerative disease. Several studies have indicated that DNA methylation levels are linked to PD, and genes related to the immune system are significantly enriched in PD-related methylation modules. Although recent studies have provided profound insights into the DNA methylation of both RA and PD, no shared co-methylation relationships have been identified to date. Therefore, we sought to identify shared co-methylation relationships linked to RA and PD. Here, we calculated the Pearson’s correlation coefficient (PCC) of 225,239,700 gene pairs and determined the differences and similarities between the two diseases. The global co-methylation change between PD cases and controls was larger than that between RA cases and controls. We found 337 gene pairs with large changes that were shared between RA and PD. This co-methylation relationship study represents a new area of study for both RA and PD and provides new ideas for further study of the shared biological mechanisms of RA and PD.

Keywords: Parkinson's disease (PD), rheumatoid arthritis, DNA methylation, Pearson's correlation coefficients, disease

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disease (Adkar et al., 2017; Nair et al., 2017; Choudhary et al., 2018; Liu and Page, 2018; Mizoguchi et al., 2018) with multiple environmental risk factors, including lifestyle factors, hormones, infections, and smoking, as well as interactions between genes and the environment (Klareskog et al., 2009; Scott et al., 2010; Karlson and Deane, 2012; Aleyd et al., 2016; Choudhary et al., 2018; Soulaidopoulos et al., 2018). Parkinson’s disease (PD) (Antony et al., 2013; Goldman, 2014), like Alzheimer’s disease (AD) (Zhou et al., 2018a), is a common neurodegenerative disease. The most important pathological change of PD is the degeneration of dopamine (DA) neurons in the midbrain, which causes a significant decrease in the striatum DA content and results in disease (Antony et al., 2013). The exact cause of this pathological change is still unclear. Genetic factors, environmental factors, aging, and oxidative stress may be involved in the degenerative death of PD dopaminergic neurons (Goldman, 2014).
In recent years, many studies have paid close attention to epigenetic mechanisms (Cui et al., 2018), especially DNA methylation (Liu et al., 2013; Davegårdh et al., 2018; Meehan et al., 2018). Epigenomic differences can represent phenotypic differences resulting from environmental exposure (Viatte et al., 2017). Increasingly, DNA methylation has been investigated as a potential diagnostic biomarker (Nair et al., 2017). DNA methylation occurs when a methyl group is added to cytosine DNA nucleotides to form 5-methylcytosine. This process is catalyzed by methyltransferases (Nair et al., 2017). A large number of studies have shown that DNA methylation can cause changes in chromatin structure, DNA conformation, DNA stability and DNA-protein interactions, to control gene expression (Hannon et al., 2016; Rao et al., 2018).

In 2013, Liu et al. analyzed the genome-wide DNA methylation differences in peripheral blood leukocytes (PBLs) from RA patients and normal controls (Liu et al., 2013). They applied the Causal Inference Test, with genotype as a causal factor, DNA methylation as a potential mediator and RA as the outcome, to identify the differential methylation positions (DMPs) associated with the RA phenotype. They found 264 unique single nucleotide polymorphisms (SNPs) and nine unique DMPs, comprising 535 SNP-DMP pairs, within the major histocompatibility complex region. Those SNP-DMP pairs represent potential methylation-mediated relationships between SNPs and RA disease risk, revealing that DNA methylation is a potential mediator of genetic risk. Many studies have indicated there are strong associations between neurodegenerative diseases and autoimmune diseases (Zhou et al., 2018b). In 2017, Chuang et al. found that PD status had a profound association with DNA methylation levels in blood and saliva, and three out of six PD-related CpG clusters in blood were significantly enriched for genes related to the immune system (Chuang et al., 2017).

Although recent studies have provided profound insights into the DNA methylation of RA and PD, they have not considered the interaction relationships between genes. RA and PD, as complex diseases, may be caused by the interaction of multiple genes (Yi et al., 2017). Based on this concept, in this study we analyzed one-to-one gene co-methylation relationships to further dissect common molecular mechanisms between RA and PD.

MATERIALS AND METHODS

DNA Methylation Data for RA
In this study, we obtained RA methylation data (GSE42861) from the NCBI GEO database. The platform was GPL13534 (Illumina HumanMethylation450 BeadChip, including 485,577 probes). GSE42861 contains 354 RA samples and 335 normal control samples. Bisulphite converted DNA was from PBLs. The matrix file of GSE42861 includes the information for samples and normalized beta values (from 0 to 1) for each sample.

DNA Methylation Data for PD
We also obtained PD methylation data (GSE111629) from the NCBI GEO database. The platform was also GPL13534. The matrix file includes 572 samples (335 PD samples and 237 normal control samples from whole-blood DNA) and normalized beta values (from 0 to 1) for each sample.

Mapping of Methylation Positions to Genes
In this study, we analyzed genes on a one-to-one basis to compare the global co-methylation landscapes. We used the same data processing for RA and PD. We mapped 485,577 probes in the GPL13534 platform to 21,225 genes. If there were multiple probes in a gene, we calculated the average beta value of those probes as the methylation value of the gene. The final RA dataset used for subsequent analysis consisted of 21,225 genes and 689 samples (including 354 RA samples and 335 normal control samples), and the final PD dataset included 21,225 genes, 335 PD samples, and 237 normal control samples.

The Co-methylation Coefficient
We used the Pearson's correlation coefficients (PCC) to measure the co-methylation relationships between gene pairs. We defined the PCC as:

\[ r(i,j) = \frac{1}{n-1} \sum_{k=1}^{n} \left( \frac{b_{i,k} - \bar{b}_i}{S_i} \right) \left( \frac{b_{j,k} - \bar{b}_j}{S_j} \right) \]

where \( r(i,j) \) is the PCC between gene i and gene j; \( n \) is the number of samples; \( b_{i,k} \) is the methylation beta value of gene i in sample k; \( b_{j,k} \) is the methylation beta value of gene j in sample k; \( \bar{b}_i \) is the mean of the methylation beta value of gene i in sample; \( \bar{b}_j \) is the mean of the methylation beta value of gene j; \( S_i \) is the standard deviation (SD) of the methylation beta value of gene i; and \( S_j \) is the SD of the methylation beta value of gene j. When \( r(i,j) > 0 \), gene i and gene j represent a positive correlation, that is, gene i and gene j are highly co-methylated. When the \( r(i,j) < 0 \), gene i and gene j represent a negative correlation. In other words, the methylation levels of gene i and gene j are exactly opposite.

RESULTS

Comparison of the Distribution of the PCC Between RA and PD
We calculated the PCC of 225,239,700 gene pairs for RA cases and RA controls. The mean PCC of the RA cases was 0.174 and the SD was 0.250, while the mean PCC of the RA controls was 0.209, and the mean PCC of the PD controls was 0.414 and the SD was 0.260. We then analyzed their distribution and plotted them on a graph (Figure 1A). Both the RA cases and controls showed a similar unimodal distribution of the global PCC.

We also calculated the PCC for the dataset of PD cases and controls. The mean PCC of the PD cases was 0.334 and the SD was 0.209, while the mean PCC of the PD controls was 0.414 and the SD was 0.194. Their distribution is shown in Figure 1B, where the global change of PD cases and controls is larger than that of RA cases and controls.

Identifying Large Change Gene Pairs for RA
We divided the PCC in eight intervals, named strongest negative correlation (−1 to −0.75), strong negative correlation...
large changes of the co-methylation relationship from controls to cases. These are shown in Supplementary File S1.

**Identifying Large Change Gene Pairs for PD**

For PD, we used the same method to divide the PCC into eight intervals, and changes of the co-methylation relationship from controls to cases are shown in Table 2.

From the PD controls to the PD cases, 15 gene pairs recorded the largest changes of the co-methylation relationship, crossing five intervals. These included two gene pairs (HIST1H2BH and NFKBIE, and HIST1H3F and NFKBIE) where the PCC changed from the strongest positive correlation to a weak negative correlation, and 13 gene pairs where the PCC changed from a strong positive correlation to a strong negative correlation.
Furthermore, 2,468 gene pairs crossed four intervals, including one gene pair from a weak negative correlation to a strong positive correlation, six gene pairs from the weakest positive correlation to the strongest negative correlation, 594 gene pairs from a weak positive correlation to a strong negative correlation, and 60 gene pairs from the weakest positive correlation to a strong negative correlation. In total, 196,311 gene pairs crossed no less than three intervals in PD, shown in Supplementary File S2. There was obviously a larger difference in the partial co-methylation relationship between controls and cases in PD than in RA.

Shared Gene Pairs Between RA and PD

To identify the genetic mechanisms shared in both RA and PD, we extracted the gene pairs that changed across no less than three intervals. We extracted 1,893 gene pairs from RA and 196,311 gene pairs from PD. We found there were 337 gene pairs (in Supplementary File S3) shared between RA and PD. We extracted 28 gene pairs with an absolute value of difference >0.8 from RA and 4,748 gene pairs from PD. Nine of these gene pairs (shown in Table 3) were shared in RA and PD, and these nine gene pairs were included in the 337 gene pairs extracted above.

Gene Ontology and KEGG Pathway Analysis of Shared Gene Pairs

To further analyze the functional characteristics of the shared gene pairs, we extracted 220 unique genes from the 337 gene pairs by removing duplicate genes, and then annotated them to GO (Gene Ontology) categories and KEGG pathways using DAVID. This found 70 GO categories and four KEGG pathways that were significantly \( P < 0.05 \) annotated. Among them, 11 immune-related GO categories included GO: 0002250: adaptive immune response, GO: 0050853: B cell receptor signaling pathway, GO: 0006952: defense response, GO: 0031295: T cell co-stimulation, GO: 0006955: immune response, GO: 0042832: defense response to protozoan, GO: 0030183: B cell differentiation, GO: 0001772: immunological synapse. In addition, these genes were also annotated into two immune-related KEGG pathways (hsa04662: B cell receptor signaling pathway and hsa05340: primary immunodeficiency). As discussed above, 220 unique genes from the 337 gene pairs play important roles in autoimmune regulation and brain development.

### Table 2 | The number of gene pairs in eight sections for PD cases and controls.

| Interval | PD_case |
|----------|---------|
| [-1, -0.75) | 53,560 (88.04287%) |
| [-0.75, -0.5) | 7,274 (11.95713%) |
| [-0.5, -0.25) | 6,062 (84.95924%) |
| [-0.25, 0) | 4,277 (6.21640%) |
| [0, 0.25) | 9 (0.01308%) |
| [0.25, 0.5) | 2 (0.00291%) |
| [0.5, 0.75) | 16,983 (0.01308%) |
| [0.75, 1) | 314 (0.00291%) |

### Table 3 | Nine gene pairs were shared between RA and PD.

| Gene1 | Gene2 | RA | PD |
|-------|-------|----|----|
| Case_corr | Control_corr | Case_corr | Control_corr |
| ARHGAPI7 | TSHZ3 | 0.382 | −0.561 |
| CDK20 | GRB2 | 0.481 | −0.484 |
| CDK20 | PLEKHF2 | 0.458 | −0.342 |
| CRY1 | GRB2 | 0.599 | −0.252 |
| ETV1 | GRB2 | 0.388 | −0.473 |
| GATA3 | GRB2 | 0.279 | −0.544 |
| GRB2 | LCA5 | 0.505 | −0.36 |
| GRB2 | PVR | 0.692 | −0.32 |
| GRB2 | TSHZ3 | 0.441 | −0.616 |

| Control_corr | Case_corr | Control_corr | Case_corr |
| 0.384 | −0.497 | 0.517 | −0.492 |
| 0.492 | −0.408 | 0.552 | −0.264 |
| 0.63 | −0.237 | 0.59 | −0.64 |
| 0.469 | −0.41 | 0.698 | −0.114 |
| 0.397 | −0.588 |
DISCUSSION

In summary, we calculated the PCC of each gene pair and compared the distribution of the global PCC between cases and controls. We found a similar distribution between cases and controls. Nevertheless, we still found 1,893 gene pairs in RA and 196,311 gene pairs in PD whose co-methylation relationship changed a great deal. Moreover, there were 337 gene pairs that changed across no less than three intervals, and nine gene pairs with an absolute value of difference > 0.8, shared between RA and PD. By Gene Ontology and KEGG pathway analysis, we found 220 unique genes from the 337 gene pairs that play important roles in autoimmune regulation and brain development. There are therefore potential pathogenic and genetic mechanisms shared between RA and PD.

In our study, to assess the reliability of the results and the repeatability of the co-methylation relationships, and to determine how many samples are required for a repeat experiment, we randomly selected five groups of genes with 100 genes in each gene group without returning to the samples. We randomly selected samples from each group (the sample size ranged from 2 to 150) and calculated the PCC of each gene pair. To examine the repeatability, we measured the similarity between replicates of the same operator by selecting the same number of samples from the remaining samples and recalculating the PCC. For each of the five groups, the PCC between the first experiment and the second replicate is shown in Figures 2A (RA cases) and 2B (RA controls). In both RA cases and controls, we found that the PCC increased with increasing sample size. Most of the PCCs were > 0.9 when the sample size was > 150. In other words, the PCC is repeatable when the sample size is relatively large.

AUTHOR CONTRIBUTIONS

SH and RF wrote the main manuscript text. SH, GT, and LX analyzed the data. SH, HP, and FK discussed and improved the article. SH, FK, and YJ conceived and designed the experiments.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Grant No. 91746113). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.01001/full#supplementary-material

Supplementary File S1 | 1,893 gene pairs crossing no less than three intervals with large changes of the co-methylation relationship from RA controls to cases.

Supplementary File S2 | 196,311 gene pairs crossing no less than three intervals with large changes of the co-methylation relationship from PD controls to cases.

Supplementary File S3 | 337 gene pairs shared between RA and PD.

REFERENCES

Adkar, S. S., Brunger, J. M., Willard, V. P., Wu, C. L., Gersbach, C. A., Guilak, F., et al. (2017). Genome engineering for personalized arthritis therapeutics. Trends Mol. Med. 23, 917–931. doi: 10.1016/j.molmed.2017.08.002

Aleyd, E., Al, M., Tuk, C. W., van der Laken, C. J., and van Egmond, M. (2016). IgA complexes in plasma and synovial fluid of patients with rheumatoid arthritis induce neutrophil extracellular traps via FcαRI. J. Immunol. 197, 4552–4559. doi: 10.1049/jimmunol.1502353

Antony, P. M., Diedrich, N. J., Krüger, R., and Balling, R. (2013). The hallmarks of Parkinson’s disease. FEBS J. 280, 5981–5993. doi: 10.1111/febs.12335

Choudhary, N., Bhatt, L. K., and Prabhavalkar, K. S. (2018). Experimental animal models for rheumatoid arthritis. Immunopharmacol. Immunotoxicol. 40, 1–8. doi: 10.1080/08923973.2018.1434793
Chuang, Y. H., Paul, K. C., Bronstein, J. M., Bordelon, Y., Horvath, S., and Ritz, B. (2017). Parkinson’s disease is associated with DNA methylation levels in human blood and saliva. Genome Med. 9:76. doi: 10.1186/s13037-017-0466-5

Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., et al. (2018). MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. Nucleic Acids Res. 46, D371–D374. doi: 10.1093/nar/gks1025

Davegårdh, C., García-Calzón, S., Bacos, K., and Ling, C. (2018). DNA methylation in the pathogenesis of type 2 diabetes in humans. Mol. Metab. 14, 12–25. doi: 10.1016/j.molmet.2018.01.022

Goldman, S. M. (2014). Environmental toxins and Parkinson’s disease. Annu. Rev. Pharmacol. Toxicol. 54, 141–164. doi: 10.1146/annurev-pharmtox-011613-135937

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., et al. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat. Neurosci. 19, 48–54. doi: 10.1038/nn.4182

Karlson, E. W., and Deane, K. (2012). Environmental and gene-environment interactions and risk of rheumatoid arthritis. Rheum. Dis. Clin. North Am. 38, 405–426. doi: 10.1016/j.rdc.2012.04.002

Klareskog, L., Catrina, A. I., and Paget, S. (2009). Rheumatoid arthritis. Lancet 373, 659–672. doi: 10.1016/S0140-6736(09)60008-8

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. 31, 142–147. doi: 10.1038/nbt.2487

Liu, Z., and Page, M. (2018). A novel gene and pathway-level subtyping analysis scheme to understand biological mechanisms in complex disease: a case study in rheumatoid arthritis. Genomics. doi: 10.1016/j.ygeno.2018.02.012. [Epub ahead of print].

Meehan, R. R., Thomson, J. P., Lentini, A., Nestor, C. E., and Penningts, S. (2018). DNA methylation as a genomic marker of exposure to stress and environmental agents. Curr. Opin. Chem. Biol. 45, 48–56. doi: 10.1016/j.copcb.2018.02.006

Mizoguchi, F., Slowikowski, K., Wei, K., Marshall, J. L., Rao, D. A., Chang, S. K., et al. (2018). Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. Nat. Commun. 9:789. doi: 10.1038/s41467-018-02892-y

Nair, N., Wilson, A. G., and Barton, A. (2017). DNA methylation as a marker of response in rheumatoid arthritis. Pharmacogenomics 18, 1323–1332. doi: 10.2217/pgs-2016-0195

Rao, S., Chiu, T. P., Kribelbauer, J. F., Mann, R. S., Bussemaker, H. J., and Rohs, R. (2018). Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. Epigenet. Chromatin 11:6. doi: 10.1186/s13072-018-0174-4

Scott, D. L., Wolfe, F., and Huizinga, T. W. (2010). Rheumatoid arthritis. Lancet 376, 1094–1108. doi: 10.1016/S0140-6736(10)60826-4

Soulaidopoulos, S., Nikiphorou, E., Dimitroulas, T., and Kitas, G. D. (2018). The role of statins in disease modification and cardiovascular risk in rheumatoid arthritis. Front. Med. 5:24. doi: 10.3389/fmed.2018.00024

Viatte, S., Plant, D., and Raychaudhuri S. (2013). Genetics and epigenetics of rheumatoid arthritis. Nat. Rev. Rheumatol. 9, 141–153. doi: 10.1038/nrrheum.2012.237

Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., et al. (2017). RAID v2.0: an updated resource of RNA-associated interactions across organisms. Nucleic Acids Res. 45, D115–D118. doi: 10.1093/nar/gkw1052

Zhou, M., Zhang, Z., Zhao, H., Bao, S., Cheng, L., Sun, J., et al. (2018b). An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. Mol. Neurobiol. 55, 3684–3697. doi: 10.1007/s12035-017-0572-9

Zhou, M., Zhao, H., Wang, X., Sun, J., and Su, J. (2018a). Analysis of long noncoding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer’s disease. Brief Bioinform. doi: 10.1093/bib/bby021. [Epub ahead of print].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a conflict of interest.