Genome-wide association study identifies genetic susceptibility loci about DNA repair activated by oxidative stress

DongYeon Nam

Department of Bioinformatics and Life Science, Soongsil University, Seoul, Korea.

E mail: eodyddmstn@gmail.com

Introduction

DNA damage is a relatively common event in cell life and can lead to mutation, cancer, and the death of cells and organisms [1]. The accumulation of DNA damage caused by oxidative stress has been linked to aging and to the etiology of numerous age-related diseases [2]. Thus, the DNA repair process is a very important factor in cell life. However, as this process regulates multiple downstream processes and cellular functions, there is a lack study that can identify significant loci due to high cost of conducting research experimentally.

the stress-activated protein kinase, c-Jun N-terminal kinase (JNK), phosphorylates SIRT6 on serine 10 in response to oxidative stress [2]. SIRT6 is activated by oxidative stress to promote DNA double-strand break (DSB) repair [2].

DUSP14, Dual Specificity Phosphatase 14, dephosphorylates and inactivates JNK pathway [3]. Thus, DUSP14 has important role in regulating the JNK pathway in DNA repair in one of the earliest steps. And TADA2A, Transcriptional Adapter 2-Alpha, plays a role in chromatin remodeling. Chromatin remodeling is also necessary part of the DNA repair process [4]. But these two genes (DUSP14, TADA2A) are not well studied for DNA repair.

In this study, we presented results of SNPs associated with transcription of DUSP14 and TADA2A. The result suggests not only genetic basis of regulating mechanism of DNA repair but also uninvestigated insight as a novel genetic marker.

Methods & Material

Human genome databases

The Geuvadis RNA-seq data was utilized to identify expression quantitative trait loci (eQTLs). These include mRNA sequence data on lymphoblastoid cell lines from the 1000 Genome Project [5] and investigate 5 populations British (n = 94), Finns (n = 95), Toscani (n = 93), CEPH (n = 91) and Yoruba (n = 89). Yoruba population was excluded in order to avoid false positive associations that produced by heterogeneous genetic background from continental differences. Gene expression of samples was calculated as sum of reads per kilobase per million mapped reads (RPKM) for transcripts of all gene. We excluded the sex chromosome in analysis and genotypes with missing rate upper 0.05 or minor allele frequency under 0.05, Hardy-Weinberg equilibrium test under $1x10^{-6}$ were removed for quality control. 5,916,563 SNPs were remained and conducted for final analysis.

Genome-wide association Analysis

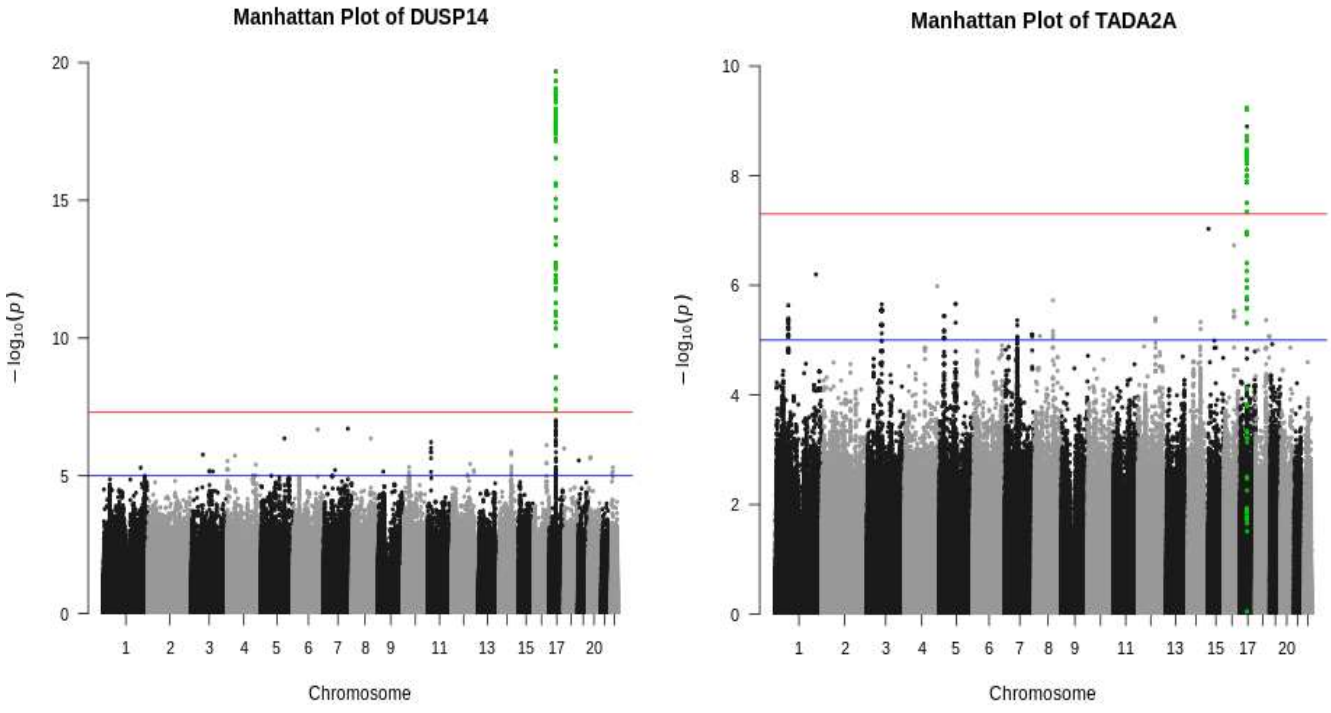
Standard linear regression was implemented to identify expression quantitative trait loci (eQTLs) of DUSP14, TADA2A. PLINK v1.90 which is a whole genome association analysis toolset was used for all statistical analyses [6]. PLINK v1.90 saves the location of the SNPs as GRCh37/hg19. UCSC Lift Over [7] was used to match the current popular location version GRCh38/hg38. Genome-wide significance threshold (5×10^{-8}) was employed to evaluate multiple tests. Linkage

Disequilibrium (LD) blocks were made by HaploView[8]. Manhattan plots were drawn using qqman package in R program [9] and Scatter plot was designed by R plot function. HaploReg v4.1 was employed to determine histone ChIP-seq and regulatory motifs changes. ChIA-PET analysis (In GSE33664 – K562 specific EP) [10], identified the promoter enhancer interaction and found that the eQTL had regulation effect on the corresponding region.

Results

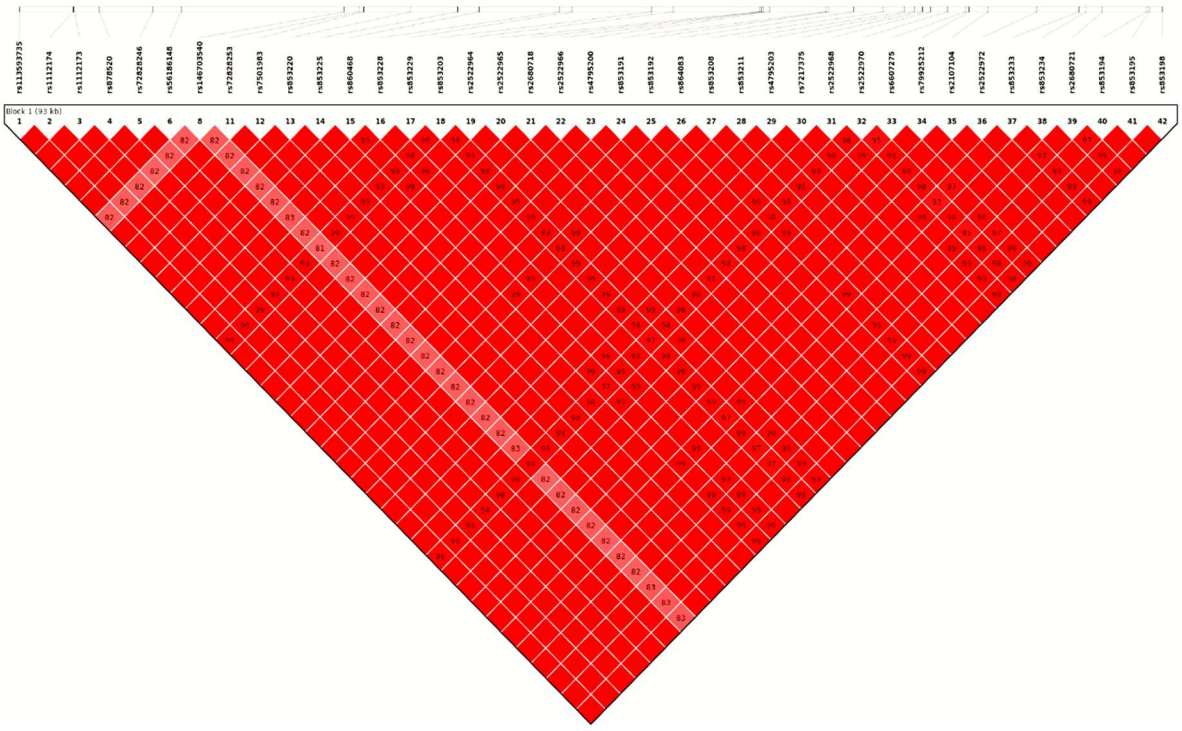
DUSP14 gene is located on chr17:37,489,891-37,513,498 (GRCh38/hg38). TADA2A gene is located on chr17:37,407,224-37,465,655 (GRCh38/hg38). Genome-wide association analysis revealed 96 eQTLs associated with expression of DUSP14 (Supplementary Table 1).

55 eQTLs were identified to be associated with expression of TADA2A (Supplementary Table 2). The eQTLs found in the DUSP14 were Cis-eQTL. Because eQTLs are located from the front of about 50kb to the back of 90kb of the gene. Similarly, eQTLs identified in TADA2A were located form the front of about 30kb to the back of 37kb, indicating Cis-eQTL. Two genes shared 54 eQTLs. So, these shared eQTLs were also Cis-eQTLs. These eQTLs were detected in Manhattan plot (Figure. 1) and confirmed to be in strong linkage in HaploView. Linkage Disequilibrium blocks consisted of 2 blocks (Figure. 2). Through eQTL analysis and LD analysis, 2 eQTLs (rs853195, rs865483) were extracted from each LD blocks, which were observed to be the promoter region of DUSP14 (Table 1). Pearson correlation test was conducted between the expression values of the two genes. As a result, it was confirmed that there was a correlation (Figure. 3).

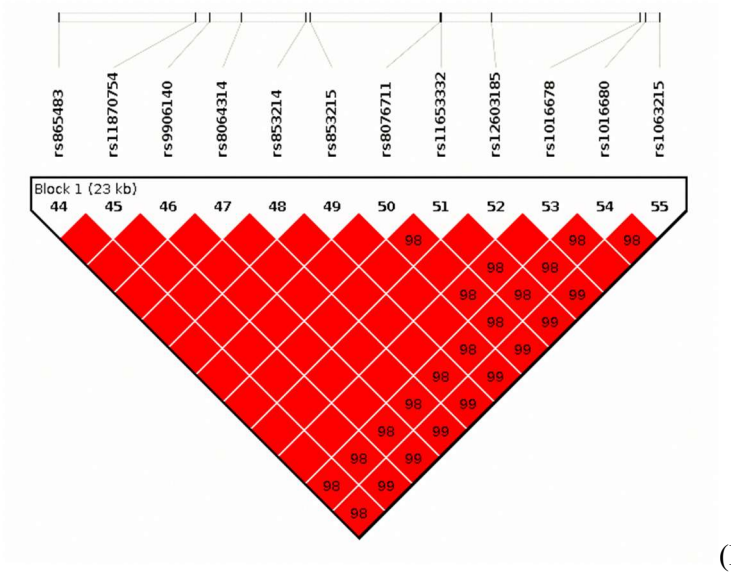


(A) (B)

Figure.1 Manhattan plots for genome-wide association with (A) DUSP14 and (B) TADA2A. The blue line records a threshold for SNPs ($P = 1 \times 10^{-5}$) and red line indicates a significance threshold ($P = 5 \times 10^{-8}$). The green dots are the ones shown in (B) TADA2A that satisfied significance threshold ($P = 5 \times 10^{-8}$) in (A) DUSP14.



(A)



(B)

Figure. 2 Linkage Disequilibrium blocks identified the association between the RNA expression of TADA2A and the base position. LD was checked with the cutoff of R square score ($r^2 > 0.8$). Single LD block represented that the region was likely to be inherited to the next generation dependently. (A) LD block 1 chr17:37,398,591-37,489,236 (GRCh38/hg38), (B) LD block 2 chr17:37,491,071-37,514,833 (GRCh38/hg38)

Table. 1 SNPs that has been significantly identified simultaneously in the Genome-wide association analysis of the two genes. (SNP: Variant identifier, Position: base-pair (GRCh38/hg38), Allele: Minor allele, BETA: Regression coefficient)

SNP	Gene	Position	Allele	BETA	P-value
rs853195	TADA2A	17:37,488,149	A/C	0.579	4.71E-09
	DUSP14	17:37,488,149	A/C	1.385	5.27E-19
rs865483	TADA2A	17:37,491,071	C/A	0.550	3.15E-08
	DUSP14	17:37,491,071	C/A	1.375	1.30E-18

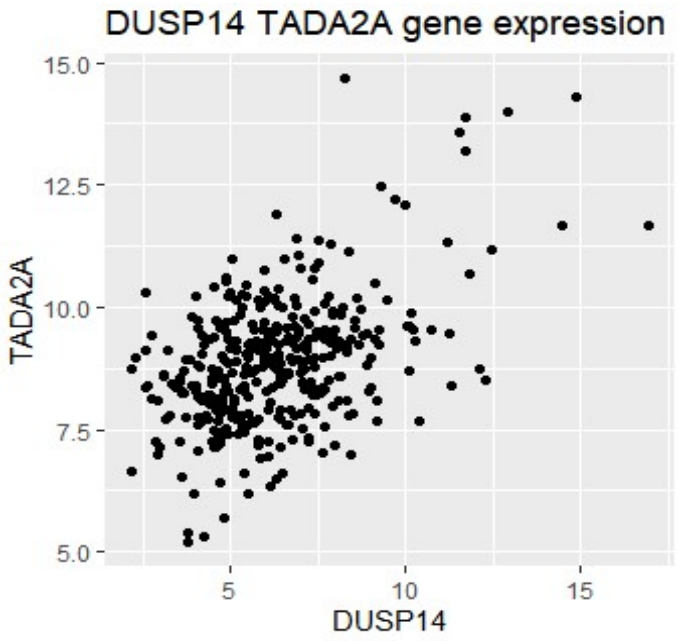


Figure. 3 Scatter Plot of the expression values of DUSP14 and TADA2A. (alternative hypothesis: true correlation is not equal to 0; 95 percent confidence interval: 0.3919 ~ 0.5496; sample estimates: 0.4745; p-value < 2.2 x 10^{-16})

Discussion

In current study, RNA-seq data of 373 individuals on lymphoblastoid cell was analyzed and detected that expression quantitative trait loci (chr17:37,487,939 to chr17:37,514,833) were positioned across TADA2A and DUSP14 while forming two linkage disequilibrium blocks. Rs853195 is located in LD block 1 and rs865483 is located in LD block 2. rs853195(chr17:37,488,149) and rs865483(chr17:37,491,071) are regions identified by histone Chip-Seq (H3K4me3, H3K9ac) in GM12878 Lymphoblastoid Cells and many other groups, too (HaploReg v4.1). Two histone marks on histone H3, lysine 4 trimethylation (H3K4me3) and lysine 9 acetylation (H3K9ac), co-localize on active gene promoters and are associated with active transcription [11]. Thus, rs853195 and rs865483 are indicated as important genetic marker which might be function as promoter and regulate the expression of TADA2A and DUSP14. Moreover, in ChIA-PET analysis (In GSE33664 – K562 specific EP), chr17:37,487,410-37,491,446 were detected to be in the promoter region [10].

GH17J037486, Gene Hancer identifier for the two genes in GeneCard [12], shown the Transcription Factor Binding Site in chr17:37,486,683-37,494,837. YY1 is one of the TFs that binds to the site. YY1 (Yin Yang 1) is a transcriptional repressor protein in humans that is encoded by the YY1 gene [13-14]. In Haploreg v4.1, it was denoted that allelic change in rs853195 reduced the binding affinity of YY1. This indicates the possibility that YY1 be less bound to binding site, and then DUSP14's transcription repression is not carried out normally. This evidence suggests that this locus involving rs853195 is functionally important regulating mechanism than other loci.

According to PathCards [15], DUSP14 is found to be related to JNK pathway which belong to the mitogen-activated protein kinase family. DUSP14 is involved in the inactivation of MAP (mitogen-activated protein) kinase by dephosphorylation [3].

JNK phosphorylates SIRT6 on serine 10 in response to double-strand breaks (DSBs) or other DNA damage, and this step is required for efficient repair of DSBs [2]. JNK plays an important role in DNA repair in one of the earliest steps. DUSP14 expression is not properly controlled and the process of phosphating serine is not performed normally, which can cause problems that DNA repair is not properly carried out and accumulated. To allow repair of double-strand breaks in DNA, the chromatin must be remodeled. TADA2A is a component of the ATAC complex, which has histone acetyltransferase activity on histones H3 and H4 and plays a role in chromatin remodeling. Acetylation of histone H4 plays a critical role in directing changes both in chromatin organization and in promoting recruitment of DSB repair proteins to sites of DNA damage [16].

Furthermore, it has not been investigated exactly how TADA2A functionally relates to DNA repair. But it is identified that the expression of DUSP14 and TADA2A are positively correlated (Figure3.). The expression of DUSP14 is not suppressed properly by the eQTL(rs853195), which may cause problems with the JNK pathway, which affects the important initial stage of DNA repair. Therefore, it would be worth further investigating the relationship between the corresponding eQTL and the expression of DUSP14 and TADA2A.

Reference

[1] Aziz Sancar, Laura A. Lindsey-Boltz (2004). Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. AR of Biochemistry. Vol 73:39-85

[2] Michael Van Meter (2016). JNK Phosphorylates SIRT6 to Stimulate DNA Double-Strand Break Repair in Response to Oxidative Stress by Recruiting PARP1 to DNA Breaks. Cell Reports. Vol 16(10): 2641-2650

[3] Jain Ha, Sayeon Cho (2019). Phosphorylation Dynamics of JNK Signaling: Effects of Dual-Specificity Phosphatases (DUSPs) on the JNK Pathway. IJMS. Vol 20(24)

[4] Liu B, Yip RK, Zhou Z (2012). "Chromatin remodeling, DNA damage repair and aging". *Curr. Genomics*. Vol **13** (7): 533–47.

[5] Siva, Nayanah. "1000 Genomes project." (2008): 256-256.

[6] Shaun Purcell, Benjamin Neale (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Vol 81(3): 559-575

[7] Maximilian Haeussler, Ann S Zweig (2019). The UCSC Genome Browser database: 2019 update. *ucleic Acids Research*, Vol 47 (D1-08): D853–D858

[8] Barrett, Jeffrey C., et al. "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics 21.2 (2005): 263-265.

[9] Stephen D, Turner (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv 005165

[10] Jonas Paulsen, Einar A. Rodland (2014). A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions Nucleic Acids Research. Vol 42(18): e143

[11] Leah A. Gates, Jiejun Shi (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. Journal of Biological Chemistry. Vol. 292: 14456-14472

[12] Safran, Marilyn, et al. "GeneCards Version 3: the human gene integrator." Database 2010 (2010).

[13] Thiaville MM, Kim J (2011). "Oncogenic potential of yin yang 1 mediated through control of imprinted genes". Critical Reviews in Oncogenesis. Vol **16** (3–4): 199–209.

[14] Zhu W, Lossie AC, Camper SA, Gumucio DL (1994). "Chromosomal localization of the transcription factor YY1 in the mouse and human". Mammalian Genome. Vol **5** (4): 234–6.

[15] Belinky, Frida, et al. "PathCards: multi-source consolidation of human biological pathways." *Database* 2015 (2015).

[16] Surbhi Dhar, Ozge Gursoy-Yuzugullu (2017). "The tale of a tail: histone H4 acetylation and the repair of DNA break". The Royal Society Publishing. Vol 372(1731): 1471-2970