Identifying expression quantitative trait loci for pro-inflammatory response triggered by TIFA with genome-wide association study

Harim Chun, 20152629, School of Systems Biomedical Science, Soongsil University

I. Introduction

When a pathogen invades our body, a series of responses occurs such as apoptosis, inflammatory response or phagocytosis. Especially, in an inflammatory response, NOD1 and TIFA play an important role [4]. They trigger nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kappa B) which is a protein regulating multiple innate and adaptive immune functions [1]. TIFA and NOD1 have an independent role in inflammatory response when the gram-negative bacteria infect the cell. Especially TIFA can detect the proliferation of the bacteria in the cytosol and stimulate the activation of the NF-kB in a late step [4].

TIFA gene which encodes a protein named TIFA is in chromosome 4. Certain variation including SNPs would affect the mRNA expression and alter the regulatory motif. Due to these changes, TIFA would be predicted to be malfunctioned or overexpressed which can lead to a certain disease. Therefore, it is needed to be discovered which SNPs are associated with the TIFA gene and which effect would be made in accordance with SNPs. The current study was aimed to identify the expression quantitative trait loci (eQTLs) for the gene TIFA which might regulate the TIFA and its function.

II. Material & Methods

2.1. Material

This study was conducted to identify eQTLs with 373 Europeans' expression data of the genes encoding TIFA in Lymphoblastoid cell lines which were examined by the Geuvadis RNA-sequencing project [10]. Phenotype data includes gene expression data which was calculated by the reads per kilobase per million mapped reads (RPKM) for all transcripts from the genes according to the 1000 Genomes Project. In this data, the Base Position (BP) was assigned based on the Genome Reference Sequence (GRCh37/hg19) and 22 chromosomes, except the sex chromosomes, were used. Also, genotype data contains information about SNPs and individual genotypes. The data with Minor allele frequency under 5% and Hardy Weinberg Equilibrium (HWE) threshold under 1×10^{-6} were removed during the quality control and 5,913,563 SNPs were used for the overall study.

2.2. Methods

Most of the analysis was operated with PLINK ver. 1.90. Linear regression analysis in accordance with the minor allele counts was executed to find eQTLs of TIFA and the threshold value for the p-value was set to 5×10^{-8} with PLINK. Linkage Disequilibrium (LD) blocks constructed based on the algorithm of Gabriel et al [7]. and by the D' and r^2 value with the Haploview (Broad Institute). R programming and `qqman` which is one of the packages in R were used to visualize the various statistical data. Manhattan plots, boxplot about the quantity of the gene expression by the genotype and the bar plot of the allele counts were plotted by the R program.

Certain SNPs were identified whether they are in the gene (Causative SNPs) or outside of the gene

(Linked SNPs) with web crawling like Python programming referring to the NCBI dbSNP database. Each SNPs' IDs were extracted from the association file (ALL_SNP.chr4.TIFA.assoc.linear) and used for the running python code. (Fig. 1). The identified eQTLs' regulatory function to the gene, especially TIFA, were inspected using the Chip-Seq data from the HaploReg v4.1(Broad Institute) and selected if its regulatory chromatin states exist on the Lymphoblastoid Cells. The eQTLs' tissue-specific gene expression, exclusively lymphoblastoid tissue, were examined with GTEx data. Gene (NCBI), Gene Cards (Crown Human Genome Center), and LNCipedia v5.2 were used for searching the function of the various genes. Interactions between the genes by the enhancers and promoters were investigated by the Gene Hancer in the UCSC Genome Browser.

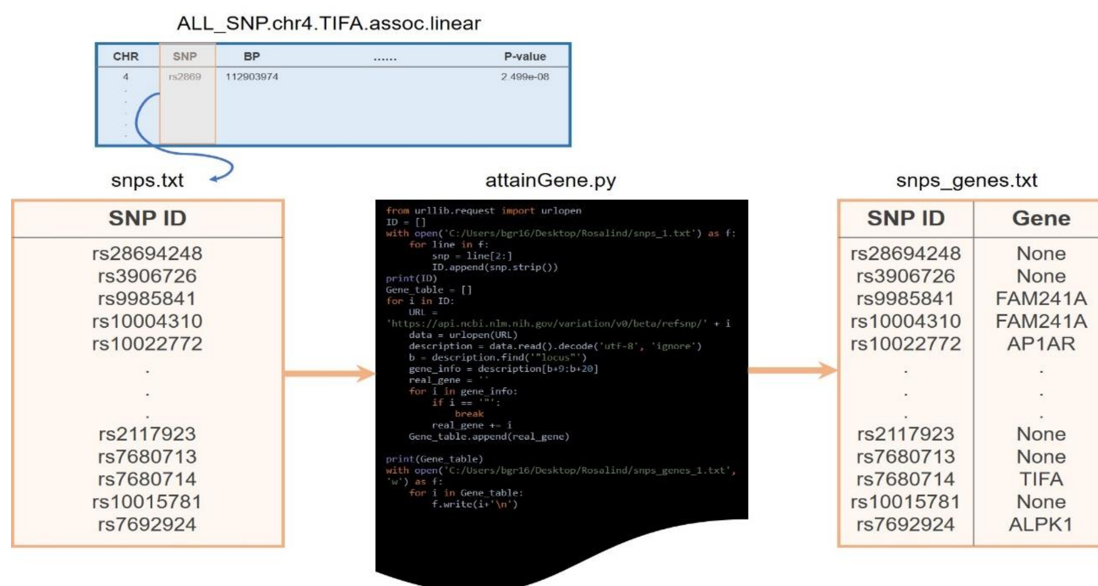

Figure 1. Python Programming Scheme of attaining gene name from the dbSNP. When the text file containing SNPs' ID, which is extracted from the association file were input, attainGene.py would be executed and access to the https://api.ncbi.nlm.nih.gov/variation/v0/beta/refsnp/"SNPid" according to each SNPid. After the execution, each SNPs would obtain the information whether they are in the gene or not and saved into snps_genes.txt file.

III. Results

The linear regression analysis with the gene TIFA and the all SNPs' gene expression disclosed 221 eQTLs and were saved into the file "ALL_SNP.chr4.TIFA.assoc.linear" (p-value < 5×10^{-8}). All eQTLs were located on the chromosomes 4 and were cis-eQTLs within 1Mb from the target gene, TIFA. These cis-eQTLs located in chr4: 112,903,974-113,220,802 (316,828base pairs, Fig. 2. (a)). And each SNPs were assigned to a specific gene if they are in the gene by `attainGene.py` and assigned genes were LOC105377370, FAM241A, AP1AR, TIFA, and ALPK1(supplementary Table. 1). With these eQTLs, 7 LD Blocks were constructed by the Haploview and discovered 9 eQTL signals according to its own p-value including two single SNPs that are not contained to any blocks (Fig. 2. (b), Table. 1, Supplementary Fig. 1). Among them, the sixth block indicates the lowest p-value which is mostly comprised of SNPs in the TIFA gene. Using Haploreg v4.1, 4 eQTL signals out of 9 eQTL signals, including 2 single SNPs, were revealed that might act as an enhancer, a promoter, or a DNase when they are acetylated or methylated in various cell lines including lymphoblastoid cell lines. Four eQTLs were predicted to alter the binding affinity of regulatory motifs when a major allele to minor allele mutation occurred (Supplementary Table. 2, 3). These 4

eQTLs are in the TIFA and ALPK1 gene and have a negative beta coefficient by the additive linear regression model. Among the 221 SNPs, rs701759 was reported on the dbSNP that it can trigger stop gained mutation with A→T substitution. Also, TIFA shows regulatory interactions with enhancers of ALPK1 and FAM241A. 8 SNPs in linked enhancers showed high p-value over 2.0×10^{-4} from the linear regression analysis with the gene TIFA.

Table1.
9 eQTL signals calculated by p-value. 7 eQTL signals are from every 7 Blocks and the others are not contained to any blocks.

Chr	Block	SNP	Position	Minor Allele	Beta	p-value
4	1	rs28694248	112903974	A	-6.038	2.449e-08
4	2	rs10015781	112940112	A	-6.866	1.919e-10
4	3	rs4834195	113018722	G	-5.925	9.112e-09
4	4	rs12504956	113081382	G	-7.038	1.959e-13
4	5	rs7667292	113084978	A	-6.726	1.367e-12
4	6	rs3202899	113196906	C	-8.082	2.783e-17
4	None	rs4834253	113205473	G	-7.336	1.395e-12
4	7	rs4834254	113205543	T	-7.882	6.893e-16
4	None	rs6533603	113220802	C	-7.537	3.709e-13

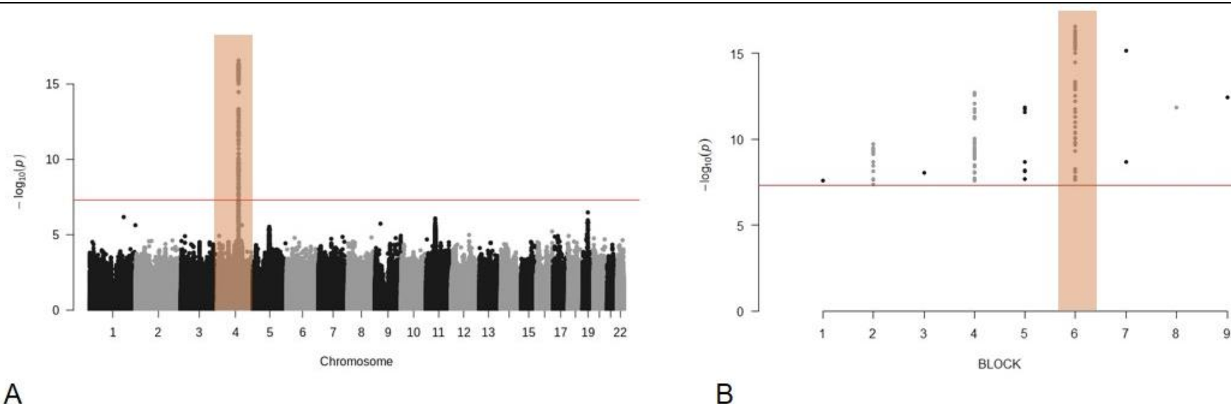

Figure 2. Manhattan plot for genome-wide association of SNPs with mRNA expression of TIFA gene (A). and Each blocks' $-\boldsymbol{log_{10}(p-value)}$ plot (B). The red line indicates the p-value threshold (p-value < $\mathbf{5 \times 10^{-8}}$). Dots on the 8 and 9 Block are single SNP which were not included to any of blocks. And orange colored box indicates the location with low p-value

IV. Discussion

The current study identified 221 SNPs that could regulate the mRNA expression of TIFA genes on chromosome 4. These eQTLs are clustered within 32Kb and considered as cis-eQTLs. Also, these eQTLs were in the 5 genes or not. Most of the genes' function has been known except LOC105377370 which is a lncRNA. Haploreg database indicates that eQTLs in this gene LOC105377370 have no regulatory histone modification and the gene's function was also unknown on the LNCipedia version 5. 2.. With a limited database, it is hard to uncover its function well, therefore, this gene's regulatory function is expected to be discovered through various experiments. Using the Haploview and p-value, this study picked out 4 major signals, rs3202899, rs4834254, rs4834253, rs6533603. One eQTL, rs3202899, might act as an enhancer or a promoter if the T→C mutation occurred and alter zfp105's (Regulatory Motif) binding affinity to be increased. One of the ZFP105's aliases is ZNF35(Zinc Finger Protein) and this protein is a DNA-binding

protein that regulates the gene expression [11]. Also, the beta-coefficient of rs3202899 was -8.082. It implies that when people have a C allele on the chr4:113196906(GRCh37/hg19), they would have more probability to express the TIFA gene than the others. rs4834253 also predicted to act as an enhancer, a promoter or a DNase with the A→G mutation in lymphoblastoid cell lines and alter the binding affinity of Myc_disc2 to be increased which can regulate the apoptosis [8]. Since this eQTL was predicted that have DNase activity, this gene would be more expressed. Also, when Myc_disc2's affinity increased, the cell's apoptotic reaction would be seized, and proliferation would be increased [8]. TIFA gene is usually activated when a pathogen invades into a cell, and when the cell infected, this cell must be eliminated or undergone a proinflammatory reaction. Low Myc_disc2 would increase the apoptosis and eliminate the infected cells. However, when a high level of binding affinity continues due to the mutation, pathogen invaded cell would not be eliminated in our body and proliferated. Therefore, it is considered that many people have an A allele rather than a G allele.

rs4834254 might also act as an enhancer, a promoter or a DNase with the 'C' to 'T' mutation in lymphoblastoid cell lines and alter the binding affinity of nuclear factor kB_disc3(NF-kappaB_disc3) which produces various cytokines and controls cell's survival to be increased [9]. Especially, in the immune response pathway, it can trigger the inflammatory reactions in the cell. When a bacterial pathogen, especially gram-negative bacteria, infects the cell, each NOD1 and ALPK1-TIFA would sense the Pathogen-Associated Molecular Pattern (PAMP), peptidoglycan, Heptose 1,7-Bisphosphate (HBP), and ADP-Heptose, like a sentinel and trigger pro-inflammatory response with NF-kappaB [2][3][5][6]. Usually, the gram-negative bacterium produces HBP [3][4]. When an ALPK1's NTD site binds to HBP, it phosphorylates TIFAs to be oligomerized to form a TIFAsome [1]. NOD1-dependent inflammation is produced at the early step and TIFA-dependent inflammation is triggered at the late step (Fig. 3). The late step inflammation with TIFA is also important as the early step because TIFA can detect the proliferation of the bacteria in the cell [4]. The gram-negative bacterium enters the cell by forming the vacuole, escape the vacuole, and free in the cytosol when the environment is favorable to the pathogen. Then TIFA and ALPK1 notice this event and provoke the late step inflammation in the cell [4]. In this study, the beta-coefficient of rs4834254 was -7.882 which indicates the low gene expression when a major allele to minor allele mutation appeared, but the binding affinity of NF-KappaB_disc3 was expected to be increased. When the gene expression of TIFA is diminished, TIFA would more likely not to detect the proliferation of the pathogen in the cytosol and only NOD1 would detect the early step's bacteria and produce inflammation. When this phenomenon persists, the cell might produce the early step's inflammation mainly and the late step's inflammation or recognition of pathogen's reproduction would be decreased which can lead to damage to the cell. Also, when the binding affinity of NF-KappaB_disc3 increased, inflammation would be increased and might provoke diverse chronic inflammation or autoimmune disease. According to the GeneHancer(GH), TIFA shows regulatory interactions with ALPK1 and FAM241A. rs4834254 is on the promoter site of the TIFA and is connected with ALPK1's enhancer site which implies that TIFA and ALPK1 are related. ALPK1 has a crucial function to the TIFA. ALPK1 can detect the PAMP and phosphorylates the TIFA to form a TIFAsome to provoke the inflammatory action. Even rs6533603 was predicted to act as an enhancer or promoter, this eQTL signal was not in the linked enhancer. Also, SNPs in the linked enhancer were not related to TIFA. Though ALPK1 and TIFA are linked from the functional aspect, there was no linkage from the eQTLs aspect.

rs701759 was predicted to have stop gained mutation on the dbSNP. However, in this data, a T allele was absent on this SNP and this eQTL is predicted to not seize the transcription. It was proved by a bar plot of the allele counts and a boxplot of the expression data (Fig. 4).

In conclusion, this study acknowledged 221 unique eQTLs for the TIFA gene in chromosome 4. Among the eQTLs, 9 eQTLs were selected by the p-value and 4 eQTLs had a regulatory function. These 4 eQTLs might be important signals for a TIFA-dependent immune response. Further studies are demanded to discover TIFA mechanisms.

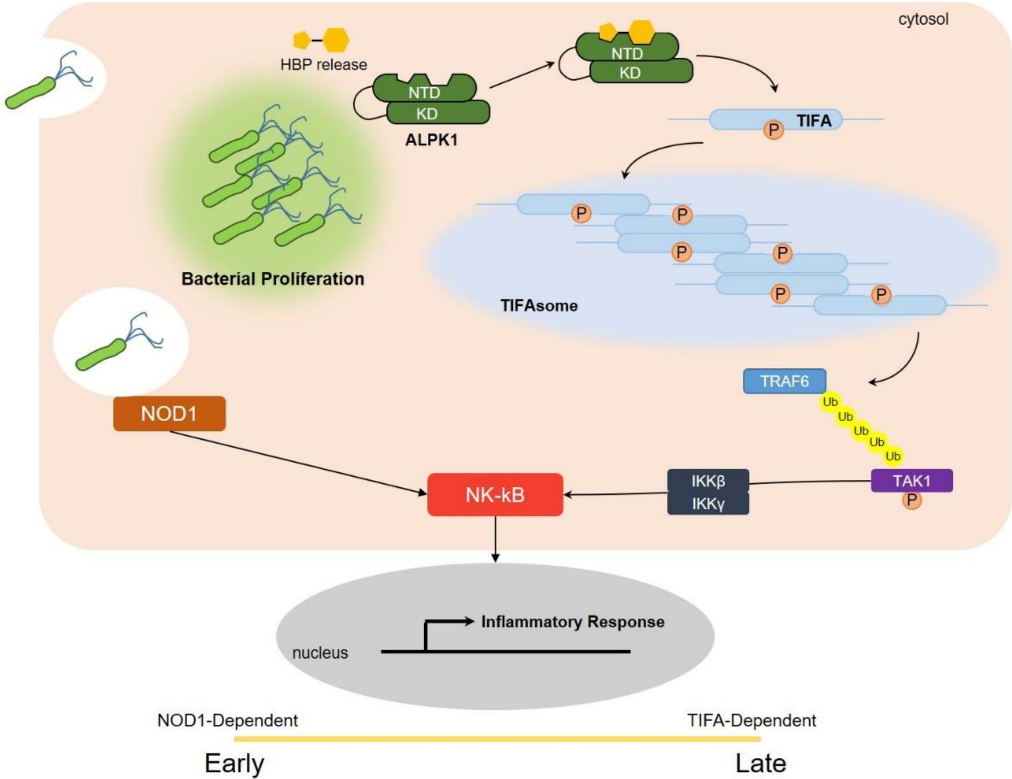

Figure 3. Overall Pathway of the Inflammatory response when the gram-negative bacterium infects the cell. There are two main proteins NOD1 and TIFA that can trigger an inflammatory response with NF-kB. ALPK1 phosphorylates the TIFAs to be oligomerized. The NOD1-Dependent response occurs at the early step and TIFA-dependent response events at the late step.

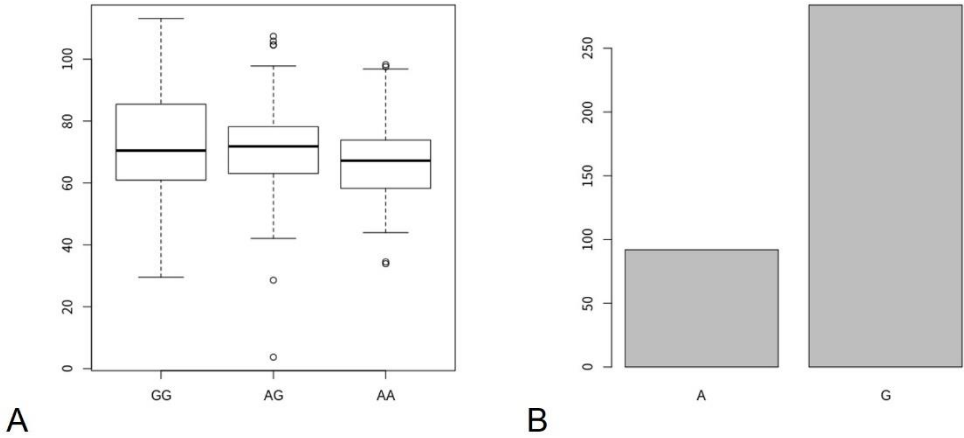

Figure 4. A boxplot was made by the expression value according to the Haplotype of rs701759 (A). and the allele counts of the subjects in the site of rs701759 was used to construct a bar plot (B).

V. Reference

[1] Zhou, P., She, Y., Dong, N., Li, P., He, H., Borio, A., ... & Xu, Y. (2018). Alpha-kinase 1 is a cytosolic innate immune receptor for bacterial ADP-heptose. *Nature*, *561*(7721), 122.

[2] Gaudet, R. G., & Gray-Owen, S. D. (2016). Heptose sounds the alarm: Innate sensing of a bacterial sugar stimulates immunity. PLoS pathogens, 12(9), e1005807.

[3] Zimmermann, S., Pfannkuch, L., Al-Zeer, M. A., Bartfeld, S., Koch, M., Liu, J., ... & Kosma, P. (2017). ALPK1-and TIFA-dependent innate immune response triggered by the helicobacter pylori type IV secretion system. *Cell reports*, *20*(10), 2384-2395.

[4] Gaudet, R. G., Guo, C. X., Molinaro, R., Kottwitz, H., Rohde, J. R., Dangeard, A. S., ... & Gray-Owen, S. D. (2017). Innate recognition of intracellular bacterial growth is driven by the TIFA-dependent cytosolic surveillance pathway. *Cell reports*, *19*(7), 1418-1430.

[5] García-Weber, D., Dangeard, A. S., Cornil, J., Thai, L., Rytter, H., Zamyatina, A., ... & Arrieumerlou, C. (2018). ADP-heptose is a newly identified pathogen-associated molecular pattern of Shigella flexneri. *EMBO reports*, *19*(12).

[6] Pfannkuch, L., Hurwitz, R., Traulsen, J., Sigulla, J., Poeschke, M., Matzner, L., ... & Meyer, T. F. (2019). ADP heptose, a novel pathogen-associated molecular pattern identified in Helicobacter pylori. *The FASEB Journal*, fj-201802555R.

[7] Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... & Liu-Cordero, S. N. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*(5576), 2225-2229.

[8] Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., ... & Young, R. A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*, *151*(1), 56-67.

[9] Liu, T., Zhang, L., Joo, D., & Sun, S. C. (2017). NF-κB signaling in inflammation. *Signal transduction and targeted therapy*, *2*, 17023.

[10] Lappalainen, T., Sammeth, M., Friedländer, M. R., AC‘t Hoen, P., Monlong, J., Rivas, M. A., ... & Barann, M. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature, 501(7468), 506.

[11] Pengue, G., Cannada-Bartoli, P., & Lania, L. (1993). The ZNF35 human zinc finger gene encodes a sequence-specific DNA-binding protein. FEBS letters, 321(2-3), 233-236.