Integrative Transcriptome-wide Analysis of Atopic Dermatitis for *in silico* Drug-repositioning

Jaeseung Song[1], Daeun Kim[1], Sora Lee[1], Junghyun Jung[2,3], Jong Wha J. Joo[2], Wonhee Jang[1*]
[1]*Department of Life Sciences, Dongguk University, Seoul, 04620, Republic of Korea*
[2]*Department of Computer Science and Engineering, Dongguk University-Seoul, 04620, South Korea*

Abstract

Atopic dermatitis (AD) is one of the most common inflammatory skin disorder, which is damaging the quality of life. Multiple genetic risk factors were suggested by previous studies; however, the entire pathogenesis of AD is not perfectly uncovered yet. We conducted a transcriptome-wide association study (TWAS) with FUSION software to estimate both transcriptomic and genomic features of AD. We detected significant associations between 31 expression quantitative loci regions and 25 genes in 9 chromosomes after Bonferroni correction ($P < 0.05/52860$). Our results included *FLG*, a well-known genetic marker for AD with 4 novel associated genes: *P4HA2*, *LINGO4*, *RBM17*, and *RFX5*. Next, five transcriptomic data sets were merged for comparison with TWAS results. Using connectivity map to TWAS and transcriptome data, enrichment scores were calculated and combined into a product score which was used for evaluation of resulted perturbagens. In this study, we propose the first integrative approaches for an AD by combining TWAS and transcriptome meta-analysis. Together, our findings could deepen the understandings of the underlying pathogenic mechanisms of AD and have potential applications in future treatment for AD.

Overview

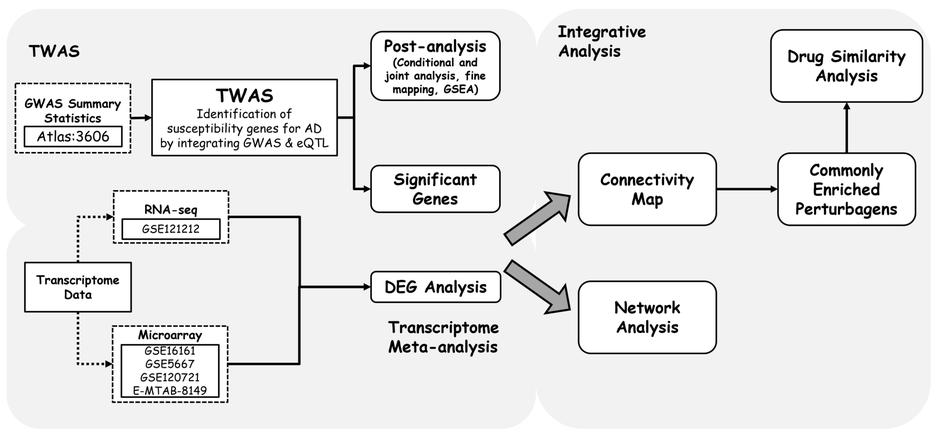

Figure 1. A flowchart of overall analyses. GWAS summary statistics and transcriptome data were retrieved from GWAS Atlas (Atlas ID: 3606) and GEO or ArrayExpress (RNA-seq: GSE121212, microarray: GSE16161, GSE5667, GSE120721, E-MTAB-8149) respectively. A TWAS was conducted using FUSION software to identify significantly associated genes and by using the signals from TWAS, TWAS-GSEA was performed. Five transcriptomic data were merged with SVA R package and the DEGs were obtained. Significantly associated genes from TWAS and the DEGs from meta-analysis were then integrated into a single network. Also, those signatures were analyzed with connectivity map (CMAP) for *in silico* drug-repositioning. Subsequently, resulted chemicals were used for constructing drug network based on mechanisms of action.

Transcriptome-wide Association Study

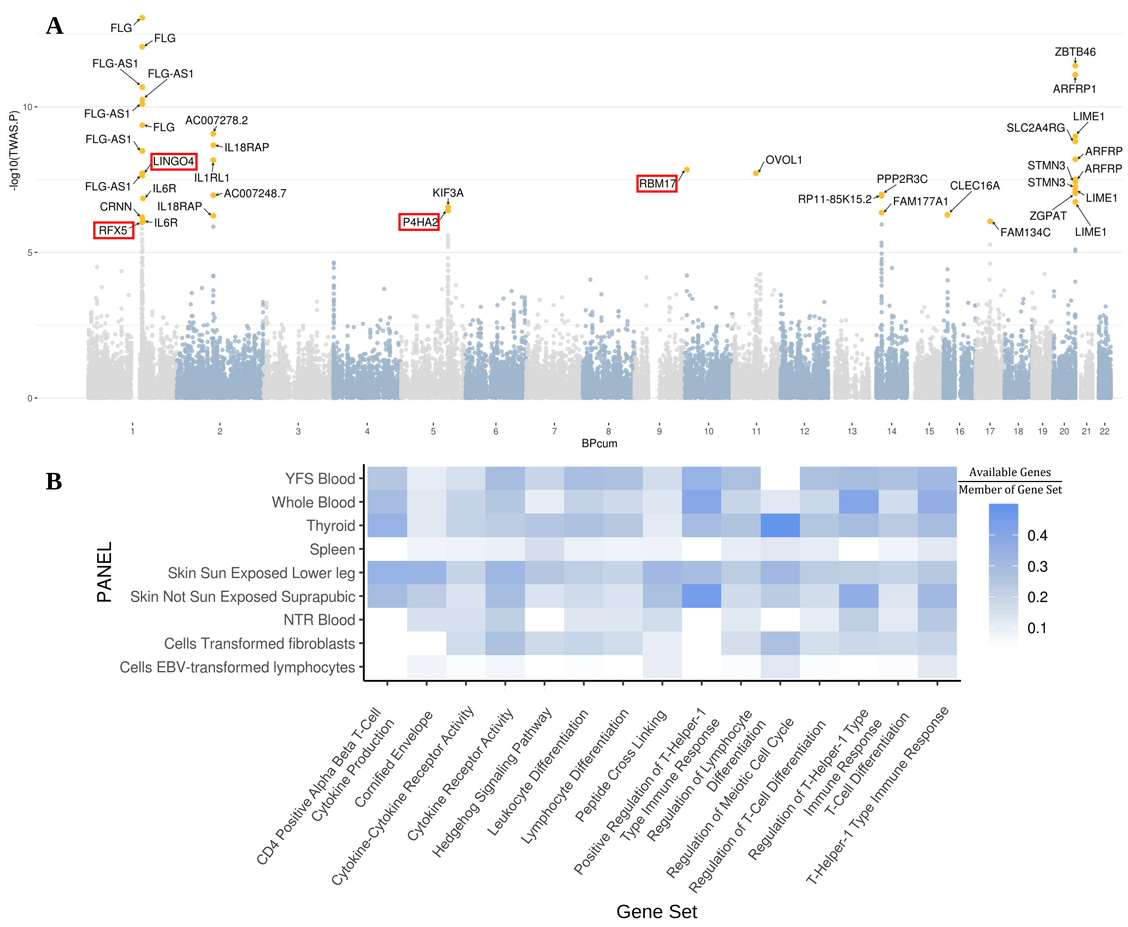

Figure 2. Results of TWAS and TWAS GSEA analysis. A) A Manhattan plot of TWAS results. For the TWAS analysis, pre-computed expression weights from 7 different tissues which assumed to be related with the pathogenesis of AD were used for the analysis. Two blood data from 2 large-scale studies (Netherlands Twin Register, Young Finns Study), and 7 GTEx expression data from different tissue sources (transformed fibroblast, suprapubic skin, lower leg skin, spleen, EBV transformed lymphocytes, thyroid, whole blood) were used as pre-computed expression weights. TWAS was performed with FUSION software and obtained signals from 52860 genes. Among them, 25 genes which remained significant after Bonferroni correction (P-value $< 0.05/52860$) were regarded as significantly associated genes and marked with yellow dots at the plot. Well-known genetic risk factors such as *FLG*, *OVOL1* or *IL6* were resulted as the significantly associated genes for AD in transcriptome-wide scale. Interestingly, there were 4 novel genes: *LINGO4*, *RFX5*, *P4HA2*, and *RBM17* which were not focused on previous studies were identified and they were marked with red boxes at the plot. B) A heatmap of the TWAS-GSEA enrichment. The gene sets which were enriched significantly (FDR < 0.01) in at least one panel were displayed. The color of each cell represents the number of TWAS signal which were included in the gene set divided by the total number of genes included in the gene set. The darker cells indicate the higher proportion of available genes in the gene sets.

Transcriptome Meta-analysis

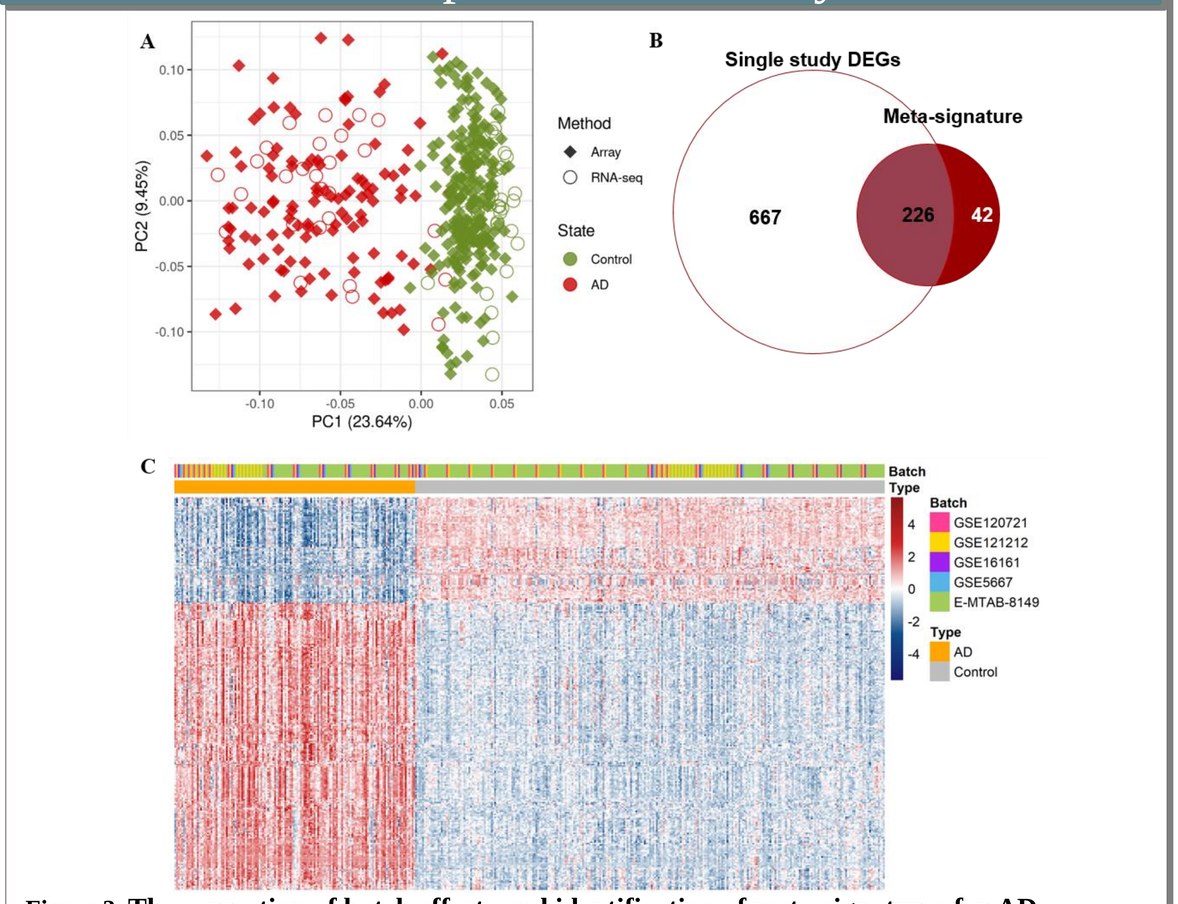

Figure 3. The correction of batch effects and identification of meta-signatures for AD. (A) A scatter plot displaying the PCA results using all genes after the batch effect correction. The shape of the points indicate the samples from each dataset. Green and red color correspond to the healthy control samples and AD samples, respectively. B) A Venn-diagram comparing the DEGs from single-studies with meta-signatures. (C) A heatmap of expression profiles of meta-signatures across the samples.

Identifying Potential Drugs by *in silico* Analysis

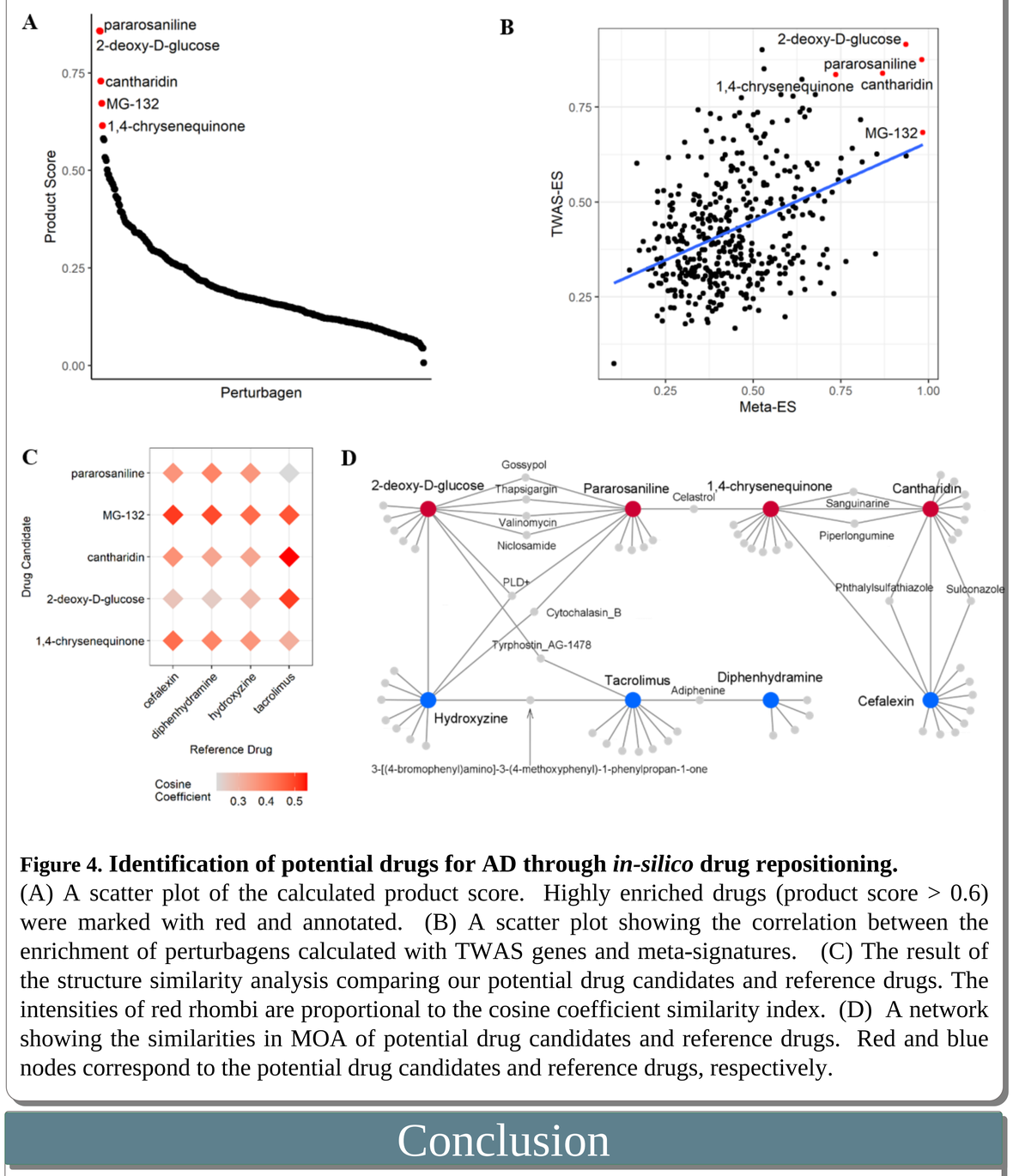

Figure 4. Identification of potential drugs for AD through *in-silico* drug repositioning. (A) A scatter plot of the calculated product score. Highly enriched drugs (product score > 0.6) were marked with red and annotated. (B) A scatter plot showing the correlation between the enrichment of perturbagens calculated with TWAS genes and meta-signatures. (C) The result of the structure similarity analysis comparing our potential drug candidates and reference drugs. The intensities of red rhombi are proportional to the cosine coefficient similarity index. (D) A network showing the similarities in MOA of potential drug candidates and reference drugs. Red and blue nodes correspond to the potential drug candidates and reference drugs, respectively.

Conclusion

- Four significantly associated genes with AD: *LINGO4*, *RFX5*, *P4HA2*, and *RBM17* were newly identified with TWAS and the enrichment of TWAS signals in immune and/or inflammatory responses were observed.
- Five potential drug for AD: pararosaniline, cantharidin, 2-deoxy-D-glucose, MG-132, and 1,4-chrysenequinone were resulted from *in silico* drug-repositioning.

Acknowledgement

This work has supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2018R1D1A1B07042854).

dongguk UNIVERSITY
Developmental biology Laboratory
Dongguk University, Seoul, Korea