BIOMOLECULES & THERAPEUTICS

Original Article
Biomol Ther 30(1), 98-116 (2022)

Discovery of Cellular RhoA Functions by the Integrated Application of Gene Set Enrichment Analysis

Kwang-Hoon Chun*

Gachon Institute of Pharmaceutical Sciences, College of Pharmacy, Gachon University, Incheon 21936, Republic of Korea

Abstract

The small GTPase RhoA has been studied extensively for its role in actin dynamics. In this study, multiple bioinformatics tools were applied cooperatively to the microarray dataset GSE64714 to explore previously unidentified functions of RhoA. Comparative gene expression analysis revealed 545 differentially expressed genes in RhoA-null cells versus controls. Gene set enrichment analysis (GSEA) was conducted with three gene set collections: (1) the hallmark, (2) the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and (3) the Gene Ontology Biological Process. GSEA results showed that RhoA is related strongly to diverse pathways: cell cycle/growth, DNA repair, metabolism, keratinization, response to fungus, and vesicular transport. These functions were verified by heatmap analysis, KEGG pathway diagramming, and direct acyclic graphing. The use of multiple gene set collections restricted the leakage of information extracted. However, gene sets from individual collections are heterogenous in gene element composition, number, and the contextual meaning embraced in names. Indeed, there was a limit to deriving functions with high accuracy and reliability simply from gene set names. The comparison of multiple gene set collections showed that although the gene sets had similar names, the gene elements were extremely heterogeneous. Thus, the type of collection chosen and the analytical context influence the interpretation of GSEA results. Nonetheless, the analyses of multiple collections made it possible to derive robust and consistent function identifications. This study confirmed several well-described roles of RhoA and revealed less explored functions, suggesting future research directions.

Key Words: RhoA, Gene Set Enrichment Analysis (GSEA), GSE64714, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, Hallmark pathway, Gene ontology

INTRODUCTION

The Rho family of GTPases is a subfamily of the Ras superfamily and consists of more than 20 members in humans (Haga and Ridley, 2016). Most Rho GTPases modulate cytoskeletal dynamics, cell movement, vesicle trafficking, and cell adhesion to maintain polarity and motility (Ridley, 2006; Hodge and Ridley, 2016). The Rho GTPases work as molecular switches: when activated by numerous extracellular signals, guanine nucleotide exchange factors stimulate the replacement of the inactive GDP moiety with GTP, forming an active GTP–Rho complex, whereas GTPase-activating proteins enhance GTPase activity, leaving the protein in an inactive GDP-bound form (Haga and Ridley, 2016).

The human Rho GTPases are divided into eight subfamilies. The Rho subfamily consists of three members: RhoA, RhoB, and RhoC. RhoA is related to various pathological conditions (Narumiya and Thumkeo, 2018). RhoA and RhoC are upregulated or engaged in various tumors, including gastric cancer, testicular cancer, squamous cell carcinoma, hepatocellular carcinoma, and pancreatic adenocarcinoma (Suwa *et al*., 1998; Clark *et al*., 2000; Simpson *et al*., 2004; Faried *et al*., 2006; Wang *et al*., 2007; Porter *et al*., 2016; Wang *et al*., 2016). Recent reports have identified mutated RhoA in several tumors, such as diffuse gastric cancer and angioimmunoblastic T cell lymphoma (Kakiuchi *et al*., 2014; Sakata-Yanagimoto *et al*., 2014; Wang *et al*., 2014; Yoo *et al*., 2014; Zhou *et al*., 2014; Zhao *et al*., 2015). In contrast, RhoB is generally reported to inhibit oncogenic progression, and its expression is downregulated in cancer cells and tissues (Du and Prendergast, 1999; Chen *et al*., 2000; Adnane *et al*., 2002; Mazieres *et al*., 2004; Sato *et al*., 2007). Dysregulation of the RhoA/ROCK pathway induces oxidative stress and promotes cardiovascular diseases, such as vasospasm, arteriosclerosis, ischemia/

Open Access **https://doi.org/10.4062/biomolther.2021.075**

Received Apr 19, 2021 **Revised** Jun 19, 2021 **Accepted** Jul 9, 2021
Published Online Aug 25, 2021

***Corresponding Author**
E-mail: khchun@gachon.ac.kr
Tel: +82-32-820-4951

reperfusion injury, hypertension, pulmonary hypertension, and heart failure (Satoh *et al*., 2011; Shimokawa *et al*., 2016). Several reports have shown that RhoA can have an ameliorating or aggravating effects in diabetic complications, although most studies have indicated that RhoA/Rho activation aggravates diabetic conditions. Thus, it is urgent to identify RhoA-mediated signaling pathways and develop regulators to modulate pathological conditions.

The molecular details of RhoA activities have only been partially characterized. RhoA can regulate actin cytoskeleton remodeling (Ridley and Hall, 1992), vesicle trafficking in exocytosis or Golgi-to-endoplasmic reticulum transport (Ridley, 2006), and glucose transport (Duong and Chun, 2019). These functions are exerted primarily through downstream effector molecules, such as Rho-kinases (ROCKs) (Leung *et al*., 1995) and Diaphanous-related formin 1 (Watanabe *et al*., 1999). ROCKs activate LIM kinase and inhibit myosin light chain phosphatase to enhance the conversion of monomeric G-actin into polymeric fibrous F-actin (Kimura *et al*., 1996; Maekawa *et al*., 1999). Through the formation of F-actin, RhoA can plays diverse cellular roles by activating transcription regulators, including serum response factor/myocardin related transcription factor A, activator protein 1, Nuclear factor NF-kappa-B, transcriptional coactivator YAP/Tafazzin, β-catenin, and hypoxia inducible factor-1α (Kim *et al*., 2018).

In this study, the Gene Express Omnibus database was searched for RhoA knockdown or knockout systems. One study conducted by Garcia-Mariscal *et al*. (2018) fell into the category (further described in the Materials and Methods). Although the role of RhoA was examined at the genomic level, their study focused on identifying the role of retinol in keratinocyte differentiation, without comprehensive analyses of the other functions. This study aimed to uncover unknown functions by evaluating the role of RhoA in various functional aspects. For this purpose, Gene Set Enrichment Analysis (GSEA) (https://www.gsea-msigdb.org/gsea/index.jsp) was performed using various molecular signatures, and the results were compared. GSEA is a computational method to compute the statistical significance of a defined set of genes in a given experimental environment (Mootha *et al*., 2003; Subramanian *et al*., 2005). In GSEA, a gene set is a group of genes that share a common function or property. The GSEA database officially provides eight major curated gene set collections in the Molecular Signatures Database (MSigDB) (Liberzon *et al*., 2011), which can be chosen according to the analytical purpose. The release of MSigDB (ver. 7.0) has more than 20,000 gene sets as of the time of this study.

Three different gene set collections were applied: the hallmark gene set (Liberzon *et al*., 2015), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway gene set (Kanehisa *et al*., 2017), and the Gene Ontology (GO) Biological Process (BP) gene set (Ashburner *et al*., 2000). As these gene sets have distinct advantages and disadvantages, using them in combination allows their strengths to be leveraged while compensating for their weaknesses. The outcomes from each gene set collection were compared and analyzed to identify novel putative functions of RhoA.

MATERIALS AND METHODS

Microarray dataset

The microarray dataset was obtained by searching the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) using "RhoA" as a keyword. Only one dataset (GSE64714) satisfied the following criteria: (1) RhoA depletion or overexpression was an independent variable, (2) it contained genome-wide RNA expression, and (3) it had three or more samples, and (4) complete microarray raw data were available. GSE64714 is based on the Affymetrix Mouse Genome 430 2.0 array (platform ID: GPL1261) and includes 10 samples of RhoA-null or wild-type keratinocytes from freshly isolated (*in vivo*) or cultured primary mouse keratinocytes (*in vitro*). As the *in vivo* data were prepared in duplicate, further analyses were performed with the triplicate *in vitro* data.

Data preprocessing and quality testing

The original CEL data files were read using the Affy package (Gautier *et al*., 2004) in R software (Version 3.5.2) (R Core Team, 2018). Background correction, normalization, and summarization were performed for all 45,101 probes (45,037 target genes and 64 in-house markers) using the robust multiarray average (RMA) method (Irizarry *et al*., 2003). Principal component analysis (PCA) of this normalized expression data was performed using the "prcomp" function.

Screening of differentially expressed genes

Probes were filtered from the normalized expression dataset with relatively high variance using the "nsFilter" function in genefilter (Gentleman *et al*., 2018) (variance cutoff=0.5), resulting in 18,993 probes selected. The overall distribution of the filtered gene expression data was displayed using an MA plot, in which the mean log2 gene expression from six groups was used as a representative expression level. The *p* value of each gene was determined by *t*-testing using the Limma package (Smyth *et al*., 2005) in R software to obtain the differentially expressed genes (DEGs). The false discovery rate (FDR) was calculated using the Benjamini–Hochberg method to adjust the *p* value. The FDR value versus log2 fold change (FD) of DEGs was displayed using a volcano plot. Affymetrix mouse genome 430 2.0 array annotation data (mouse4302.db) (Carlson, 2016) were used for gene annotation.

Heatmap display

The relative gene expression was visualized using the pheatmap package (Kolde, 2019). Expression values were centered and scaled in the row direction (scale="row").

Gene set enrichment analysis

A total of 37,999 normalized probes with annotated gene symbols were subjected to the lmFit and eBayes functions of the Limma package. Statistics for duplicated genes with identical symbols were combined into mean values, producing 20,563 individual genes. GSEA was performed using the fgsea function in the fgsea package (Sergushichev, 2016), applying the hallmark (50 gene sets), the KEGG pathway (186 gene sets), and the GO BP (7,350 gene sets) gene set collections from MSigDB (Liberzon *et al*., 2015) (version 7.0) individually as signature gene sets with 1000 permutations.

For GO BP, DEGs with adjusted *p* values <0.05 were selected, and GO term enrichment was tested using topGo. Fifty refined and concise "hallmark gene sets" were derived from multiple founder sets. The KEGG pathway database is another collection of manually curated pathway maps. This

database offers a helpful way to represent visualized pathway diagrams of selected gene sets based on the molecular interaction, reaction, and relationship networks. GO is the most widely used database of gene function information. GO provides a hierarchically organized set of terms consisting of three classes: Molecular Function, Cellular Component, and Biological Process. The full description of members included in the gene sets is provided in List S1.

After measuring the gene–phenotype association score for each gene in a dataset, the degree of overrepresentation was calculated and assigned an "enrichment score." For demonstration, several pathways with the highest normalized enrichment scores (NES) were selected and displayed using the ggplot2 package (Wickham, 2016). Enrichment plots were generated using fgsea. The mean gene expression levels in a gene set were displayed in a scatter plot and boxplot using ggplot2.

Visualization of pathways using the KEGG pathway database

Gene lists in a given KEGG pathway ("SNARE interaction in vesicular transport" (KEGG id#: 04130) and "DNA replication" (KEGG id#: 03030)) were entered into the KEGG pathway database (https://www.genome.jp/kegg/pathway.html) to map DEGs in biological pathways. Significantly upregulated or downregulated genes (p<0.05) were colored red or blue. Genes without significance were colored green.

Construction of a directed acyclic graph between the enriched GO BP terms

Gene sets in the GO are organized in hierarchical parent-child relationships with each other. These hierarchical relationships can be visualized using a graph in which terms representing gene sets can be depicted as nodes, and relationships between gene sets are represented as edges. Specifically, this graph has directed parent-child relationships between nodes, but does not have cyclic loop structures internally, so it is called a "directed acyclic graph" or DAG. DAG can efficiently mark the "part of" relationship between gene sets as moving across levels of abstraction. The root node, the highest level of abstraction, contains more general gene members, while leaves are more specific members. The directed acyclic graph (DAG) was constructed from the 51 most enriched GO BP terms (p<0.01 and |NES|>2) using GOView web-based software (Shoop *et al*., 2004) (http://www.webgestalt.org/2017/GOView/) and AEGIS software (Zhu *et al*., 2019). The GO IDs used in GOView analyses were obtained from the MSigDB.

Similarity analyses

The similarity of gene sets was measured using the Jaccard distance. The more similar two sets are, the smaller the Jaccard distance is. The Jaccard distance was obtained by subtracting the Jaccard index from 1 and has a value between 0 and 1. The formula to calculate the Jaccard index is

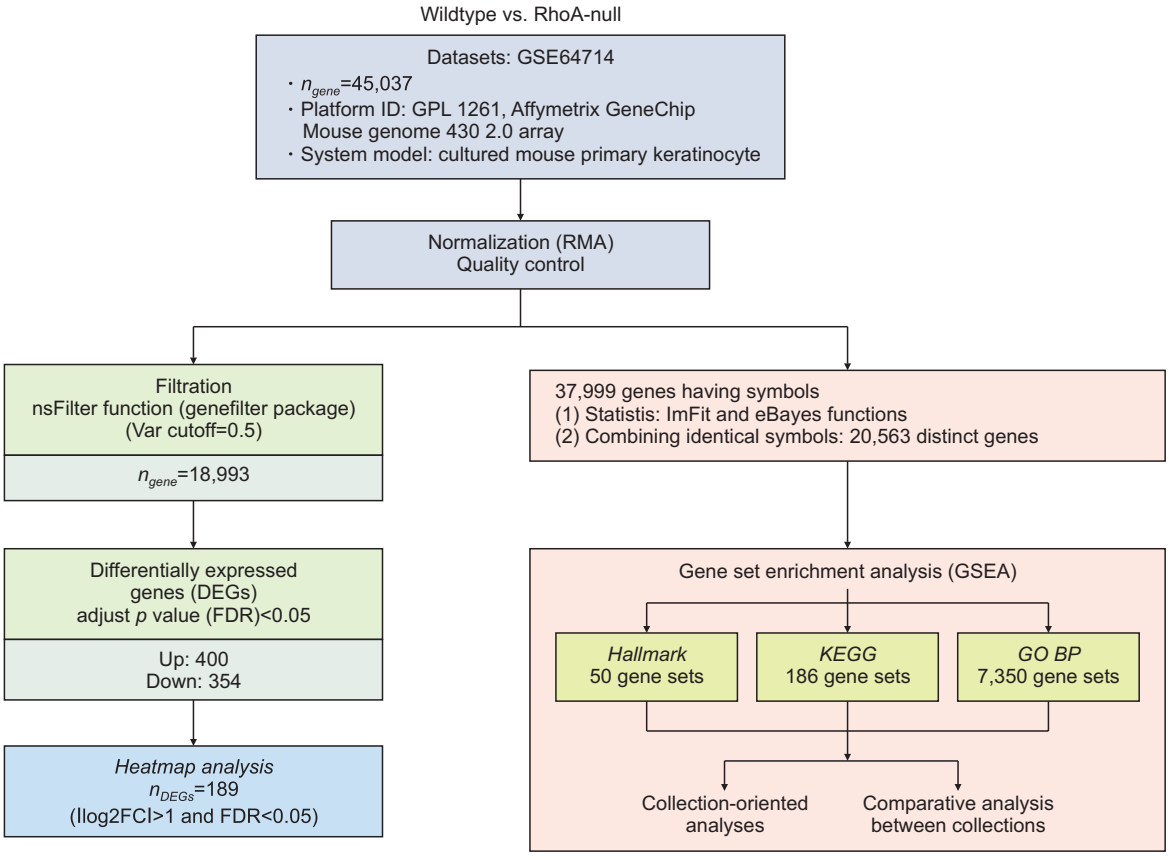


Fig. 1. Study flow diagram. The chart describes the methods used and the features extracted during analyses of the GSE dataset GSE64714.

Jaccard index=(*the number of genes in both gene sets*)/(*the number in either gene set*).

The distances between multiple gene sets were visualized with a correlation plot generated using the corrplot package (Taiyun and Viliam, 2017) in R software.

RESULTS

Analyses of DEGs regulated by RhoA

Microarray data were obtained from the GEO database (GSE64714) to explore RhoA-regulated targets. The *in vitro* data from RhoA-negative mouse keratinocytes contained gene expression information for 45,101 genes. The overall scheme is illustrated in Fig. 1. Gene expression levels were normalized by applying the RMA method (Fig. 2A). The PCA plot showed that the expression profile was clearly differentiated between the RhoA-null and wild-type control data (Fig. 2B), thus indicating that the consequent DEG analyses could be properly performed. The 18,993 genes were obtained by filtering out genes with low expression or high variation, and the quality of data to be used was confirmed using the MA plot (Fig. 2C). The distribution of gene expression values (log2 fold change) with adjusted *p* values (FDR) was presented using a volcano plot (Fig. 2D). The result indicated that RhoA-deficiency upregulated several genes having large differential changes. The subsequent *t*-test produced 754 DEGs with FDR<0.05 (400 upregulated, 354 downregulated), and the 50 most significantly regulated genes are listed in Table 1. The statistical information for 754 probes is provided in Supplementary Table 1. A heatmap of 189 genes with the criterion of log2|FC|>1 and FDR<0.05 is shown in Fig. 2E.

Functional enrichment analysis using three distinct gene set collections

GSEA was performed to analyze the functions of RhoA. In this study, three distinct gene set collections were applied to deduce robust functions and use collection-dependent features: the “hallmark,” the “KEGG pathway,” and the “GO BP” collections. The results are provided in Supplementary Table 2.

FDR values of gene sets were plotted versus normalized enrichment scores (NESs) to evaluate the quality of the GSEA

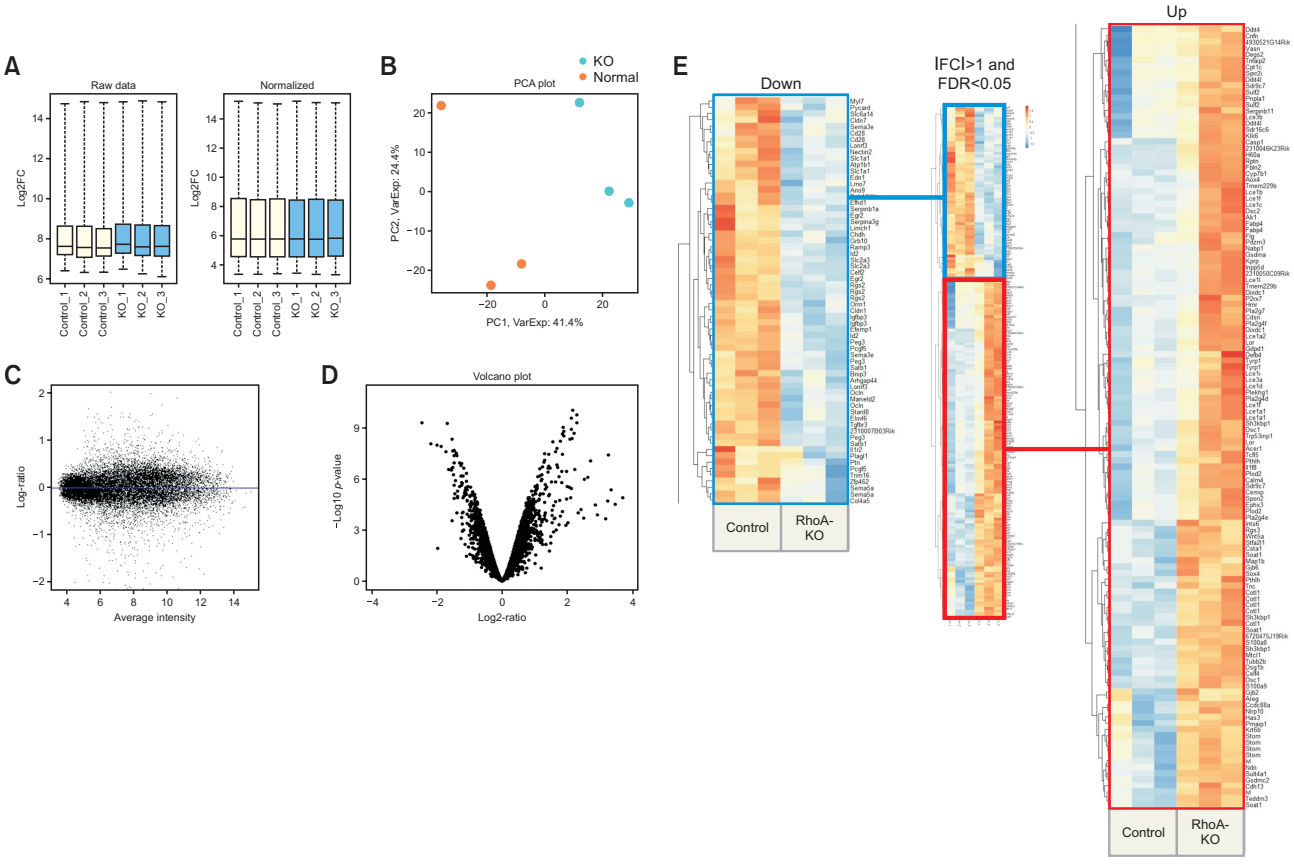


Fig. 2. Preprocessing and identification of DEGs. Datasets of three wild-type samples (control) and three RhoA-null samples (KO) were preprocessed and analyzed for DEGs. (A) Box plots of log2 expression values of the GSE64714 dataset before (left) and after RMA normalization (right). (B) PCA plot representing the differential gene expression patterns of RhoA-null and wild-type samples in two dimensions; x-axis=PC1: PCA component 1; y-axis=PC2: PCA component 2. (C) MA plot for 18,993 genes after filtration. The x-axis represents the average log2 intensity of genes, and the y-axis represents the log2 fold change. (D) Volcano plot for DEGs. DEG, differentially expressed gene. (E) A total of 189 DEGs (log2|FC|>1 and FDR<0.05) were selected and clustered hierarchically based on distances. The row for each cluster was divided into two parts (up- and downregulation) and are displayed using a heatmap.

Table 1. Differentially expressed genes in the RhoA-null condition (50 most significant genes)

Probe ID	Symbol	LogFC	AveExpr	t	*p* value	FDR	B
1431592_a_at	*Sh3kbp1*	2.16	8.61	–13.8	8.61E-11	1.56E-06	14.4
1460337_at	*Sh3kbp1*	2.3	8.48	–13.2	1.64E-10	1.56E-06	13.8
1456203_at	*Dsc1*	2.11	8.6	–12.9	2.61E-10	1.66E-06	13.4
1433924_at	*Peg3*	–2.45	7.35	12.3	4.94E-10	1.75E-06	12.9
1417356_at	*Peg3*	–1.69	6.71	12.3	5.52E-10	1.75E-06	12.8
1437811_x_at	*Cotl1*	2.28	8.39	–12.4	4.89E-10	1.75E-06	12.9
1448756_at	*S100a9*	1.87	10.4	–11.8	9.50E-10	2.58E-06	12.3
1436838_x_at	*Cotl1*	2.12	10.7	–11.7	1.14E-09	2.70E-06	12.2
1417695_a_at	*Soat1*	1.89	8.72	–11.5	1.52E-09	3.22E-06	11.9
1435760_at	*Csta1*	1.84	9.68	–11.2	2.17E-09	3.74E-06	11.6
1454264_at	*2310046K23Rik*	2.2	10.2	–11.2	2.10E-09	3.74E-06	11.6
1442339_at	*Stfa2l1*	1.89	8.12	–11	3.02E-09	4.79E-06	11.3
1433795_at	*Tgfbr3*	–1.54	9.13	10.7	4.46E-09	5.64E-06	10.9
1423494_at	*Teddm3*	2.09	7.78	–10.7	4.39E-09	5.64E-06	11
1421460_at	*Dsc1*	2.13	7	–10.8	4.10E-09	5.64E-06	11
1448397_at	*Gjb6*	1.79	8.13	–10.5	6.02E-09	7.15E-06	10.7
1419394_s_at	*S100a8*	1.56	10.8	–10.4	6.68E-09	7.46E-06	10.6
1425415_a_at	*Slc1a1*	–2.19	6.04	10.3	8.43E-09	8.90E-06	10.4
1437052_s_at	*Slc2a3*	–2	6.66	10.1	1.04E-08	1.03E-05	10.2
1423062_at	*Igfbp3*	–1.87	10.6	9.99	1.26E-08	1.20E-05	9.99
1439878_at	*Ivl*	1.68	9.61	–9.84	1.58E-08	1.43E-05	9.78
1448873_at	*Ocln*	–1.51	9.03	9.64	2.16E-08	1.78E-05	9.49
1453511_at	*2310007B03Rik*	–1.47	5.3	9.65	2.11E-08	1.78E-05	9.51
1458268_s_at	*Igfbp3*	–1.73	9.73	9.55	2.46E-08	1.87E-05	9.37
1436236_x_at	*Cotl1*	1.64	9.74	–9.58	2.37E-08	1.87E-05	9.41
1420401_a_at	*Ramp3*	–1.41	8.16	9.42	3.04E-08	2.17E-05	9.18
1429053_at	*Mtcl1*	1.31	6.49	–9.41	3.09E-08	2.17E-05	9.16
1455519_at	*Dsg1b*	1.42	7.7	–9.31	3.59E-08	2.35E-05	9.02
1422324_a_at	*Pthlh*	1.85	6.25	–9.32	3.55E-08	2.35E-05	9.03
1448745_s_at	*Lor*	3.24	7.94	–9.26	3.91E-08	2.47E-05	8.94
1419248_at	*Rgs2*	–1.38	8.34	9.23	4.11E-08	2.52E-05	8.89
1447830_s_at	*Rgs2*	–1.48	7.23	9.18	4.38E-08	2.60E-05	8.83
1417696_at	*Soat1*	1.31	8.25	–8.8	8.11E-08	4.67E-05	8.26
1420431_at	*Rptn*	2.71	5.77	–8.76	8.65E-08	4.83E-05	8.2
1421606_a_at	*Sult4a1*	1.29	5.86	–8.65	1.04E-07	5.67E-05	8.02
1427682_a_at	*Egr2*	–1.47	5.7	8.58	1.17E-07	6.00E-05	7.91
1429564_at	*Pcgf5*	–1.3	10.1	8.59	1.14E-07	6.00E-05	7.94
1416007_at	*Satb1*	–1.25	8.3	8.54	1.24E-07	6.20E-05	7.86
1454883_at	*Gsdmc2*	1.68	5.98	–8.44	1.48E-07	7.23E-05	7.69
1419247_at	*Rgs2*	–1.27	8.54	8.41	1.54E-07	7.33E-05	7.65
1416002_x_at	*Cotl1*	1.43	8.03	–8.33	1.77E-07	8.21E-05	7.52
1425801_x_at	*Cotl1*	1.23	9.42	–8.31	1.82E-07	8.25E-05	7.49
1451924_a_at	*Edn1*	–1.53	7.79	8.2	2.21E-07	9.14E-05	7.31
1417355_at	*Peg3*	–1.51	6.67	8.21	2.18E-07	9.14E-05	7.33
1451154_a_at	*Celf2*	–1.18	5.55	8.2	2.20E-07	9.14E-05	7.32
1450633_at	*Calm4*	1.69	11.9	–8.22	2.15E-07	9.14E-05	7.34
1417697_at	*Soat1*	1.4	7.26	–8.17	2.31E-07	9.20E-05	7.27
1432269_a_at	*Sh3kbp1*	1.47	7.56	–8.17	2.32E-07	9.20E-05	7.26
1423071_x_at	*6720475J19Rik*	1.15	7.83	–8.12	2.55E-07	9.87E-05	7.18
1455898_x_at	*Slc2a3*	–1.64	5.32	8.1	2.64E-07	1.00E-04	7.14

Genes were sorted in ascending order of the FDR.

results (Fig. 3, left). Gene sets with highly positive or negative NESs had lower FDR values, resulting in bell-shaped symmetric distributions. Parts of significantly regulated gene sets were presented with their NESs (Fig. 3, right).

The GSEA results using the "hallmark" collection showed that the "p53 pathway" was upregulated significantly (FDR<0.1, NES>1.5), whereas "E2F targets," "G2M checkpoint," "Myc targets v1," "mTORC1 signaling," "Myc targets

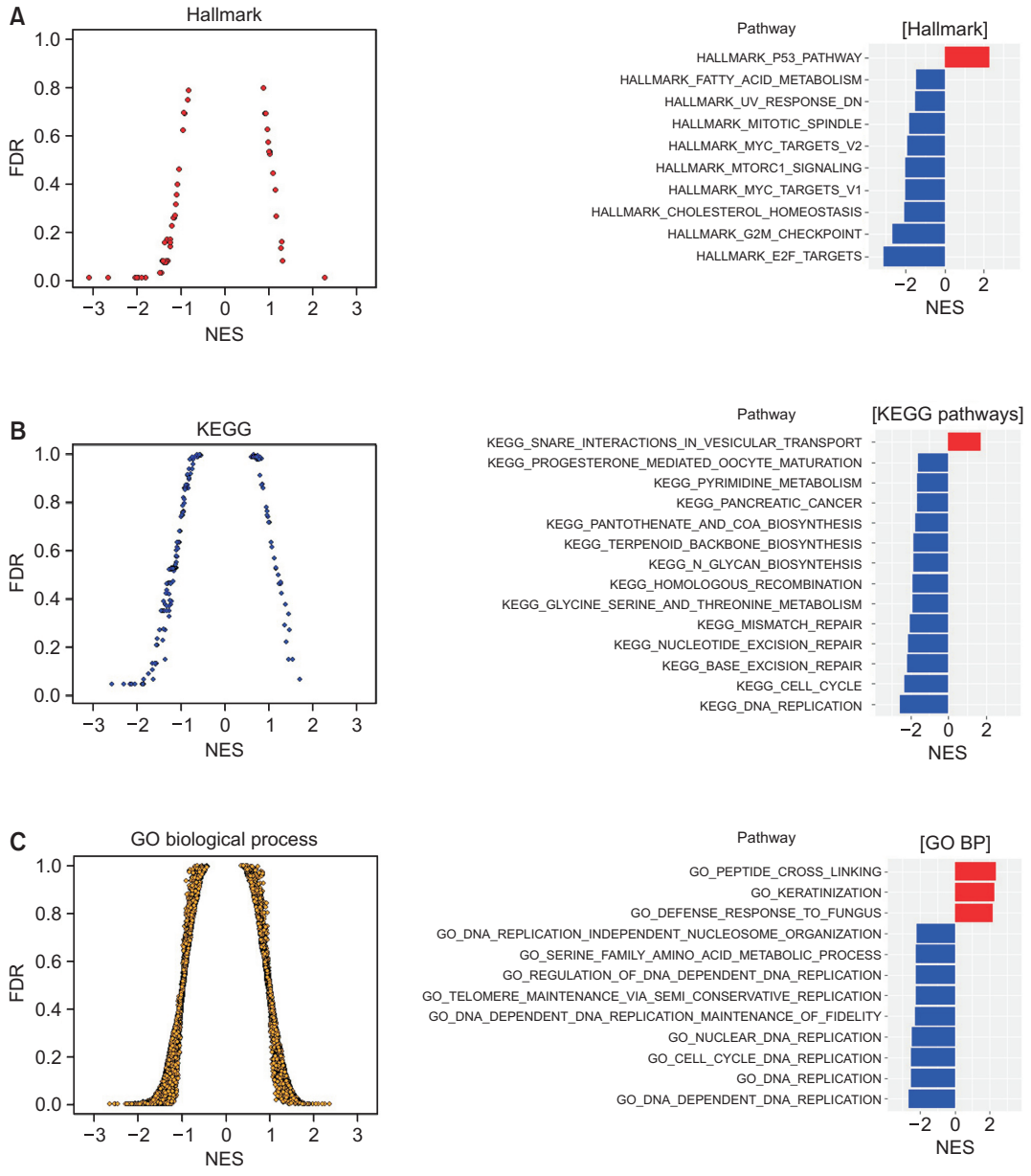


Fig. 3. Quality assessments of three collections in GSEA. GSEA was performed for 20,563 genes using the fgsea package using (A) the "hallmark" (50 gene sets), (B) the "KEGG" (186 gene sets), and (C) the "GO BP" (7,350 gene sets) collections. In each pathway, the FDR of gene sets obtained from GSEA was plotted against the NES (left). Gene sets with a FDR<0.01 were displayed using bar plots (right). Only 12 gene sets of "GO BP" were selected and displayed because of space limitations. Red; positive NES, blue; negative NES.

v2," and "mitotic spindle" were downregulated significantly (FDR<0.1, NES<–1.5) (Fig. 3A). However, the "KEGG pathway" analysis showed "SNARE interactions in vesicular transport" as upregulated (FDR<0.1, NES>1.5). The same analysis showed downregulation of "DNA replication," "cell cycle," "base excision repair," "mismatch repair," "glycine serine and threonine metabolism," "homologous recombination," "N glycan biosynthesis," "terpenoid backbone biosynthesis," "pantothenate and CoA biosynthesis," and "pyrimidine metabolism" (FDR<0.1, NES<–1.5) (Fig. 3B). The "GO BP" pathways provided results of >7000 curated gene sets (Fig. 3C), among which "peptide cross-linking," "keratinization," and "defense response to fungus" were the most positively affected gene sets (NES>2.2). Negatively regulated pathways (NES<–2.2) were "DNA-dependent DNA replication," "DNA replication," "cell cycle DNA replication," "nuclear-dependent DNA replication maintenance of fidelity," "telomere maintenance via semi-conservative replication," "serine family amino acid metabolic process," "DNA replication-independent nucleosome organization," "regulation of DNA-dependent DNA replication," and "chromatin remodeling at centromere."

Enrichment plots of the selected gene sets

The enrichment plots were inspected to visualize the rel-

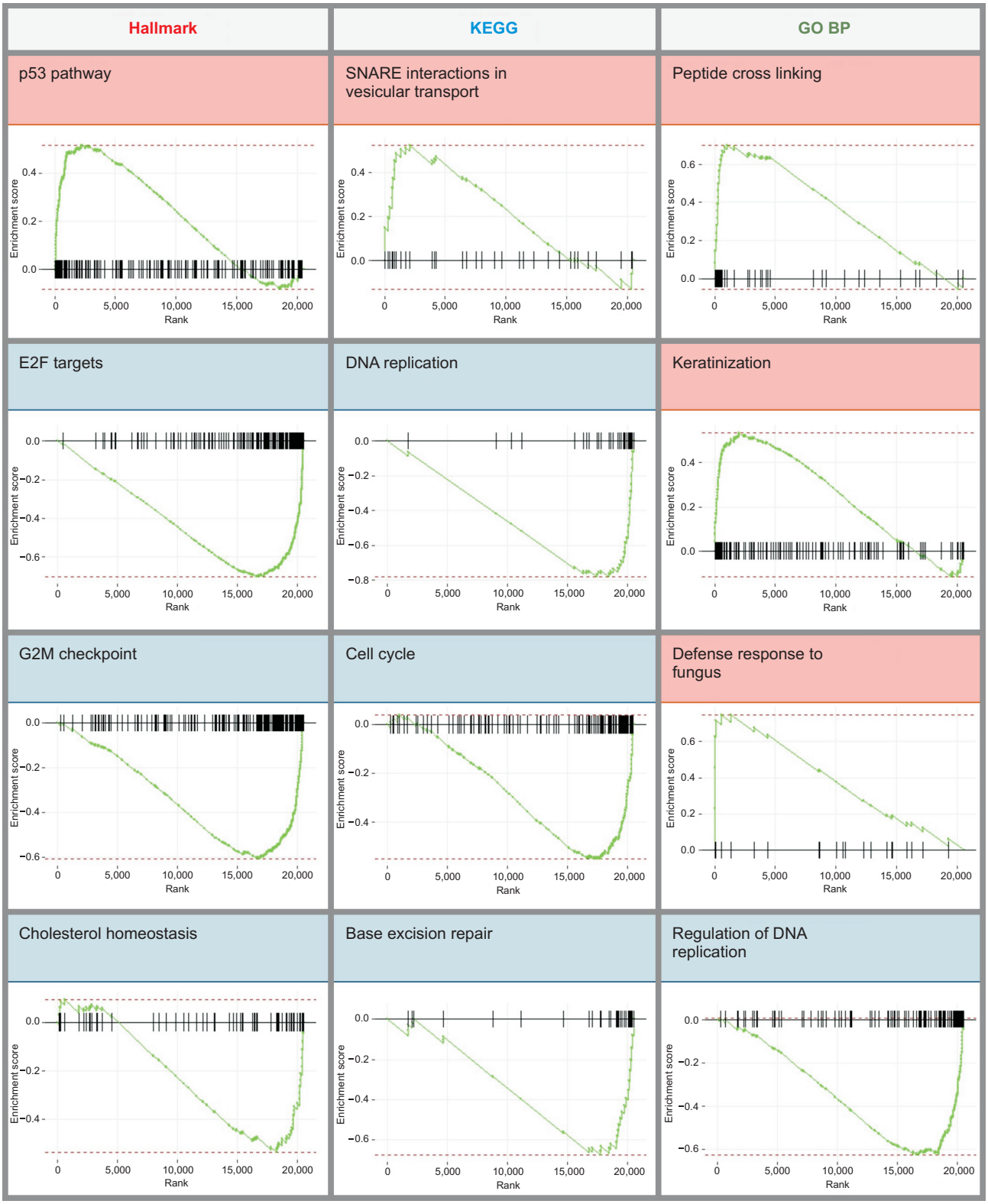


Fig. 4. Enrichment plots in three collections. Enrichment plots for significantly up- or downregulated pathways for the "hallmark," the "KEGG," and the "GO BP" collections.

evance of the selected gene sets to RhoA activity: "p53 pathway" (hallmark), "SNARE interactions in vesicular transport" (KEGG pathway), and "peptide cross-linking," "keratinization," and "defense response to fungus" (GO BP) showed positive NESs (Fig. 4, red). In contrast, "E2F targets," "G2M checkpoint," and "cholesterol homeostasis" (hallmark), "DNA replication," "cell cycle," and "base excision repair" (KEGG pathway), and "regulation of DNA replication" (GO BP) showed negative NESs (Fig. 4, blue). These gene sets with high NES values showed strong asymmetry, indicating significant posi-

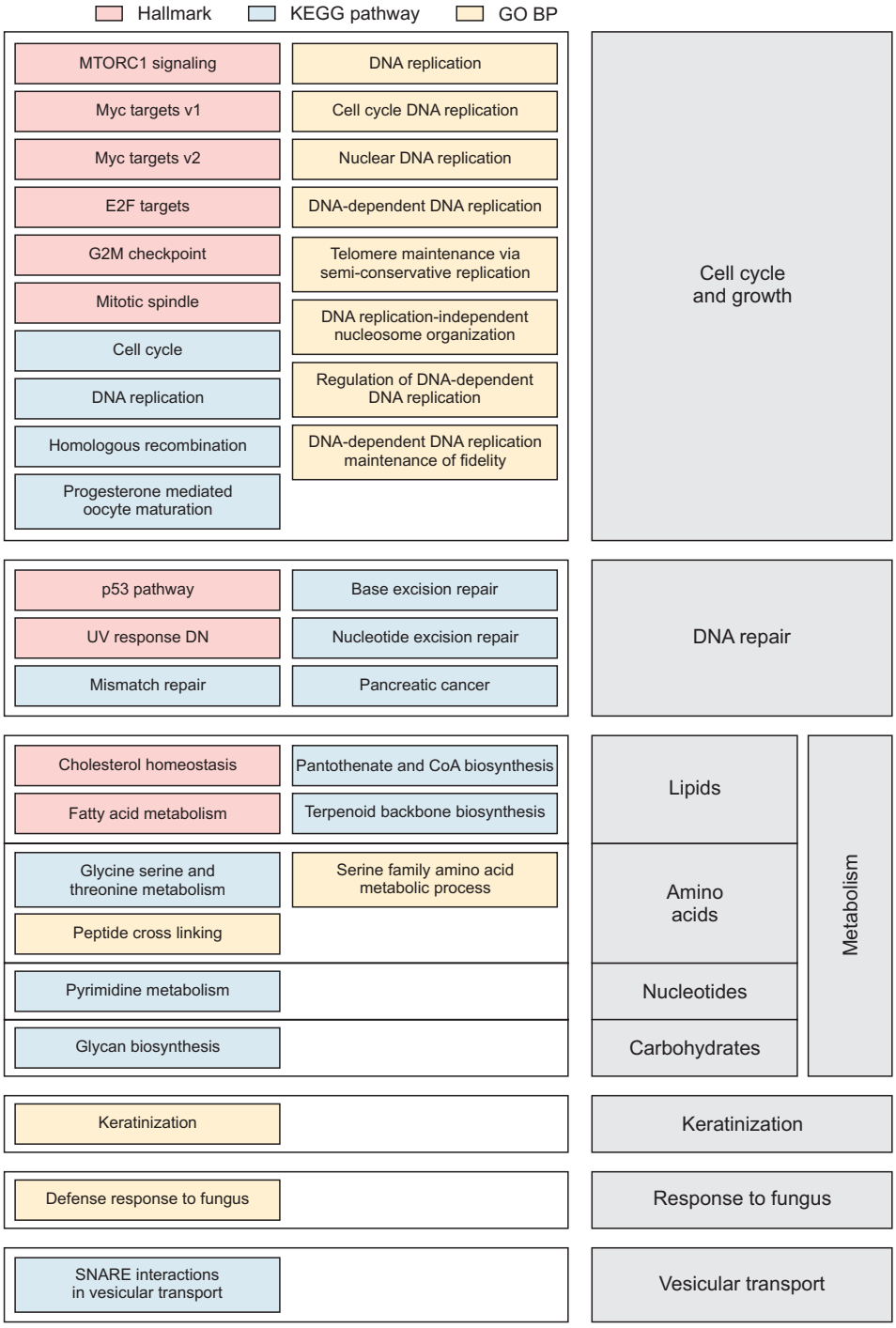


Fig. 5. Overview of RhoA's function inferred from GSEA. Highly scored gene sets from three collections were organized and combined according to their functional similarity (left boxes). Summarized common functions were noted on the right side. Cell cycle and growth-related functions were the most highly scored. Gene sets from the hallmark, KEGG pathway, and GO BP are colored in pink, blue, and yellow, respectively.

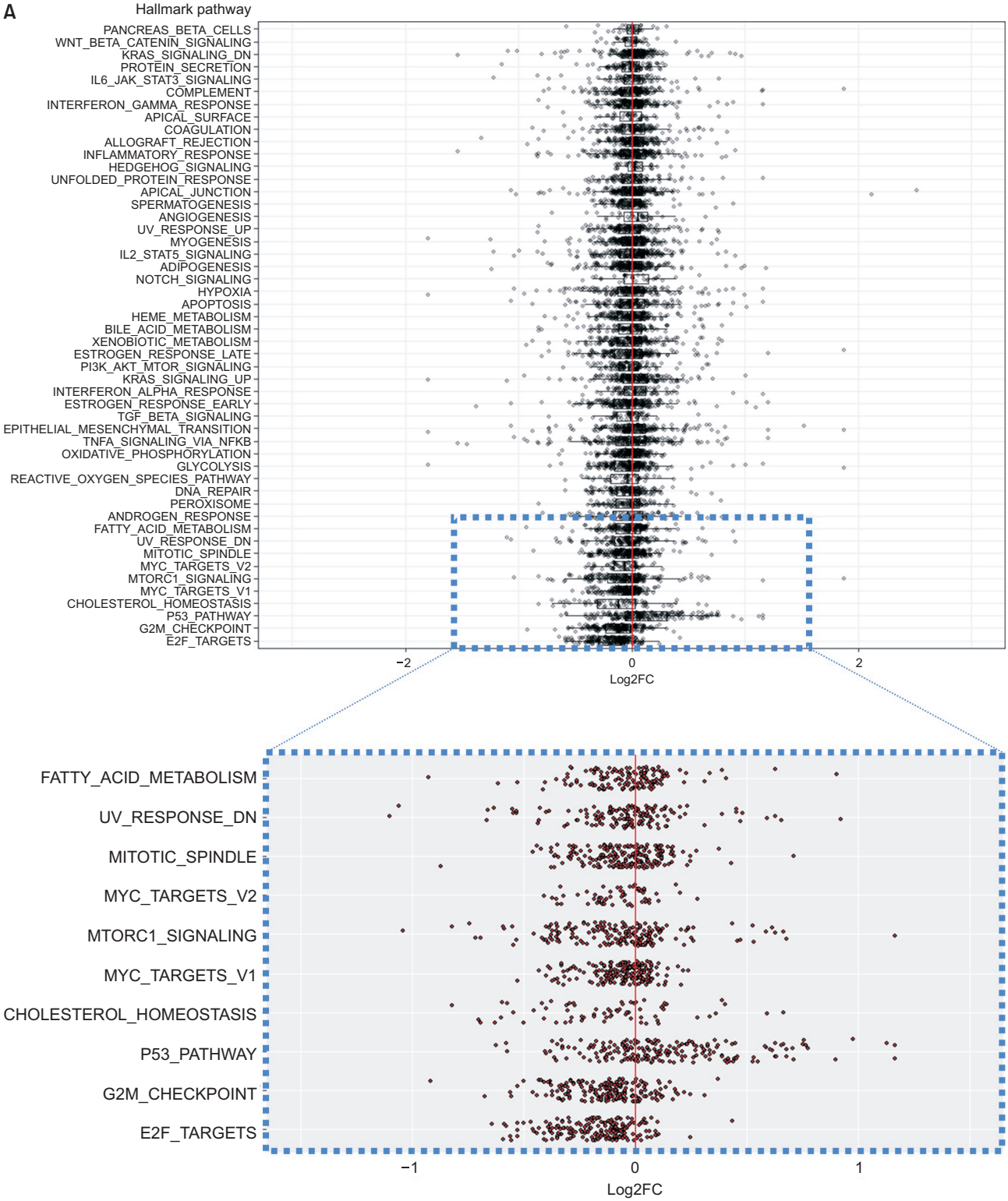


Fig. 6. Evaluation of GSEA results of the "hallmark" pathways. (A) To inspect the performance of GSEA, the distribution of log2FC values of individual genes was displayed. All 50 gene sets of the hallmark pathway collection were examined. Each dot represents a mean log2FC value of an individual gene element from a triplicate sample. Gene sets are sorted along the y-axis by placing the set with the lowest FDR value on the bottom. The 10 most affected gene sets were further enlarged for the detail in the magnification box below. A vertical red line indicates genes showing no change in expression between RhoA-KO and the control. (B) The hierarchical clustering and heatmap display of most upregulated (p53 pathway) or downregulated (E2F targets pathway) genes by RhoA deletion. Gene names are listed on the right side and treatment is marked on the bottom of the colored map. The color bar indicates the log2FC values (RhoA-KO vs. control).

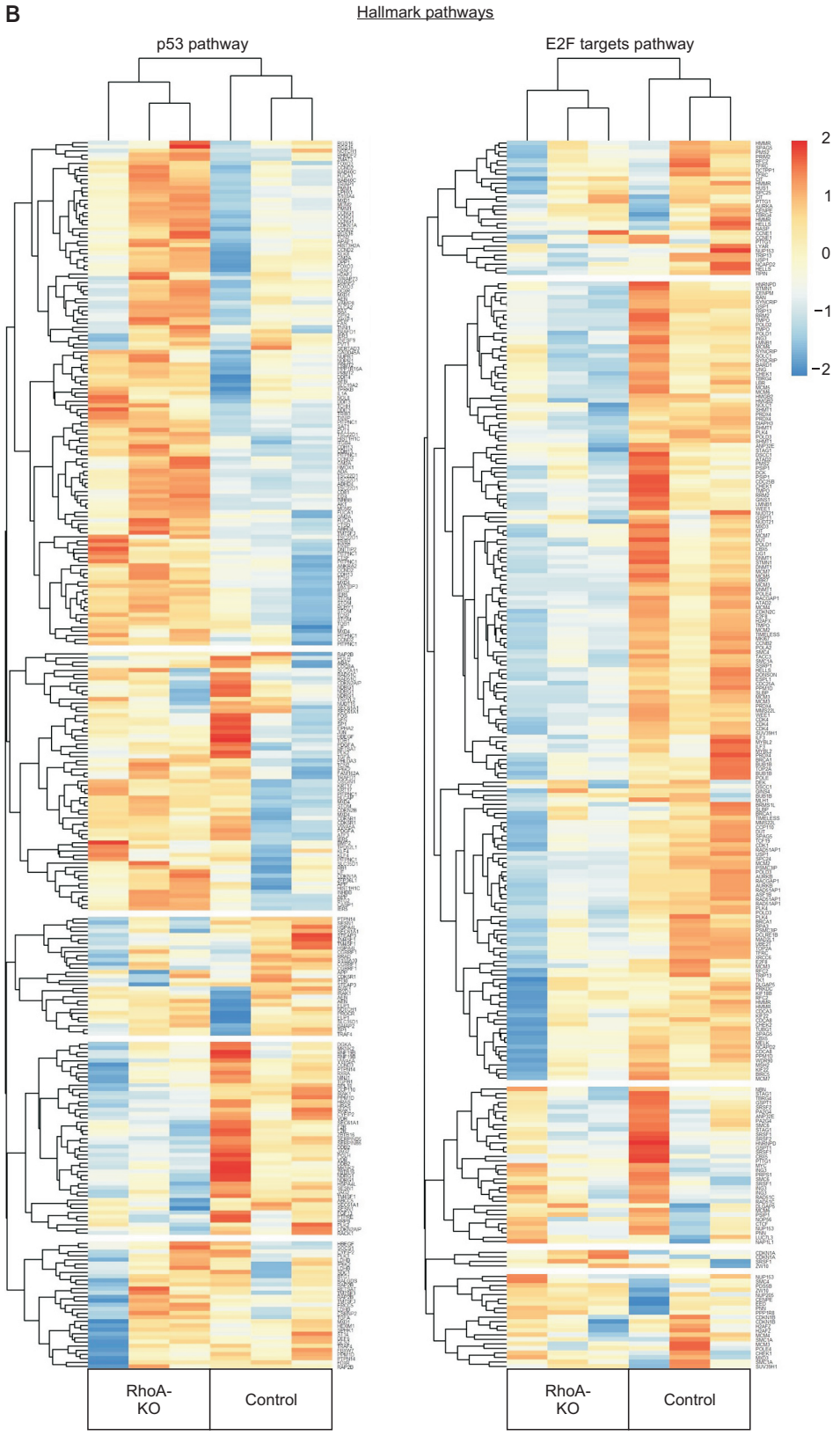


Fig. 6. Continued.

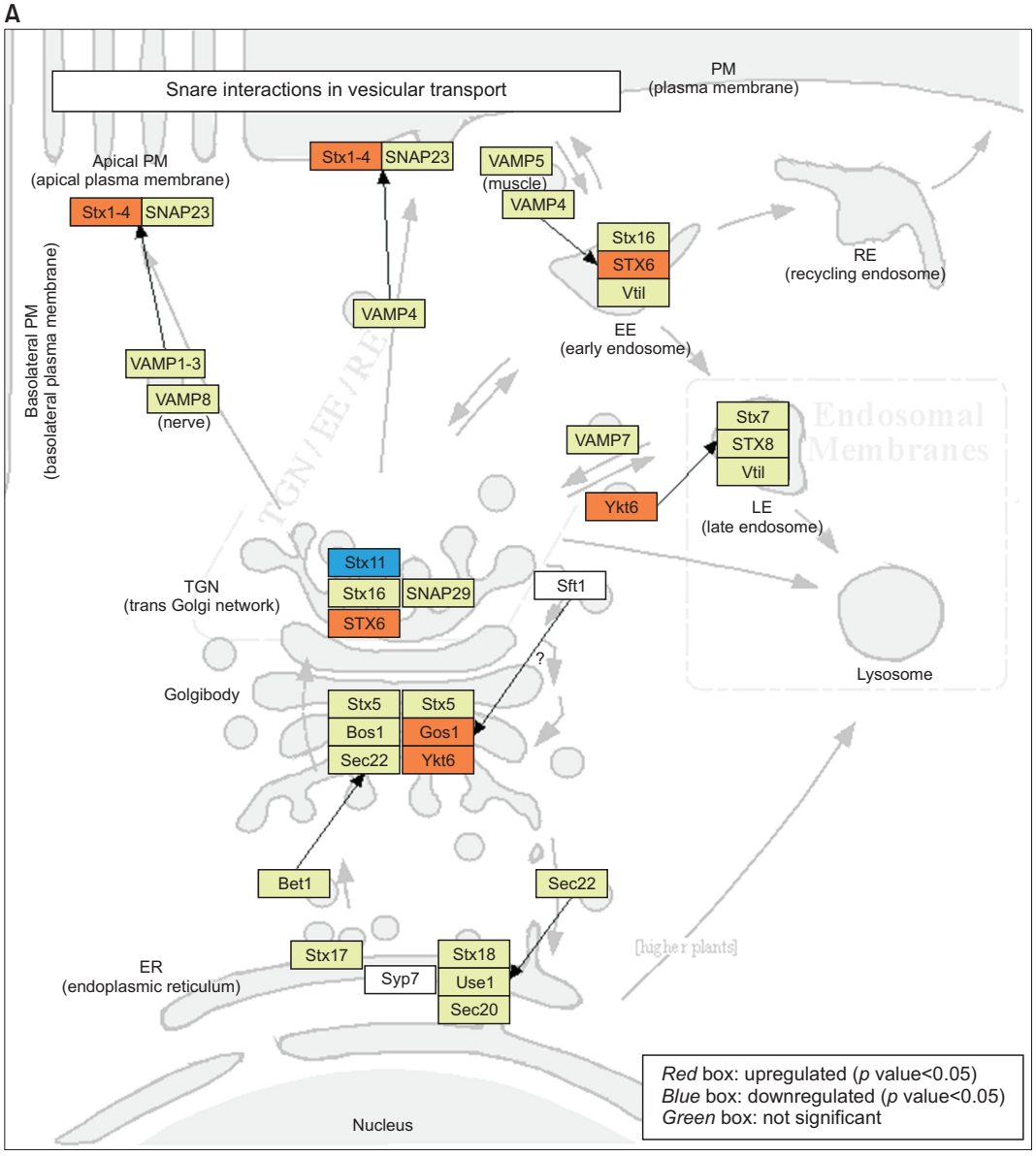


Fig. 7. Mapping GSEA results on KEGG pathways. For two pathways obtained from GSEA on the KEGG pathways, DEGs were expressed in the signaling pathway diagram of (A) "SNARE interaction in vesicular transport" (KEGG id#: 04130) and (B) "DNA replication" signaling pathways (KEGG id#: 03030). Significantly upregulated or downregulated genes (p<0.05) by RhoA depletion are colored red or blue. Genes without significant change are colored green.

tive or negative control.

Interestingly, these gene sets shared minimal similarity between collections, indicating that analysis of a single collection could reveal only a partial view of RhoA functions. Thus, comparative analyses among multiple collections followed to provide a robust method for the comprehensive identification of functions, including those collection-specific functions that may reduce any leakage in discovering functions.

Inference of RhoA functions

As each collection has a different number of gene sets (from 50 to 7350) and distinct characteristics, it was impossible to compare the results obtained for each collection directly. Nevertheless, by comparing a small set of gene sets presented in Fig. 3, robust functions could be derived based on functional similarity. This analysis showed that RhoA is a major regulator of cell cycle/growth and DNA repair (Fig. 5). These findings are consistent with numerous reports that RhoA is related closely to cancer progression. Recent evidence has suggested the role of RhoA in the DNA repair system (reviewed in (Cheng *et al*., 2021)). The analysis also showed that RhoA is involved in the metabolism of macromolecules, such as fatty acids, cholesterol, amino acids, nucleotides, and carbohydrates.

Meanwhile, several functions were detected only in a single collection. For instance, functions on "keratinization" and "response to fungus" were obtained only from the GO BP–spe-

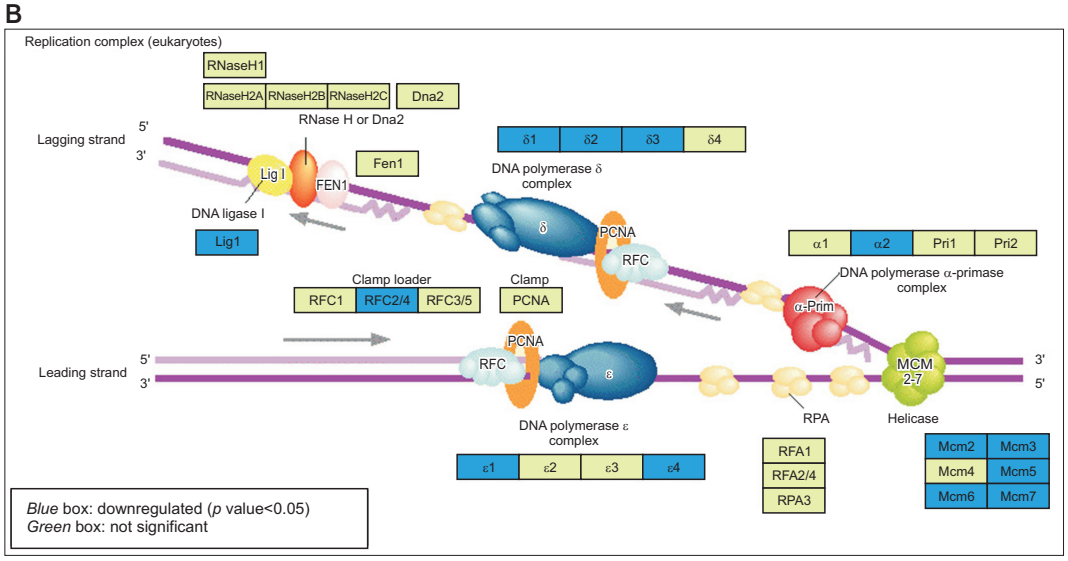


Fig. 7. Continued.

cific gene sets. Further research is needed to characterize the role and mechanisms of RhoA in these processes, and it would be interesting to identify whether these functions are regulated in a cytoskeleton remodeling–dependent or independent manner.

Evaluation of significant gene sets in the "hallmark" pathway collection

The overall distribution of fold changes was visualized to evaluate the quality of GSEA of the "hallmark" pathway (Fig. 6A). Fifty gene sets were sorted in order of |NES| vertically, and the log-scaled fold changes were expressed using point and bar plots. It was confirmed that gene sets with large |NES| (bottom) were distributed away from the centerline (red dotted line). The magnified plots of the 10 most enriched gene sets (Fig. 6A, lower panel) showed that the gene sets with large |NES| have more biased distributions from the centerline. This result indicates that GSEA effectively derived significant functions related to RhoA deletion.

The efficiency of GSEA was validated by comparing individual gene expression levels in a heatmap, which showed that genes related to the "p53 pathway" were generally upregulated in the RhoA-null group, whereas genes related to the "E2F Targets" were downregulated (Fig. 6B). Both point/bar and heatmap plots indicate that RhoA may negatively regulate the p53 pathway and positively regulate E2F-related cell growth under basal conditions.

Mapping analysis of the "KEGG pathway" collection

The KEGG pathway database provides a handy analysis tool that enhances our understanding of gene interaction by visualizing the molecular interaction network (Kanehisa *et al.*, 2017). In the "SNARE interaction in vesicular transport" gene set (pathway id 04130), *Gosr1*, *Stx1a*, *Stx2*, *Stx6*, and *Ykt6* were upregulated in the RhoA-null group (red), whereas *Stx11* was downregulated (blue) (p<0.05) (Fig. 7A). In the "DNA replication" gene set (pathway id 03030), all genes with significance were downregulated in the RhoA-null group (blue) (p<0.05): *Lig1*, *Mcm2*, *Mcm3*, *Mem5*, *Mem6*, *Mem7*, *Pola2*, *Pold1*, *Pold2*, *Pold3*, *Pole*, *Pole4*, *Rfc2*, and *Rfc4* (Fig. 7B). Thus, RhoA likely suppresses SNARE-mediated vesicular transport and promotes cell proliferation in normal conditions.

Functional analysis of "GO BP" based on a DAG

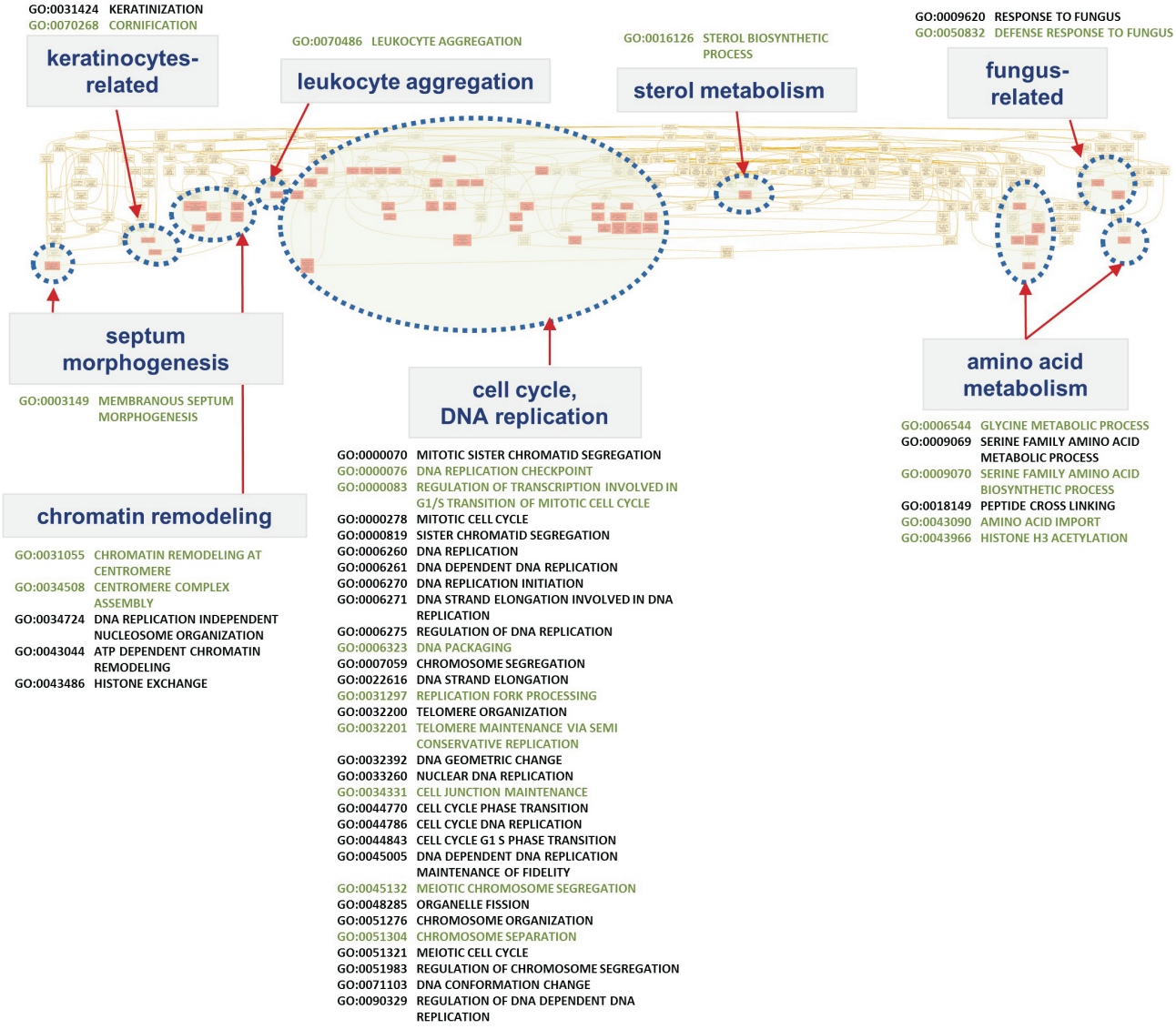
Unlike the "hallmark" and "KEGG pathway," "GO BP" provided a large number of functionally redundant gene sets, which made it difficult to summarize the enriched results. One valuable GO pathway property is the hierarchical structure among GO terms in which nodes represent pathways and edges describe a parent–child relationship. Genes in a descendant GO term are a subset of its parent term and more specific. This directional relationship between GO terms can be visualized using a DAG. To resolve the redundancy issue, the hierarchical structure of 51 enriched GO terms was visualized based on the DAG using GOView. The result showed that GO terms could be classified into eight different functional groups (Fig. 8A, circled), and most terms were included in groups related to cell proliferation and DNA replication.

Subsequently, the geometric distribution of the enriched GO terms was examined using AEGIS software (Zhu *et al.*, 2019). Enriched gene set pathways were displayed on a "focus and context" graph representation in a root-bound layout. (Fig. 8B). In the left focus graph, 51 enriched GO terms called focus pathways were shown as leaf anchors (blue closed circles) with their ancestors (purple closed circles) under the root anchor of "BP" (GO:0008150). The graph confirmed that the enriched GO pathways produced several clusters with their relatives. The context graph on the right side shows the number of the full ontology at each level. The node counts at each level are displayed with the number of ancestor and descendant terms of interest (purple bars) compared to the whole other GO terms (gray bars). The gene sets with various distances from the root were evenly selected.

Cross-evaluation of gene sets between distinct collections

In this study, GSEA resulted in gene sets with low similarity

A Directed acyclic diagram of GO BP $p<0.01$, INESI>2



B GO IDs of 51 representative gene sets used in the construction of DAG

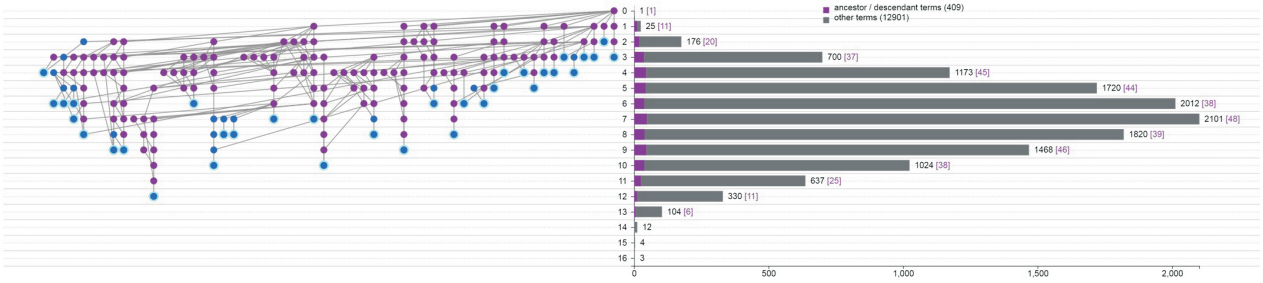


Fig. 8. A directed acyclic graph (DAG) showing the hierarchical structure of gene sets of GO BP. A DAG was constructed from 51 most enriched GO BP terms ($p<0.01$ and |NES|>2) using (A) GOView web-based software (Shoop *et al*., 2004) and (B) AEGIS software (Zhu *et al*., 2019). (A) GO BP terms are colored red, and the terms with similar functions are grouped and circled. The names of the GO BP terms are listed below the group name. The end nodes representing the lowest level are colored green. (B) A DAG constructed using AEGIS software. (Left) The focus graph renders the hierarchical structure of the inquired sub GO BP terms. The structure is expressed in a buoyant layout mode. Each node represents a GO term and each link represents a parent–child relationship. A parent node is always placed at a level higher than its children. The enriched GO BP terms are colored blue. (Right) The context graph provides full ontology under the root anchor of "biological process" (GO:0008150, 13,310 GO terms). At each level, the node counts of terms of interest are presented in the purple bar, with the whole other GO terms at the same level in gray.

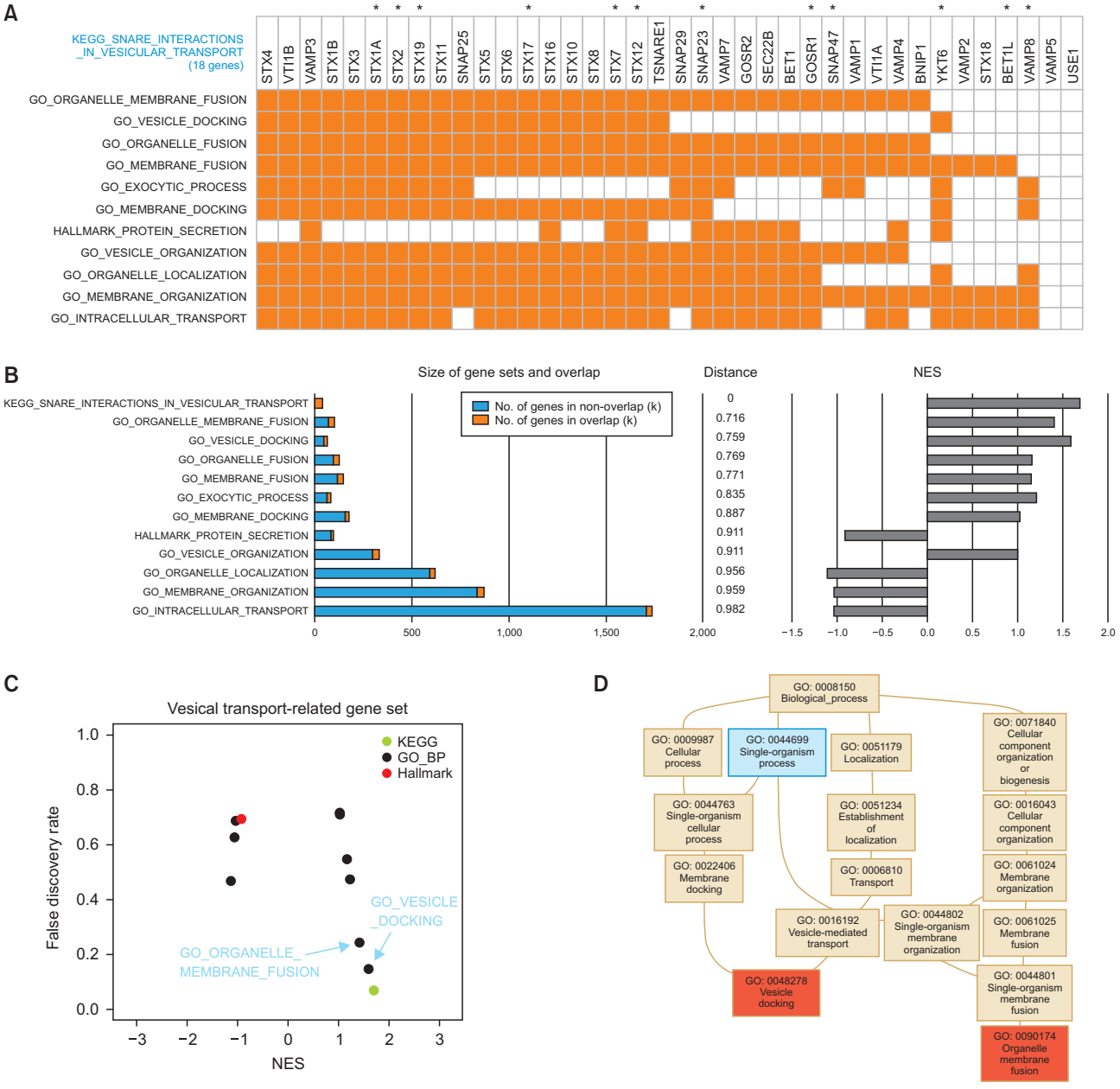


Fig. 9. Comparative analyses of the KEGG "SNARE interaction in vesicular transport" with gene sets from the hallmark and GO BP. The KEGG "SNARE interaction in vesicular transport" gene set consists of 18 genes. Gene sets similar to the KEGG were retrieved from the hallmark (one gene set) and GO BP (10 gene sets) collections. (A) Whether the searched gene sets contain the 18 genes was visualized with a heatmap in brown color. (B) The range of gene set sizes is expressed in the bar plot. The numbers of the overlapped genes are expressed in orange and those of non-overlapped genes are in blue (left). The Jaccard distance is provided to take the set size into account for comparison (middle). NES values are presented on the right side. The gene sets were sorted in the order of the Jaccard distance. (C) According to the collection, the FDR-NES plot of 12 selected gene sets shows that even similar gene sets are evaluated differently in GSEA. (D) A DAG diagram shows the resemblance between two gene sets closely related to the KEGG gene set: "GO vesicle docking" (GO: 0048278, blue) and "GO organelle membrane fusion" (GO: 0090174) in the orange-colored box. (E) Enrichment plots of the three most closely resembled GO BP gene sets.

among collections, and certain gene sets with similar names were regulated in opposing ways. For instance, "SNARE interaction in vesicular transport" (KEGG id: 04130, N_{gene}=38) had a high significance (adjusted *p* value<0.1, NES>1.5), whereas gene sets of GO BP with "SNARE" in the name showed no significance in GSEA: "SNARE complex assembly" (GO: 0035493', N_{gene}=19) (adjusted p=0.670, NES=–1.1) and "regulation of SNARE complex assembly" (GO: 0035542', N_{gene}=11) (adjusted *p*=0.792, NES=–0.97) (Supplementary Table 2).

These findings indicate that each collection's composition of gene sets can be heterogenous even if the names are similar. The elements in the gene set of one collection were

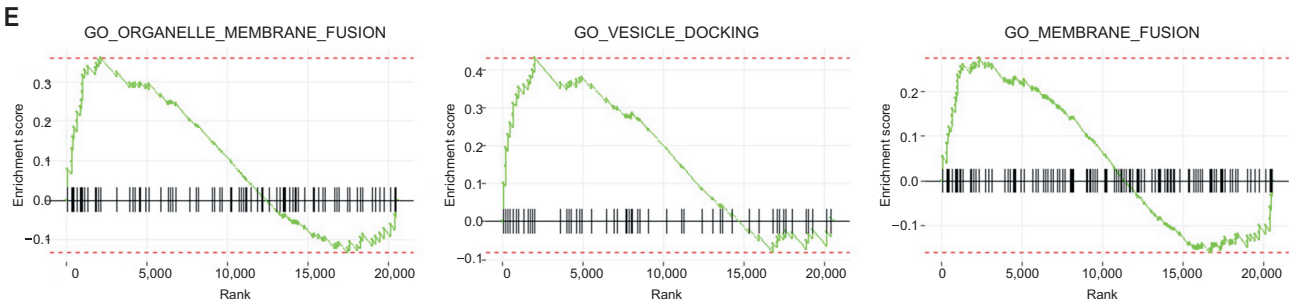


Fig. 9. Continued.

compared with those of other collections to test this hypothesis. The similarity between gene sets was measured quantitatively by calculating the distance. In the first example, KEGG "SNARE interaction in vesicular transport" was compared with GO BP gene sets with "SNARE" in the name. The Jaccard distances of the KEGG gene set between GO BP gene sets were 0.88 and 1.0 for the GO BP "SNARE complex assembly" and "regulation of SNARE complex assembly," respectively. Most of the leading-edge genes of "SNARE interaction in vesicular transport" were not included in the other two gene sets (Supplementary Fig. 1). Thus, gene sets from distinct collections may have a heterogenous composition of elements even though they have similar names, and it may be inappropriate to infer function only by the name of the gene set.

As this discordance may occur by the lack of rules to constitute genes and naming, I further searched for gene sets homologous to the KEGG "SNARE interaction in the vesicular transport" gene set (n=38 genes), using the overlap search tool provided by MSigDB. The KEGG gene set had one and ten similar gene sets in the "hallmark" and "GO BP" collections, respectively. The summary of these 11 gene sets is provided in Supplementary Table 3. The gene sets contained many genes in common with the KEGG gene set (Fig. 9A). However, the gene set sizes and GSEA NES values varied from 65 to 1741 genes and from –1.12 to 1.61, respectively (Fig. 9B). When the Jaccard distance was measured to take the set size into account, "organelle membrane fusion," "vesicle docking," "organelle fusion," and "membrane fusion" gene sets of GO BP were shown to be the closest (0.716-0.771) (Fig. 9B). In particular, "organelle membrane fusion" and "vesicle docking" showed FDR and NES values similar to the KEGG "SNARE interaction in vesicular transport" (Fig. 9C). In the DAG structure, these two gene sets (red) had "single-organism process" (GO: 0044699, blue) as a common ancestor (Fig. 9D). These gene sets with high similarity produced closely resembled enrichment plots (Fig. 9E). These findings indicate that a clear understanding of gene set functions in GSEA requires inspecting the composition of the element within the gene set and those components representing the function correctly.

Similarly, although the "p53 pathway" was the most significant gene set in hallmark gene sets, no significant gene set containing "p53" in the name was found in the other collections. A search for overlaps using MSigDB suggested that around 30 gene sets shared similarity with the hallmark "p53 pathway" gene set in three collections (Fig. 10A). The size and degree of overlap varied. The KEGG "p53 signaling pathway" had the closest Jaccard distance value of 0.906 with positive NES. However, most gene sets had negative NES values. The NES significance plot for the 10 gene sets with the largest |NES| showed that the hallmark "p53 pathway" (marked "x") failed to cluster with any similar property (Fig. 10B). When log2FC distributions of 10 gene sets (i-x) were compared, it was clear that the hallmark "p53 pathway" had a unique pattern in the distribution (Fig. 10C). Although the KEGG "p53 signaling pathway" (marked "ix") had the property most similar to the hallmark "p53 pathway," there was significant discordance in the gene set size and elements. In particular, these two gene sets' enrichment plots showed only partial similarity (Fig. 10D), with leading-edge genes only partially overlapped in two gene sets (overlapped genes: *Ccnd2*, *Ccng1*, *Cdkn1a*, *Ei24*, *Gadd45a*, *Mdm2*, and *Zmat3*).

When using public gene sets in GSEA, it is essential to be aware that (1) the size and composition of gene sets are heterogenous in each collection, (2) using multiple collections may compensate for the gaps that would result from using a single collection, (3) gene set names are not sufficient to infer function, and (4) the components of gene sets must be examined to understand the functions in the appropriate context.

DISCUSSION

Most RhoA research has focused on actin cytoskeleton regulation. In this study, I sought to investigate the functions of RhoA in an unbiased manner through bioinformatics data analysis based on microarray datasets, not limited to the known functions of RhoA. Previously, Garcia-Mariscal *et al*. (2018) focused on the role of RhoA in retinol signaling by using gene expression data. In this study, functional enrichment analyses were performed to exclude prejudice in exploring RhoA functions. Three different gene set collections were used to minimize bias that may occur by selecting specific gene sets. In short, the functions of RhoA common to the three collections were cell cycle regulation and DNA replication. Interestingly, RhoA regulated E2F pathway positively, while p53 pathway negatively. Because p53 upregulates p21 leading to the inhibition of a Cyclin dependent kinase 2 and subsequent activation of Rb transcriptional corepressor, p53 inactivation is closely linked to the activation of E2F (Sherr and McCormick, 2002). This finding suggests that RhoA is a key factor that can efficiently regulate cell growth by controlling both machineries that promote or inhibit the cell cycle. Gene set collection-specific functions were also obtained. For example, the KEGG pathway diagram confirmed that functions such as vesicular

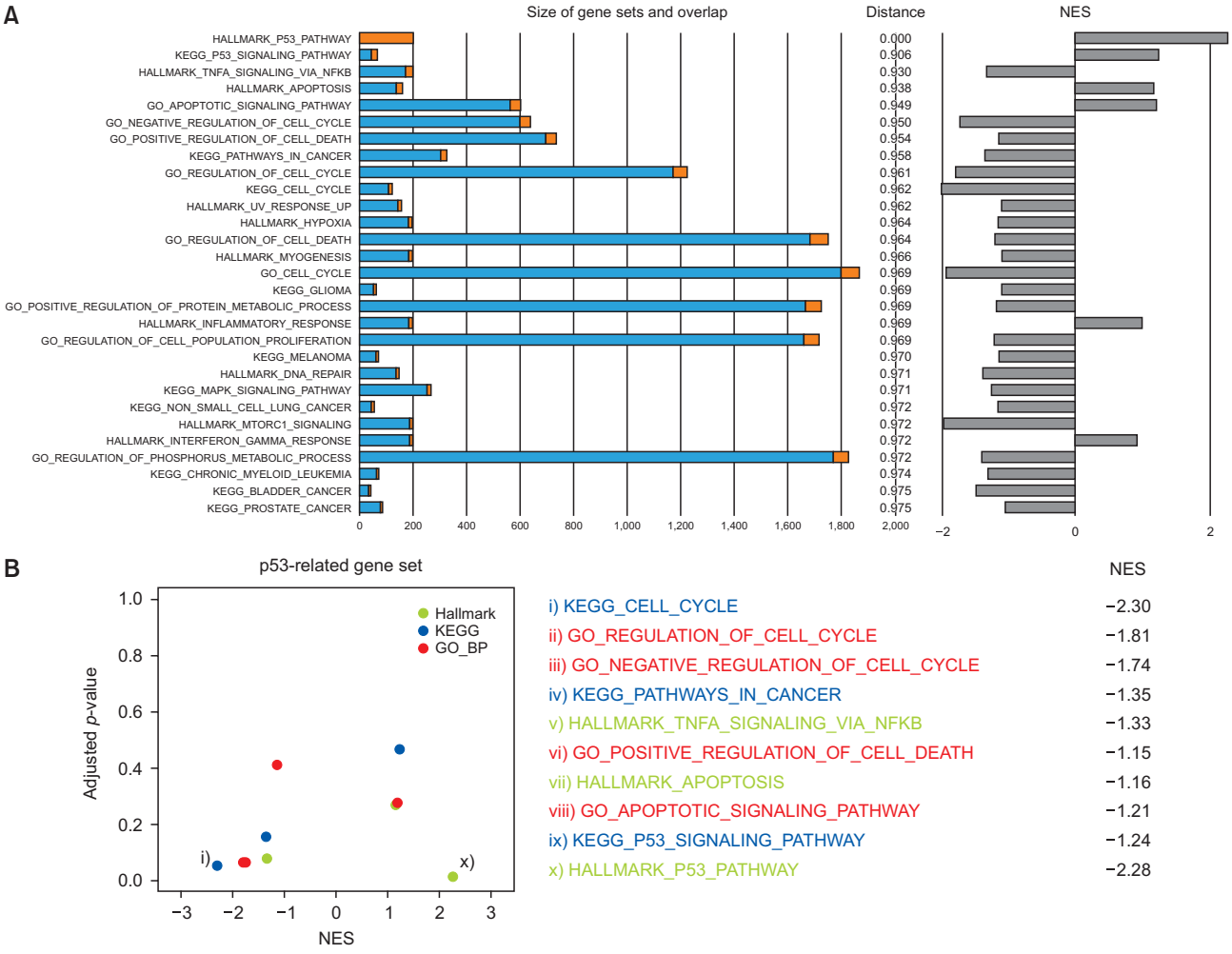


Fig. 10. Comparative analyses of the hallmark "p53 pathway" with gene sets from the KEGG and GO BP. The gene sets with similarity with the hallmark "p53 pathway" gene set (50 genes) were searched from the KEGG and GO BP collections. (A) Gene set sizes are expressed in the bar plot (left). Orange, the overlapped genes; blue, non-overlapped genes. The Jaccard distance is provided in the middle. NES values are presented on the right side. The gene sets were sorted in the order of the Jaccard distance. (B) The FDR-NES plot of the 10 selected gene sets (left). The gene set is listed in increasing NES value order, and the corresponding NES value is provided on the right. (C) The distribution of log2FC values of individual genes. The 10 selected gene sets were examined. Each dot represents a mean log2FC value of an individual gene element. Gene sets are marked with the number along the y-axis. (D) The enrichment plots for the hallmark "p53 pathway" and the KEGG "p53 signaling pathway" gene sets are compared.

transport and DNA replication could be regulated through SNARE.

Currently, the MSigBD site provides nine main collections with their sub-collections. These collections were curated by the field's experts. Thus, the appropriate collection needs to be selected based on the purpose and context of the study. Importantly, however, there is no guarantee that the chosen collection will fully cover the desired aspects of biology, as the integrity of collections has not been clearly verified. Thus, it seems important for researchers to keep in mind this potential weakness embedded in the collections.

Generally, the larger the collection, the more faithfully the function can be explored, but on the other hand, a large one likely increases redundancy between gene sets (Liberzon *et al.*, 2015). The main problem in the large-sized collection is that a bunch of redundant functions make up most of the top list and are overexposed, so lowly scored functions are often ignored even though they can be meaningful. On the other hand, a small collection presents more concise results so that researchers can quickly and easily find features. However, it has the disadvantage that researchers can potentially lose meaningful functions because the search is made with a limited number of gene sets. In this regard, rather than choosing a single collection, it would be ideal to test and compare several gene collections to see the differences in results. When using a gene set, it seems necessary to evaluate the completeness of the work through additional analysis that takes into account the characteristics of each collection.

This study demonstrated that each gene set collection comprises gene sets from different perspectives, and the results are heavily influenced by which collection is selected, so reliance on a single gene set collection may introduce bias in functional analysis.

In this study, multiple gene set collections were used to (1)

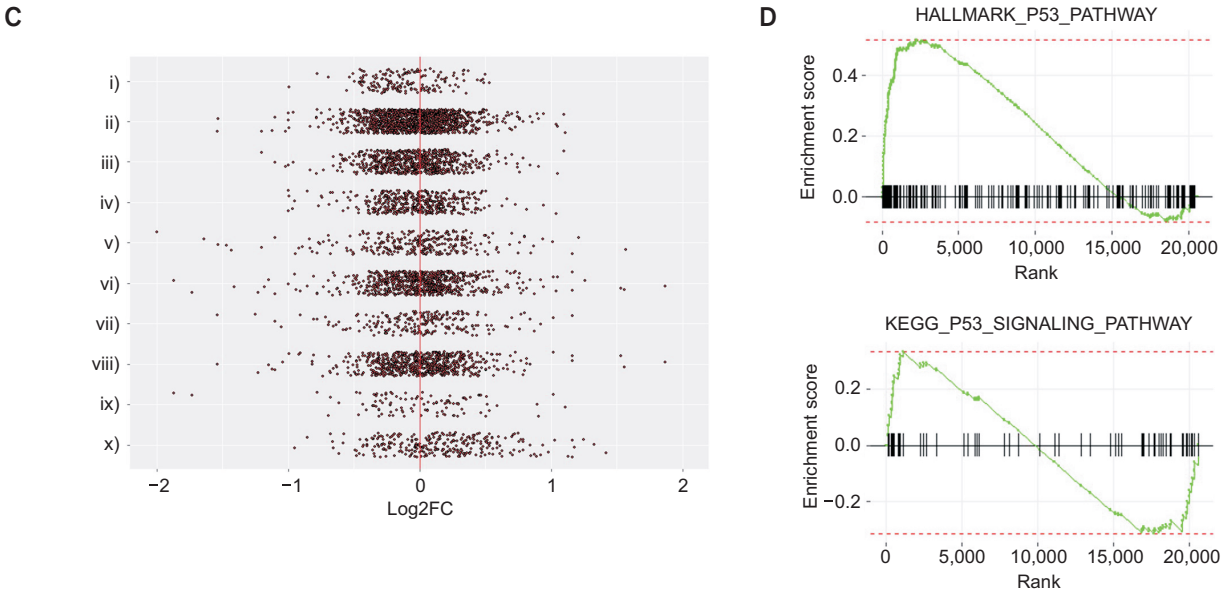


Fig. 10. Continued.

search for robust, commonly derived functions, and (2) reduce the opportunity to omit positive results because of the incompleteness of a single collection. Interestingly, the comparative experiment showed that the number and size of gene sets constituting each collection and the types of elements were diverse. It was possible to derive different functions depending on which collection was selected. Even for gene sets with similar names, NES and significance were different because of differences in composition. For example, in the Hallmark collection, only the p53 pathway showed a significant increase in expression, but other collections do not have a gene set significantly similar to this.

On the other hand, although many gene sets have a keyword related to the action of SNARE in common, there were cases where the NES values were different because of a significant difference in the type and number of constituent elements (Supplementary Fig. 1). Thus, the study design must consider the appropriate selection of gene set collections for functional enrichment analyses and interpretation. In other words, the gene set derived from the analysis is influenced greatly by the nature of the collection itself. Even gene sets with similar names contained different elements. This phenomenon seems to occur because there is no standard criterion to select the elements constituting each collection's gene set. Therefore, research bias can be minimized only when these problems are well-recognized in the actual functional analysis process. This analysis showed that, even though a gene set with significantly altered expression is derived, its function cannot be inferred simply from its name.

Even so, meaningful conclusions could be drawn if the functional characteristics of each collection were utilized correctly. For example, the KEGG pathway showed interesting results using a schematic pathway. For the first time, this study showed that RhoA consistently negatively regulates the expression of SNARE-related genes in vesicular transport. In contrast, genes involved in DNA replication are positively regulated by RhoA. RhoA functions in SNARE gene expression have not yet been elucidated, and it will be interesting to decipher the mechanism in future studies. There were also unintended outcomes in that the individual enrichment analysis yielded different results depending on the collection's characteristics, but it is worth recalling that when several collections are combined, robust functions may be derived. Also, it seems that reliable results can be obtained if the collection characteristics are utilized well.

ACKNOWLEDGMENTS

This study was supported by grants from the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (2016R1D-1A1B01012515), Republic of Korea.

REFERENCES

Adnane, J., Muro-Cacho, C., Mathews, L., Sebti, S. M. and Munoz-Antonia, T. (2002) Suppression of rho B expression in invasive carcinoma from head and neck cancer patients. *Clin. Cancer Res.* **8**, 2225-2232.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29.

Carlson, M. (2016) mouse4302.db: Affymetrix Mouse Genome 430 2.0 Array annotation data (chip mouse4302). R package version 3.2.3.

Chen, Z., Sun, J., Pradines, A., Favre, G., Adnane, J. and Sebti, S. M. (2000) Both farnesylated and geranylgeranylated RhoB inhibit malignant transformation and suppress human tumor growth in nude mice. *J. Biol. Chem.* **275**, 17974-17978.

Cheng, C., Seen, D., Zheng, C., Zeng, R. and Li, E. (2021) Role of small GTPase RhoA in DNA damage response. *Biomolecules* **11**, 212.

Clark, E. A., Golub, T. R., Lander, E. S. and Hynes, R. O. (2000) Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532-535.

Du, W. and Prendergast, G. C. (1999) Geranylgeranylated RhoB mediates suppression of human tumor cell growth by farnesyltransferase inhibitors. *Cancer Res.* **59**, 5492-5496.

Duong, K. H. M. and Chun, K. H. (2019) Regulation of glucose transport by RhoA in 3T3-L1 adipocytes and L6 myoblasts. *Biochem. Biophys. Res. Commun.* **519**, 880-886.

Faried, A., Faried, L. S., Kimura, H., Nakajima, M., Sohda, M., Miyazaki, T., Kato, H., Usman, N. and Kuwano, H. (2006) RhoA and RhoC proteins promote both cell proliferation and cell invasion of human oesophageal squamous cell carcinoma cell lines *in vitro* and *in vivo*. *Eur. J. Cancer* **42**, 1455-1465.

Garcia-Mariscal, A., Peyrollier, K., Basse, A., Pedersen, E., Ruhl, R., van Hengel, J. and Brakebusch, C. (2018) RhoA controls retinoid signaling by ROCK dependent regulation of retinol metabolism. *Small GTPases* **9**, 433-444.

Gautier, L., Cope, L., Bolstad, B. M. and Irizarry, R. A. (2004) Affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315.

Gentleman, R., Carey, V., Huber, W. and Hahne, F. (2018) genefilter: methods for filtering genes from high-throughput experiments. R package version 1.64.0.

Haga, R. B. and Ridley, A. J. (2016) Rho GTPases: regulation and roles in cancer cell biology. *Small GTPases* **7**, 207-221.

Hodge, R. G. and Ridley, A. J. (2016) Regulating Rho GTPases and their regulators. *Nat. Rev. Mol. Cell Biol.* **17**, 496-510.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264.

Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y., Ichimura, T., Ushiku, T., Funahashi, S., Tateishi, K., Wada, I., Shimizu, N., Nomura, S., Koike, K., Seto, Y., Fukayama, M., Aburatani, H. and Ishikawa, S. (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.* **46**, 583-587.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353-D361.

Kim, J. G., Islam, R., Cho, J. Y., Jeong, H., Cap, K. C., Park, Y., Hossain, A. J. and Park, J. B. (2018) Regulation of RhoA GTPase and various transcription factors in the RhoA pathway. *J. Cell. Physiol.* **233**, 6381-6392.

Kimura, K., Ito, M., Amano, M., Chihara, K., Fukata, Y., Nakafuku, M., Yamamori, B., Feng, J., Nakano, T., Okawa, K., Iwamatsu, A. and Kaibuchi, K. (1996) Regulation of myosin phosphatase by Rho and Rho-associated kinase (Rho-kinase). *Science* **273**, 245-248.

Kolde, R. (2019) pheatmap: Pretty Heatmaps. Available from: https://CRAN.R-project.org/package=pheatmap/.

Leung, T., Manser, E., Tan, L. and Lim, L. (1995) A novel serine/threonine kinase binding the Ras-related RhoA GTPase which translocates the kinase to peripheral membranes. *J. Biol. Chem.* **270**, 29051-29054.

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417-425.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. and Mesirov, J. P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740.

Maekawa, M., Ishizaki, T., Boku, S., Watanabe, N., Fujita, A., Iwamatsu, A., Obinata, T., Ohashi, K., Mizuno, K. and Narumiya, S. (1999) Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase. *Science* **285**, 895-898.

Mazieres, J., Antonia, T., Daste, G., Muro-Cacho, C., Berchery, D., Tillement, V., Pradines, A., Sebti, S. and Favre, G. (2004) Loss of RhoB expression in human lung cancer progression. *Clin. Cancer Res.* **10**, 2742-2750.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. and Groop, L. C. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267-273.

Narumiya, S. and Thumkeo, D. (2018) Rho signaling research: history, current status and future directions. *FEBS Lett.* **592**, 1763-1776.

Porter, A. P., Papaioannou, A. and Malliri, A. (2016) Deregulation of Rho GTPases in cancer. *Small GTPases* **7**, 123-138.

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/.

Ridley, A. J. (2006) Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *Trends Cell Biol.* **16**, 522-529.

Ridley, A. J. and Hall, A. (1992) The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389-399.

Sakata-Yanagimoto, M., Enami, T., Yoshida, K., Shiraishi, Y., Ishii, R., Miyake, Y., Muto, H., Tsuyama, N., Sato-Otsubo, A., Okuno, Y., Sakata, S., Kamada, Y., Nakamoto-Matsubara, R., Tran, N. B., Izutsu, K., Sato, Y., Ohta, Y., Furuta, J., Shimizu, S., Komeno, T., Sato, Y., Ito, T., Noguchi, M., Noguchi, E., Sanada, M., Chiba, K., Tanaka, H., Suzukawa, K., Nanmoku, T., Hasegawa, Y., Nureki, O., Miyano, S., Nakamura, N., Takeuchi, K., Ogawa, S. and Chiba, S. (2014) Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 171-175.

Sato, N., Fukui, T., Taniguchi, T., Yokoyama, T., Kondo, M., Nagasaka, T., Goto, Y., Gao, W., Ueda, Y., Yokoi, K., Minna, J. D., Osada, H., Kondo, Y. and Sekido, Y. (2007) RhoB is frequently downregulated in non-small-cell lung cancer and resides in the 2p24 homozygous deletion region of a lung cancer cell line. *Int. J. Cancer* **120**, 543-551.

Satoh, K., Fukumoto, Y. and Shimokawa, H. (2011) Rho-kinase: important new therapeutic target in cardiovascular diseases. *Am. J. Physiol. Heart Circ. Physiol.* **301**, H287- H296.

Sergushichev, A. A. (2016) An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* doi: 10.1101/060012.

Sherr, C. J. and McCormick, F. (2002) The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103-112.

Shimokawa, H., Sunamura, S. and Satoh, K. (2016) RhoA/Rho-kinase in the cardiovascular system. *Circ. Res.* **118**, 352-366.

Shoop, E., Casaes, P., Onsongo, G., Lesnett, L., Petursdottir, E. O., Donkor, E. K., Tkach, D. and Cosimini, M. (2004) Data exploration tools for the Gene Ontology database. *Bioinformatics* **20**, 3442-3454.

Simpson, K. J., Dugan, A. S. and Mercurio, A. M. (2004) Functional analysis of the contribution of RhoA and RhoC GTPases to invasive breast carcinoma. *Cancer Res.* **64**, 8694-8701.

Smyth, G. K., Michaud, J. and Scott, H. S. (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-2075.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545-15550.

Suwa, H., Ohshio, G., Imamura, T., Watanabe, G., Arii, S., Imamura, M., Narumiya, S., Hiai, H. and Fukumoto, M. (1998) Overexpression of the rhoC gene correlates with progression of ductal adenocarcinoma of the pancreas. *Br. J. Cancer* **77**, 147-152.

Taiyun, W. and Viliam, S. (2017) R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from: https://github.com/taiyun/corrplot/.

Wang, D., Dou, K., Xiang, H., Song, Z., Zhao, Q., Chen, Y. and Li, Y. (2007) Involvement of RhoA in progression of human hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **22**, 1916-1920.

Wang, J., Wu, Q., Zhang, L. H., Zhao, Y. X. and Wu, X. (2016) The role of RhoA in vulvar squamous cell carcinoma: a carcinogenesis, progression, and target therapy marker. *Tumour Biol.* **37**, 2879-2890.

Wang, K., Yuen, S. T., Xu, J., Lee, S. P., Yan, H. H., Shi, S. T., Siu, H. C., Deng, S., Chu, K. M., Law, S., Chan, K. H., Chan, A. S., Tsui, W. Y., Ho, S. L., Chan, A. K., Man, J. L., Foglizzo, V., Ng, M. K., Chan,

A. S., Ching, Y. P., Cheng, G. H., Xie, T., Fernandez, J., Li, V. S., Clevers, H., Rejto, P. A., Mao, M. and Leung, S. Y. (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573-582.

Watanabe, N., Kato, T., Fujita, A., Ishizaki, T. and Narumiya, S. (1999) Cooperation between mDia1 and ROCK in Rho-induced actin reorganization. *Nat. Cell Biol.* **1**, 136-143.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Yoo, H. Y., Sung, M. K., Lee, S. H., Kim, S., Lee, H., Park, S., Kim, S. C., Lee, B., Rho, K., Lee, J. E., Cho, K. H., Kim, W., Ju, H., Kim, J., Kim, S. J., Kim, W. S., Lee, S. and Ko, Y. H. (2014) A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 371-375.

Zhao, R., Liu, K., Huang, Z., Wang, J., Pan, Y., Huang, Y., Deng, X., Liu, J., Qin, C., Cheng, G., Hua, L., Li, J. and Yin, C. (2015) Genetic variants in Caveolin-1 and RhoA/ROCK1 are associated with clear cell renal cell carcinoma risk in a chinese population. *PLoS ONE* **10**, e0128771.

Zhou, J., Hayakawa, Y., Wang, T. C. and Bass, A. J. (2014) RhoA mutations identified in diffuse gastric cancer. *Cancer Cell* **26**, 9-11.

Zhu, J., Zhao, Q., Katsevich, E. and Sabatti, C. (2019) Exploratory gene ontology analysis with interactive visualization. *Sci. Rep.* **9**, 7793.