A meta-analysis to identify novel diagnostic and therapeutic targets for Dupuytren's disease

Tae Hwan Park, MD, PhD[1,2]; Dongha Kim, BS, MSc[3]; Young-Seok Lee, BS, Msc[3]; Sung Young Kim, MD, PhD[3]

1. Department of Plastic and Reconstructive Surgery, CHA Bundang Medical Center, School of Medicine, CHA University, Seongnam, Korea,
2. KCRN Research, Germantown, Maryland, 20874, and
3. Department of Biochemistry, School of Medicine, Konkuk University, Seoul, Korea

Reprint requests:
Sung Young Kim MD, PhD, Department of Biochemistry, School of Medicine, Konkuk University, Seoul 05029, Republic of Korea.
Tel: +82 2 2049 6060;
Fax: +82 2 2049 6060;
Email: palelamp@kku.ac.kr

Manuscript received: July 16, 2019
Accepted in final form: October 15, 2019

DOI:10.1111/wrr.12774

ABSTRACT

The aim of this study was to determine novel candidate genes for Dupuytren's disease by performing a meta-analysis. We identified 261 genes (111 up-regulated and 150 down-regulated) that were consistently expressed differentially in Dupuytren's disease across the studies. We performed functional enrichment on total sets of the identified 261 genes and confirmed that most of the genes were closely related to common processes of diseases in general. From the integrated studies of the gene-correlation network and the protein–protein interaction network, we identified three functional modules in the gene co-expression network and four hub gene clusters in the protein–protein interaction network that shared the same genes and represented similar biological functions, implying that the seven groups identified in the systematic analysis of these two networks might be involved in the pathogenesis of Dupuytren's disease. This work demonstrates potential in developing experimental and clinical strategies for understanding and treating Dupuytren's disease.

INTRODUCTION

Dupuytren's disease (also known as Dupuytren's contracture) is frequently observed over the age of 45 years, with male predominance and a high prevalence in northwestern Europeans of Celtic descent (40% prevalence in the Scandinavian population).[1] This condition is a chronic fibrotic disease of the palmar and digital fascial structures, characterized by nodular thickening and subsequent contracture. Deformity of the hand occurs primarily at the metacarpophalangeal and proximal interphalangeal joint level, resulting in functional disabilities. The global prevalence of this disease is estimated to range from 0.2% to 56%.[2]

Myofibroblasts are known to be the key cells responsible for both contraction and extracellular matrix deposition, but the initiating event causing myofibroblast proliferation is unresolved. Numerous published studies support various treatment modalities for potentially effective treatment of Dupuytren's disease by regulating contractile myofibroblasts.[2–10] However, its pathogenesis remains largely obscure.[11] Only over the past 30 years, many researchers suggest a strong genetic heritance in Dupuytren's disease. This is evidenced by its high prevalence in certain ethnicities and comparable disease grade or severity related to positive family history.[12]

Recent advances in high-throughput molecular approaches enable the determination of up-regulation or down-regulation of various genes. Regarding Dupuytren's disease, three microarray data sets are available in the Gene Expression Omnibus (GEO) and ArrayExpress. However, there have been inconsistent findings in each study, which is in line with our meta-analysis showing significant losses (n = 896) identified in the meta-analysis.

To develop effective treatment, molecular studies indicating potential therapeutic implications are very important to clinicians and researchers. The aim of this study was to ascertain novel candidate genes for Dupuytren's disease by performing a meta-analysis.

MATERIALS AND METHODS

Selection of eligible microarray data sets

We collected studies related to gene expression with regard to Dupuytren's disease using the PubMed database, NCBI GEO (available at http://www.ncbi.nlm.nih.gov/geo/), and ArrayExpress (available at http://www.ebi.ac.uk/arrayexpress/). For objective assessment, two reviewers assessed and extracted data. Any discrepancies between these reviewers were resolved by consensus or by consultation with a third reviewer. The search terms used were "Dupuytren's disease," "Dupuytren disease," "Dupuytren," "Dupuytren's contracture," and "gene and/or expression and/or profile." From each study that was included, we extracted the GEO accession number, platform type, sample type, and gene expression data.

Identification of differentially expressed genes by meta-analysis of microarray data sets

We performed meta-analysis of gene expression profiles in the selected microarray data sets with rank methods

(RankProd package in R) implemented in the web-based INMEX program.[13] Before data were analyzed, all gene and probe IDs from each data set were annotated as Entrez IDs for consistency, and intensity values for gene expression were log2-transformed and quantile-normalized so that their mean and unit variance was zero. RankProd (developed from the nonparametric rank product method) was used to apply a statistically rigorous algorithm, which included biological intuition of fold-change (FC) criteria.[14,15] The ranks of differentially expressed genes (DEGs) based on FC scores in all possible pair-wise comparisons to the integrated data sets were determined. With the RankProd algorithm, genes that were consistently identified as up-regulated or down-regulated DEGs in whole data sets were assigned a higher rank depending on false discovery rate (FDR)-adjusted *p*-value (threshold, ≤0.05) and FC level in each number of replicates multiplied across the given data sets; these were considered the most significantly regulated DEGs. The expression profiles of DEGs across different data sets/conditions were visualized as heat maps by implementing the "Pattern extractor" tool.

Gene ontology and Kyoto Encyclopedia of genes and genome pathway enrichment analysis

To consider the biological functions of the DEGs in Dupuytren's disease, Gene Ontology (GO) category and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway enrichment analyses were carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) bioinformatics online resource (available at http://david.abcc.ncifcrf.gov/).

The annotated gene identifiers for DEGs identified by meta-analysis were summarized into the categories of integrated GO terms and KEGG pathway with threshold of *p*-value ≤0.05 for biological significance at the cellular level.

Potential transcription factors and microRNAs analysis for the identified DEGs

For the prediction of potential transcription factors and microRNA-targeted genes among the DEGs identified by meta-analysis, we performed transcription factor and microRNA enrichment analysis by one-on-one check with database of annotated gene sets retrieved in the Molecular Signatures Database (MSigDB; available at http://www.broadinstitute.org/gsea/msigdb/index.jsp) with threshold of *p*-value ≤0.05. For statistical method and multiple test correction, hypergeometric algorithm and Benjamini–Hochberg adjustment were used in the analysis.

Gene-correlation network analysis of the identified DEGs

For prediction of functional interaction between the identified DEGs, we constructed gene-correlation network between the identified DEGs and additional genes in GeneMANIA webserver (available: http://www.genemania.org/).[16]

The interaction between the genes in the network was composed of only the co-expression except for physical/genetic interactions, pathways, co-localization, protein domain similarity, and predicted interactions. The network was filtered by weighting under GO term-biological process and by excluding all the interactions with threshold of *p*-value ≤0.05.

We performed the clustering of functional modules in the network by using the fast-greedy (I) algorithm implemented in the Cytoscape plugin, Community Clusters GLay (available: http://cytoscape.wodaklab.org/wiki/CommunityStructureLayout).[17] The overrepresented cellular functions within each module were identified by the functional enrichment analyses in DAVID program.

Protein–protein interaction network analysis

To predict biological activity of the identified DEGs at the protein level, the DEGs were imported into the protein–protein interaction (PPI) network downloaded from the Biological General Repository for Interaction Data sets (BioGRID) (available at http://thebiogrid.org/). The network was screened on a genome-wide scale in Cytoscape software (available at http://www.cytoscape.org/).

The hub proteins of functional modules were further identified in the PPI network by using a Cytoscape plugin, ClusterONE (available: http://apps.cytoscape.org/apps/clusterone).[18] The overrepresented biological functions within each module were identified by functional enrichment analyses by an online analysis tool in DAVID.

RESULTS

Microarray data sets on Dupuytren's disease cell lines selected for the meta-analysis

We extracted samples from three microarray data sets that met our criteria (see Materials and Methods section, Figure 1A). GSE31356 of the GEO series was a microarray data set extracted from NCBI GEO and originated in the palmar fascia of patients with Dupuytren's contracture by using Affymetrix Human Genome U133A array. Data sets E-GEOD-59746 and E-GEON-21221 were obtained from ArrayExpress of EMBL-EBI and were established on Affymetrix Human Genome U133 plus 2.0 and Illumina Human-6 v1.0 Expression BeadChip, respectively (Table 1).

Identification of up-regulated or down-regulated DEGs in the meta-analysis of multiple data sets

We selected DEGs based on the estimated percentage of false-positives produced by the algorithm in RankProd. We identified 261 DEGs from GSMs in which the fibroblast cell lines of Dupuytren's disease were compared with a normal control, including 111 up-regulated and 150 down-regulated genes (Figure 2). Interestingly, the majority of genes (209) in DEGs identified in the meta-analysis were not identified in any individual analysis. These 209 "gain" genes have a high potential possibility of identifying novel biomarkers related to the development and pathology of Dupuytren's disease. Heat maps were used to depict correlation in expression patterns for a subset of top 50 DEGs from the three studies (Supporting Information Figure S1). The 20 most significantly up-regulated or down-regulated DEGs, with FDR-adjusted *p*-value ≤0.05, are shown in Table 2. The up-regulated DEGs with the lowest FDR-adjusted *p*-value of 1.0E-10 were *RPS4Y1*, *EIF1AY*, *NLGN4Y*, *ZIC1*, *NLGN4X*, *HOXB13*, *TFAP2C*, *ICA1*, *HCP5*, and *HOXC10*. The down-regulated DEGs with the lowest FDR-adjusted

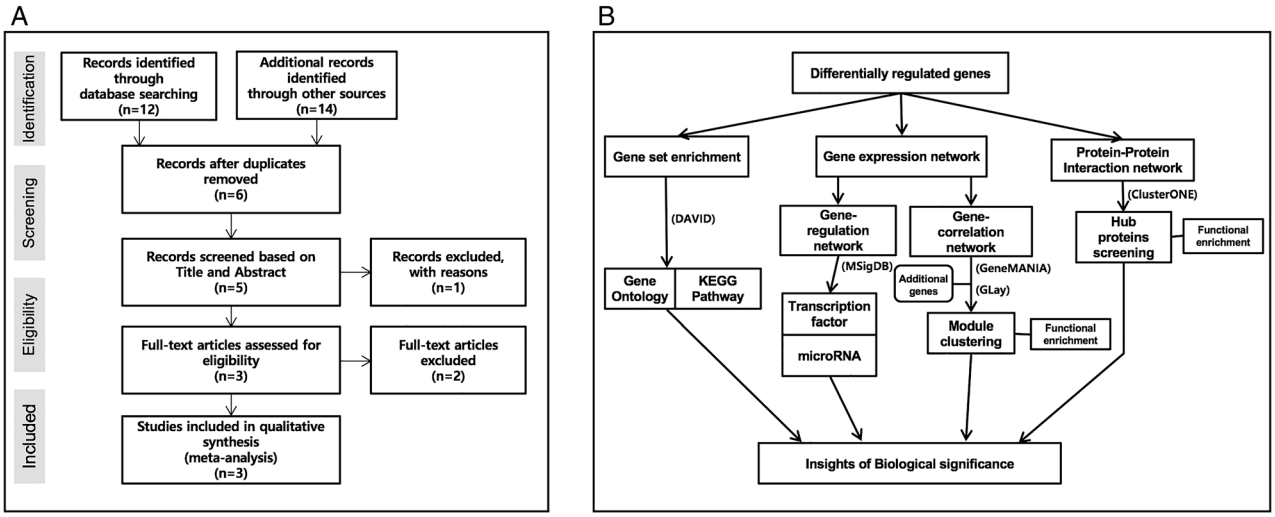

Figure 1. General workflow depicting gradational process of methodology adopted in meta-analysis of this study. (A) Selection of microarray data sets for meta-analysis of Dupuytren's disease, according to Prisma 2009 flow diagram. (B) Comprehensive analysis of DEGs identified by meta-analysis, including gene set enrichment analysis, gene expression network (gene regulation and gene co-expression) analysis, and protein–protein interaction network analysis.

Table 1. Characteristics of individual studies retrieved from GEO of NCBI for meta-analysis

	Number of samples			
GEO data set	NC	DD	Source	Platform
GSE31356	3	3	GEO of NCBI	Affymetrix Human Genome U133A array
E-GEOD-59746	2	2	ArrayExpress of EMBL-EBI	Affymetrix Human Genome U133 plus 2.0
E-GEOD-21221	6	6	ArrayExpress of EMBL-EBI	Illumina Sentrix Human-6 Expression BeadChip

NC normal control, DD dupuytren's disease, GEO gene expression omnibus, GSE gene expression series.

p-value of 1.0E-10 were *COL4A5*, *PRG4*, *NPTX2*, *SNX10*, *PLEKHA6*, *SERPINF2*, *SLC27A2*, *SIRPB1*, *LAG3*, *ADH1B*, *NOB3B*, *NEK9*, *AR*, *SLCO2A1*, *CXCL6*, and *HOXD13*.

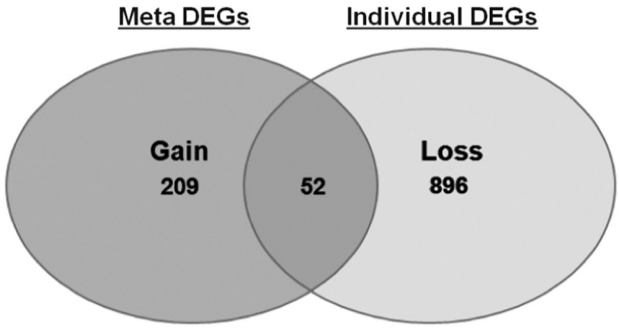

Figure 2. Differential gene expression obtained from meta-analysis for Dupuytren's disease. Venn diagram showing the number of significant associations between DEGs identified from the meta-analysis of multiple data sets (Meta-DE) and the individual analysis of each data set (individual-DE).

Among these, the up-regulated and down-regulated DEGs with the largest mean logFC were *RPS4Y1* (ribosomal protein S4, Y-linked 1) and *COL4A5* (collagen, type IV, alpha 5), respectively.

As shown in Figure 1B, the total 261 DEGs identified by the meta-analysis were studied by a systematic approach of various in silico analyses, including gene set functional enrichment analysis, gene expression network (include gene-regulation network and gene co-expression network) analysis, and PPI network analysis, to interpret the biological significance of the identified DEGs in Dupuytren's disease.

Functional enrichment analysis of the identified DEGs

The 261 DEGs identified by the meta-analysis were enriched based on biological process terms of GO categories and the KEGG pathway in DAVID (threshold of *p*-value ≤0.05) (Table 3). The most enriched terms under biological processes were "cell-cell signaling (GO: 0007267)," "defense response (GO:0006952)," and "endocrine system development (GO:0035270)." The significant

Table 2. The top 20 most strongly up- or down-regulated genes in the DEGs identified by meta-analysis

Enterz ID	Gene symbol	(–) Log_2FC	Adjusted *p*-value	Gene name
Up-regulated genes				
6,192	RPS4Y1	–4.69791433	<0.001	Ribosomal protein S4, Y-linked 1
9,086	EIF1AY	–4.365363117	<0.001	Eukaryotic translation initiation factor 1A, Y-linked
22,829	NLGN4Y	–3.573713829	<0.001	Neuroligin 4, Y-linked
7,545	ZIC1	–3.458926639	<0.001	Zic family member 1
10,481	HOXB13	–3.414739012	<0.001	Homeobox B13
57,502	NLGN4X	–3.389591655	<0.001	Neuroligin 4, X-linked
7,022	TFAP2C	–3.371684831	<0.001	Transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)
8,284	KDM5D	–2.864610713	<0.001	Lysine demethylase 5D
3,382	ICA1	–2.724068876	<0.001	Islet cell autoantigen 1
8,529	CYP4F2	–2.659292285	<0.001	Cytochrome P450 family 4 subfamily F member 2
10,866	HCP5	–2.523748206	<0.001	HLA complex P5 (nonprotein coding)
922	CD5L	–2.382866034	<0.001	CD5 molecule like
9,293	GPR52	–2.659774886	0.0125	G protein-coupled receptor 52
79,037	PVRIG	–2.468400154	0.013157895	Poliovirus receptor related immunoglobulin domain containing
3,250	HPR	–2.434383242	0.014285714	Haptoglobin-related protein
4,153	MBL2	–3.140796029	0.015	Mannose binding lectin 2
9,447	AIM2	–2.852645899	0.02037037	Absent in melanoma 2
4,283	CXCL9	–3.103790714	0.020689655	Chemokine (C-X-C motif) ligand 9
9,495	AKAP5	–2.385969866	0.04047619	A-kinase anchoring protein 5
415	ARSE	–2.596903888	0.047727273	Arylsulfatase E (chondrodysplasia punctata 1)
Down-regulated genes				
1,287	COL4A5	4.295625032	<0.001	Collagen type IV alpha 5
4,885	NPTX2	3.824155618	<0.001	Neuronal pentraxin II
29,887	SNX10	3.622452218	<0.001	Sorting nexin 10
22,874	PLEKHA6	3.520080415	<0.001	Pleckstrin homology domain containing A6
5,345	SERPINF2	3.096192826	<0.001	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium-derived factor), member 2
3,902	LAG3	2.832411702	<0.001	Lymphocyte activating 3
79,817	MOB3B	2.767421089	<0.001	MOB kinase activator 3B
367	AR	2.659081432	<0.001	Androgen receptor
91,754	NEK9	2.650549502	<0.001	NIMA-related kinase 9
6,578	SLCO2A1	2.591426035	<0.001	Solute carrier organic anion transporter family member 2A1
124	ADH1A	2.628873517	0.001724138	Alcohol dehydrogenase 1A (class I), alpha polypeptide
54,502	RBM47	2.599307626	0.002	RNA binding motif protein 47
125	ADH1B	2.959989261	0.002272727	Alcohol dehydrogenase 1B (class I), beta polypeptide
10,326	SIRPB1	2.763760401	0.002777778	Signal regulatory protein beta 1
6,286	S100P	2.85358264	0.002941176	S100 calcium-binding protein P
10,216	PRG4	4.215848302	0.003333333	Proteoglycan 4
11,001	SLC27A2	3.005231074	0.003571429	Solute carrier family 27 member 2
56,132	PCDHB3	2.917432407	0.003846154	Protocadherin beta 3
79,799	UGT2A3	2.821813459	0.023972603	UDP glucuronosyltransferase 2 family, polypeptide A3
3,198	HOXA1	2.596565951	0.030097087	Homeobox A1

Log_2FC = log_2 (class 1/class 2); FC = fold change; class1 = normal control; class2 = Dupuytren's disease.
Adjusted *p*-value = FDR-adjusted *p*-value.

Table 3. The top 10 GO and KEGG enrichment pathway of the total DEGs identified by meta-analysis

GO ID	GO term	Number of genes	*p*-value
GO:0007267	Cell–cell signaling	36	1.03E-11
GO:0006952	Defense response	33	1.58E-09
GO:0035270	Endocrine system development	9	1.10E-05
GO:0009952	Anterior/posterior pattern formation	12	1.14E-05
GO:0048608	Reproductive structure development	10	1.51E-04
GO:0006935	Chemotaxis	10	8.76E-04
GO:0048562	Embryonic organ morphogenesis	8	4.70E-03
GO:0006874	Cellular calcium ion homeostasis	9	7.75E-03
GO:0009725	Response to hormone stimulus	13	1.22E-02
GO:0050778	Positive regulation of immune response	7	2.56E-02

KEGG ID	KEGG pathway	Number of genes	*p*-value
hsa04514	Cell adhesion molecules (CAMs)	7	2.74E-02
hsa04060	Cytokine–cytokine receptor interaction	9	8.64E-02
hsa04062	Chemokine signaling pathway	6	2.29E-01
hsa04621	NOD-like receptor signaling pathway	3	2.96E-01
hsa04310	Wnt signaling pathway	4	4.96E-01

pathway overrepresented by KEGG enrichment analysis was "cell adhesion molecules (CAMs) (has 04514)." In addition, "cytokine-cytokine receptor interaction (has 04060)" and "chemokine signaling pathway (has 04062)" were also enriched.

Gene-regulation network analysis of the identified DEGs

To understand the regulation network for gene expression of the 261 DEGs that might directly affect Dupuytren's disease, we examined potential transcription factors and microRNAs that target the top 20 up-regulated and down-regulated DEGs in MSigDB (Table 4). The target gene sites of transcription factors such as NFAT, TCF3, MLLT7, and MAZ, which regulate DNA transcription into RNA, were clearly identified by the determined DEGs. The target gene sites of microRNA such as MIR-23A/B, MIR-384, MIR-508, and MIR-10A/10B, which regulate mRNA stability, were also identified by the DEGs we determined.

Gene-correlation network analysis of the top 20 up-regulated and down-regulated DEGs

We constructed a gene-correlation network composed of co-expression of 89 nodes and 352 edges by mapping the top 20 up-regulated and down-regulated DEGs into a very large set of functional interaction database in GeneMANIA webserver. The co-expression network was further divided into five functional modules by the fast-greedy algorithm of GLay plugin in Cytoscape, and the genes classified in each module were enriched by GO and KEGG pathway analysis by an online tool in DAVID (Figure 3). Among them, the significant functional modules were modules 1–3, which are highly interconnected by genes above 20 nodes. Enrichment by GO term "biological process" and KEGG pathway enrichment by genes of module 1 were "defense response (GO:0006952)," "cell adhesion (GO:0007155)," and "chemokine signaling pathway (has 04062)." GO term and KEGG pathway enrichment identified genes of module 2 as "response to wounding (GO:0009611)" and "complement and coagulation cascades (hsa 04610)." Similarly, enriched genes in module 3 were determined to be "chromosome organization (GO:0051276)" and "transition metal ion binding (GO:0046914)."

PPI network analysis of the top 10 up-regulated and down-regulated DEGs

We constructed a PPI network with 104 nodes and 174 edges by matching the top 10 up-regulated and down-regulated DEGs into a very large set of entries from the PPI database downloaded from the BIOGRID webserver (Figure 4A). The three hub proteins TFAP2C, SERPINF2, and HOXB13 in the PPI network were identified based on the density of nodes and *p*-value by ClusterONE plugin in Cytoscape (Figure 4B). While GO terms such as "regulation of cell cycle (GO:0051726)" and "regulation of transcription (GO:0009611)" were enriched by hub clusters of TFAP2C and HOXB13 of up-regulated DEGs, the GO terms such as "response to wounding (GO:0006355)" and "Response to steroid hormone (GO:0048545)" were enriched by hub cluster of SERPINF2 and COL4A5 of down-regulated DEGs.

DISCUSSION

The primary goal of this study was to determine more refined selection of several DEGs in Dupuytren's disease by a meta-analysis based on publicly available data sets. We identified 261 genes that were consistently expressed differentially in Dupuytrens disease (111 up-regulated and 150 down-regulated). Of these, 209 "gain" DEGs were identified in the meta-analysis only and not in individual studies, suggesting a high probability of them being novel biomarkers for the diagnosis of Dupuytren's disease.

Dupuytren's disease is a complex human disease resulting from a combination of hereditary and environmental factors. For the identified DEGs, understanding the comprehensive information for the topological position of the DEGs in molecular networks and the FC and *p*-value of each DEG

Table 4. The potential transcription factor and micro RNA that regulate DEGs identified by meta-analysis

Transcription factor	Target sequence	Number of genes	*p*-value
NFAT	TGGAAA	45	4.22E-16
TCF3	CAGGTG	50	5.26E-15
MLLT7	TTGTTT	44	3.86E-14
MAZ	GGGAGGRR	43	4.34E-12
microRNA	Target sequence	Number of genes	*p*-value
MIR-23A/23B	AATGTGA	10	1.60E-04
MIR-384	CTAGGAA	4	4.94E-04
MIR-508	TACAATC	4	5.24E-04
MIR-10A/10B	ACAGGGT	5	7.12E-04

are valuable parameters to evaluate the biological significance of these DEGs in Dupuytren's disease. Therefore, we approached the possible effect of the identified DEGS on Dupuytren's disease from multiple viewpoints by systematic analyses of GO and KEGG enrichment, gene expression network, and PPI network.

Functional enrichment analysis of 261 DEGs depending on GO biological process terms and KEGG pathway revealed that the DEGs were mainly classified under "cellular signaling system (GO:0007276, GO:0006874, GO:0009275, hsa04514, hsa04621, and hsa04310)," "immune response (GO:0006952, GO:0050778, hsa04060, and hsa04062)," and "development (GO:0035270, GO:0048608, and GO:0048562)," all of which are generally associated with most diseases. With the identified DEGs, exploring the potential regulatory elements such as transcription factors and their target genes may help reveal the pathogenesis of Dupuytren's disease. Gene regulation analysis showed that the DEGs significantly shared target sites of transcription factors and microRNA. Recently, it has been shown that the miR-23/27/24 family is involved in angiogenesis and endothelial apoptosis in cardiac ischemia and retinal vascular development. The miR-10 family has attracted attention because of co-expression and dysregulation of miR-10 within the *Hox* gene series of developmental regulators. The DEGs with common target sites of these microRNAs were enriched by "pattern species process (GO:0007389)" and "regulation of transcription (GO:0006355)," implying that the gene regulation network of the DEGs is closely

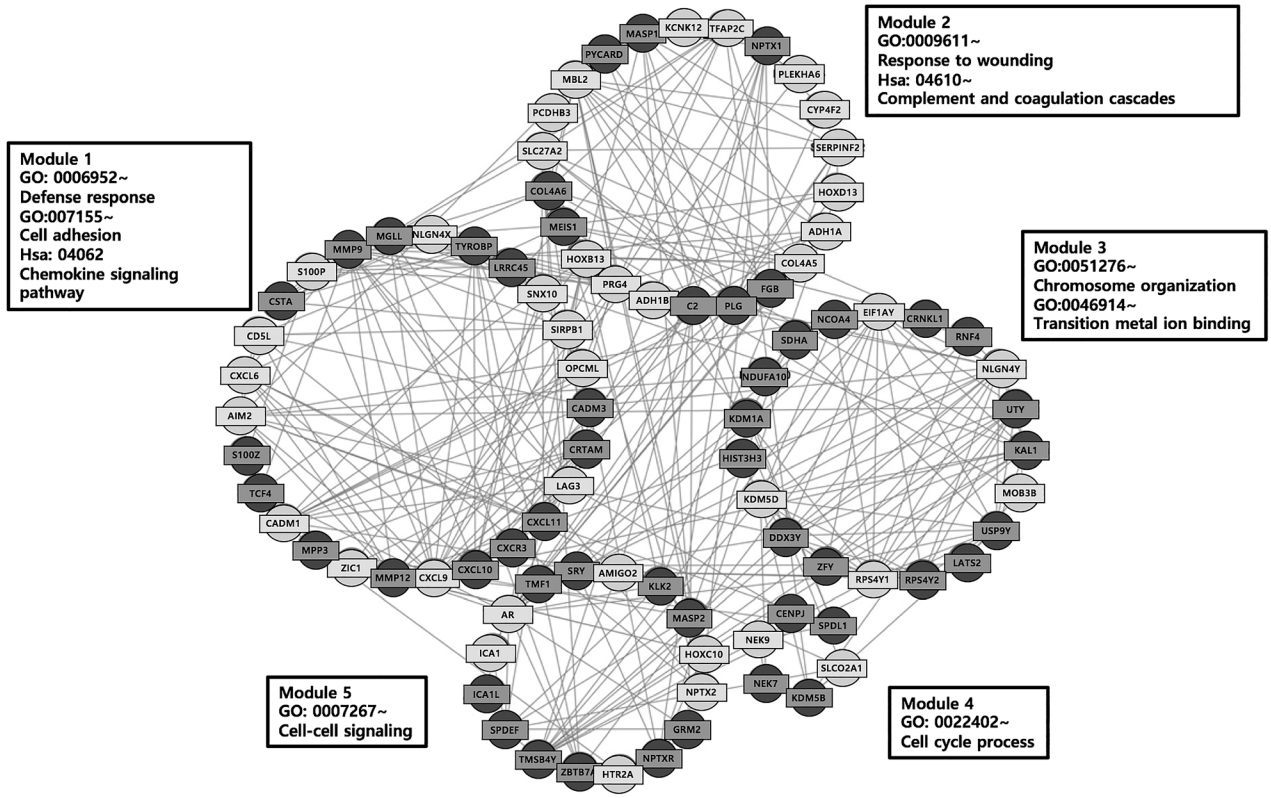

Figure 3. Gene-correlation network of the top 20 up- and down-regulated DEGs. From gene co-expression network of top 20 up- and down-regulated DEGs constructed in GeneMANIA online program, the functional modules were clustered by GLay pulgin in cytoscape. The node and edge of gene co-expression network stand for genes with the identified DGEs and interaction of the genes, respectively. In the network, the light gray nodes signify the top 20 up- and down-regulated DEGs identified by meta-analysis and the dark gray nodes signify the additional genes obtained by gene co-expression network analysis in GeneMANIA.

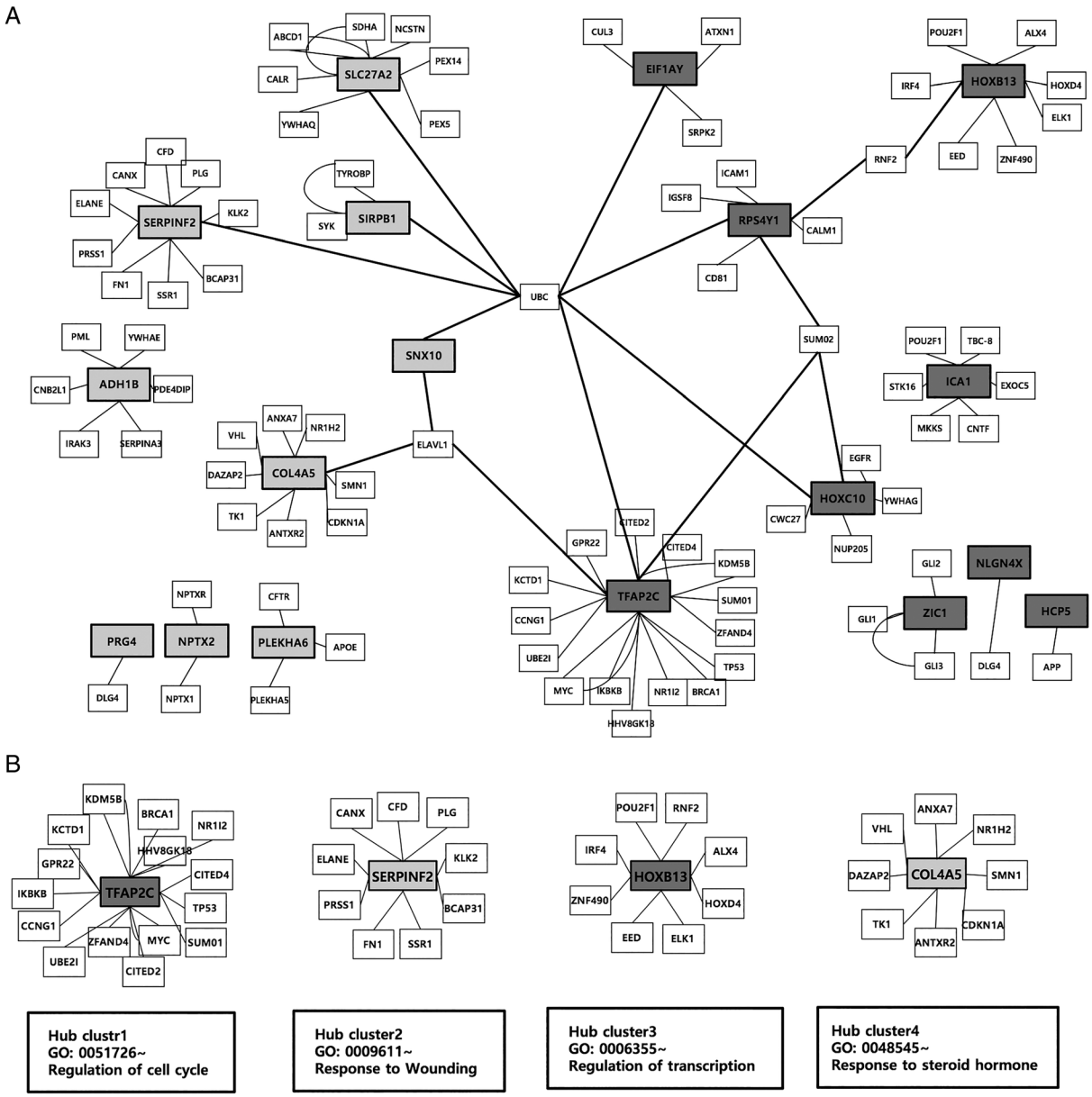

Figure 4. Protein–protein interaction network of the top 10 up- and down-regulated DEGs. (A) Protein–protein interaction network constructed by mapping the top 10 up- and down-regulated DEGs into databases downloaded from BioGRID. (B) Hub clusters that are highly connected with the neighbor proteins, in accordance with identification of ClusterONE plugin in Cytoscape. The node and edge of PPI network stand for protein encoding genes with the identified DEGs and interaction of the proteins, respectively. In the network, the dark gray nodes signify the top 10 up-regulated DEGs identified by meta-analysis and the red nodes signify the top 10 down-regulated DEGs. The light gray nodes in the network signify the additional proteins obtained by the PPI network analysis.

associated with the common inherited connective tissue disease such as Dupuytren's contracture. In case of transcription factors, the NFAT family plays a central role in the regulation of cytokine gene expression during the immune response. NFAT functions have been extensively explored in lymphocyte activation and differentiation. The Lef/Tcf-family transcription factor Tcf3 has important roles in development, stem cell function, and malignancy. During neocortical development, Tcf3 maintains the neural stem cell population by repressing Wnt-β-catenin signaling. The DEGs with common target sites of these transcription factors were enriched by "cell-cell signaling (GO:0007267)" and "regionalization (GO:0003002)," implying the possibility of somewhat similar results of microRNA analysis in Dupuytren's disease. In improving the understanding of the pathogenic mechanisms and novel therapeutic targets for the

treatment of Dupuytren's disease, it is more important to evaluate the functional enrichment of the tight coupling modules in the network rather than a literature research of the individual DEGs. We constructed gene co-expression network and PPI network of the identified DEGs at the molecular level, clustered the distinctive modules comprising DEGs showing strong interaction co-located to each other in the network, and analyzed the biological functions of each module by GO and KEGG pathway enrichment. Modules 1–3 were identified as the significantly important functional cluster in the gene co-expression network of gene level by the GO and KEGG pathway enrichment, in the way that the biological functions of three modules are largely involved in genetic disorder, aberrant wound healing, and consequent defense reaction of immune system, which are the representative characteristics of Dupuytren's disease. In parallel with that analysis, clustering and enrichment of functional modules were performed in the PPI network at the protein level to identify hub proteins with a high degree of interaction. In many studies, the hub node has been found to have a large effect on the network formed by constituents of an organic system and play important functions in the maintenance of the system. From PPI network analysis, the identified hub genes were *TFAP2C*, *SERPINF2*, *HOXB13*, and *COL4A5* which are associated with GO terms "regulation of cell cycle (GO:0051726)," "response to wounding (GO:0009611)," "regulation of transcription (GO:0006355)," and "Response to steroid hormone." The crucial point to note from the systemically coordinated analysis of two networks is that all the four hub genes in the PPI network constitute module 2 of functional modules in the gene co-expression network. To the best of our knowledge, most DEGs of the top 20 most strongly up-regulated or down-regulated genes are not validated to date, warranting further investigation. Among the DEGs in three functional modules of gene co-expression network, only *ZIC1*, known as an upstream regulator of several Wnt-genes, was subjected to pathological testing using monoclonal antibodies for Zic1,[19,20] while the down-regulation of *PRG4* was shown by two studies using real time RT-PCR.[21,22] According to Satish et al., PRG4 mRNA levels in Dupuytren's cord fibroblasts was less than 10% of that in control fibroblasts with real-time RT-PCR. In addition, using qRT-PCR, they showed the down-regulation of three candidate genes (*PRG4*, FBLN-1 transcript variant D, and type XV collagen alpha 1 chain), which are components of the extracellular matrix.[22]

Interestingly, *COL4A5* (collagen, type IV, alpha 5), identified as the most significantly down-regulated gene among total 261 DEGs was affiliated with module 2 in gene co-expression network. This gene encodes the alpha5(IV) chain of type IV collagen. This chain combines with two other types of alpha (IV) chains (the alpha3 and alpha4 chains) to make a complete type IV collagen molecule. COL4A5 protein plays an especially important role in the basement membranes of the kidney, inner ear, and eye. Mutations in *COL4A5* are responsible for X-linked Alport's syndrome, which is a basement membrane disorder characterized clinically by hereditary nephropathy. It has been reported that appearance of the myofibroblastic phenotype in Dupuytren's disease is associated with collagen type IV.[23] Very recently, Kang et al. reported the down-regulation of collagen synthesis and matrix metalloproteinase expression in myofibroblasts from Dupuytren's nodule using adenovirus-mediated relaxin gene therapy. Myofibroblastic cells with Ad-RLN demonstrated a 22% and 48% reduction in collagen I and III mRNA expression, respectively, a 50% decrease in MMP-1, 70% decrease in MMP-2, 80% decrease in MMP-9, and a 15% reduction in MMP-13 protein expression compared with cultures with viral control and saline control.[24] Dupuytren's contracture is characterized by excessive collagen deposition, which appears as cords causing an extension deficit. In Dupuytren's contracture, there is an increase in the ratio of type III to type I collagen.

RPS4Y1 (logFC = −4.6834) in module 3 was the most significantly up-regulated gene of the total DEGs identified by the meta-analysis. It is a sex-linked ribosomal protein, encoding ribosomal protein S4, a component of the 40S subunit. Besides its role as a sex-linked ribosomal protein, very little is known about the pathophysiological function of these gene alterations. Although controversial, it has been reported that haploinsufficiency of the ribosomal protein S4 genes is involved in Turner syndrome, a sex-linked genetic disorder.[25] Interestingly, Dupuytren's contracture occurs with a higher frequency in males than in females. Thus, further studies are needed to determine whether *RPS4Y1* is involved in the pathogenesis of Dupuytren's disease.

Additionally, we conducted network topology analysis for gene-correlation network data sets to identify essential genes to be involved in Dupuytren's disease (Supporting Information). For the purpose of identifying essential genes through network topology analysis, a Cytoscape plugin, which integrates calculation, evaluation, and visualization analysis for multiple centrality measures, called CytoNCA, was performed.[26] We created gene-correlation network by mapping top 20 up- and down-regulated DEGs and performed six centrality measurement (Supporting Information Table S1). We tabulated top 10 genes by each centrality measure and found that *MBL2*, *SERPINF2*, *SLCO2A1*, and *TTPA* were ranked the highest among centrality measures (Supporting Information Table S2). The mannose-binding lectin 2 (*MBL2*) gene was reported to be related to amplify high-producing genotypes of transforming growth factor beta 1 (*TGFB1*) in lung disease.[27] Badalamente et al. studied that *TGFB1* showed an intense intracellular marking pattern in Duputren's samples and thus, *MBL2* may be considered playing an important role in Dupuytren's disease, interacting with *TGFB*.[28] Serpin family F member 2 (*SERPINF2*) gene was not reported to associated with Dupuytren's disease yet, but it was identified as one of hub genes in PPI network, which implies that *SERPINF2* could be a veiled candidate gene to have Dupuytren's disease. Solute carrier organic anion transporter family, member 2A1 (*SLCO2A1*) gene was identified as a candidate in fibroblasts derived from patients with Dupuytren's disease.[29] Alpha tocopherol transfer protein (*TTPA*) encodes a protein binding a form of vitamin E with high selectivity and affinity and playing an important role in regulating vitamin E levels. Since vitamin E deficiency can cause dry skin, mutation of TTPA may be one of critical factors to outbreak Dupuytren's disease.

In conclusion, our meta-analysis revealed previously unknown genes or pathways possibly involved in Dupuytren's disease and suggested new strategies for the treatment of Dupuytren's disease.

ACKNOWLEDGMENT

This article was supported by Konkuk University in 2019.

AUTHOR CONTRIBUTIONS

S.Y.K. and T.H. P conceived and designed the study. D.K. and Y.S.L downloaded and analyzed the data. D.K., T.H.P., and S.Y.K. contributed to write the main manuscript text and prepare figures and tables. All authors reviewed manuscript.

REFERENCES

1. Gudmundsson KG, Arngrímsson R, Sigfússon N, Björnsson A, Jónsson T. Epidemiology of Dupuytren's disease: clinical, serological, and social assessment. The Reykjavik Study. *J Clin Epidemiol.* 2000; 53: 291–6.
2. Thoma A, Kaur MN, Ignacy TA, Levis C, Martin S, Duku E, et al. Health-related quality of life in patients undergoing palmar fasciectomy for Dupuytren's disease. *Plast Reconstr Surg.* 2014; 133: 1411–9.
3. Van Dam H, Elliot D. Avoidance of recurrence of CRPS Type 1 in individuals requiring further surgery for Dupuytren's contracture. *J Plast Reconstr Aesthet Surg.* 2014; 67: 878–9.
4. Mehta S, Belcher HJCR. A single-centre cost comparison analysis of collagenase injection versus surgical fasciectomy for Dupuytren's contracture of the hand. *J Plast Reconstr Aesthet Surg.* 2014; 67: 368–72.
5. Erne HC. Downgrading severe stages of Dupuytren's contracture to simplify partial aponeurectomy using percutaneous needle fasciotomy. *Plast Reconstr Surg.* 2014; 133: 79e–80e.
6. Verhoekx JSN, Mudera V, Walbeehm ET, Hovius SER. Adipose-derived stem cells inhibit the contractile myofibroblast in Dupuytren's disease. *Plast Reconstr Surg.* 2013; 132: 1139–48.
7. Branford OA, Davis M, Schreuder F. The circumflex scapular artery perforator flap for palm reconstruction in a recurrent severe case of Dupuytren's disease. *J Plast Reconstr Aesthet Surg.* 2009; 62: e589–91.
8. Ali SN, McMurtrie A, Rayatt S, Roberts JO. Ulnar-based skin flap for Dupuytren's fasciectomy. *Scand J Plast Reconstr Surg Hand Surg.* 2006; 40: 307–10.
9. Augoff K, Taboła R, Kula J, Gosk J, Rutowski R. Epidermal growth factor receptor (EGF-R) in Dupuytren's disease. *J Hand Surg Br.* 2005; 30: 570–3.
10. Baxter H, Schiller C, Johnson LH, Whiteside JH, Randall RE. Cortisone therapy in Dupuytren's contracture. *Plast Reconstr Surg (1946).* 1952; 9: 261–73.
11. Eaton C. Evidence-based medicine: Dupuytren contracture. *Plast Reconstr Surg.* 2014; 133: 1241–51.
12. Shih B, Tassabehji M, Watson JS, Bayat A. DNA copy number variations at chromosome 7p14.1 and chromosome 14q11.2 are associated with dupuytren's disease: potential role for MMP and Wnt signaling pathway. *Plast Reconstr Surg.* 2012; 129: 921–32.
13. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock REW. INMEX--a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res.* 2013; 41: W63–70.
14. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 2004; 573: 83–92.
15. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics.* 2006; 22: 2825–7.
16. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010; 38: W214–20.
17. Su G, Kuchinsky A, Morris JH, States DJ, Meng F. GLay: community structure analysis of biological networks. *Bioinformatics.* 2010; 26: 3135–7.
18. Wu H, Gao L, Dong J, Yang X. Detecting overlapping protein complexes by rough-fuzzy clustering in protein-protein interaction networks. *PLoS ONE.* 2014; 9: e91856.
19. Degreef I, De Smet L, Sciot R, Cassiman J-J, Tejpar S. Immunohistochemical evidence for Zic1 coexpression with beta-catenin in the myofibroblast of Dupuytren disease. *Scand J Plast Reconstr Surg Hand Surg.* 2009; 43: 36–40.
20. Merzdorf CS, Sive HL. The zic1 gene is an activator of Wnt signaling. *Int J Dev Biol.* 2006; 50: 611–7.
21. Shih B, Wijeratne D, Armstrong DJ, Lindau T, Day P, Bayat A. Identification of biomarkers in Dupuytren's disease by comparative analysis of fibroblasts versus tissue biopsies in disease-specific phenotypes. *J Hand Surg Am.* 2009; 34: 124–36.
22. Satish L, LaFramboise WA, O'Gorman DB, Johnson S, Janto B, Gan BS, et al. Identification of differentially expressed genes in fibroblasts derived from patients with Dupuytren's Contracture. *BMC Med Genomics.* 2008; 1: 10.
23. Berndt A, Kosmehl H, Katenkamp D, Tauchmann V. Appearance of the myofibroblastic phenotype in Dupuytren's disease is associated with a fibronectin, laminin, collagen type IV and tenascin extracellular matrix. *Pathobiology.* 1994; 62: 55–8.
24. Kang Y-M, Choi Y-R, Yun C-O, Park J-O, Suk K-S, Kim H-S, et al. Down-regulation of collagen synthesis and matrix metalloproteinase expression in myofibroblasts from Dupuytren nodule using adenovirus-mediated relaxin gene therapy. *J Orthop Res.* 2014; 32: 515–23.
25. Fisher EM, Beer-Romero P, Brown LG, Ridley A, McNeil JA, Lawrence JB, et al. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell.* 1990; 63: 1205–18.
26. Tang Y, Li M, Wang J, Pan Y, Wu F-X. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *BioSystems.* 2015; 127: 67–72.
27. Dorfman R, Sandford A, Taylor C, Huang B, Frangolias D, Wang Y, et al. Complex two-gene modulation of lung disease severity in children with cystic fibrosis. *J Clin Invest.* 2008; 118: 1040–9.
28. Badalamente MA, Sampson SP, Hurst LC, Dowd A, Miyasaka K. The role of transforming growth factor beta in Dupuytren's disease. *J Hand Surg Am.* 1996; 21: 210–5.
29. Liu H, Yin W, Liu B, Liu Y, Guo B, Wei Z. Screening of candidate genes in fibroblasts derived from patients with Dupuytren's contracture using bioinformatics analysis. *Rheumatol Int.* 2015; 35: 1343–50.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1: Supporting Information.
Supplementary Figure 1