Received: 21 November 2018 | Revised: 18 February 2019 | Accepted: 25 February 2019

DOI: 10.1111/odi.13093

ORIGINAL ARTICLE

ORAL DISEASES WILEY

# Exomic and transcriptomic alterations of hereditary gingival fibromatosis

Seong Kyu Han[1] | Jungho Kong[1] | Sanguk Kim[1,2] | Jae-Hoon Lee[2] | Dong-Hoo Han[2]

[1]Department of Life Sciences, Pohang University of Science and Technology, Pohang, Korea

[2]Department of Prosthodontics, College of Dentistry, Yonsei University, Seoul, Korea

**Correspondence**
Sanguk Kim, Department of Life Sciences, Pohang University of Sciences and Technology, Pohang, Korea.
Email: sukim@postech.ac.kr
and
Jae-Hoon Lee and Dong-Hoo Han, Department of Prosthodontics, College of Dentistry, Yonsei University, Seoul, Korea.
Emails: jaehoon115@yuhs.ac; donghoohan@yuhs.ac

**Funding information**
National Research Foundation of Korea, Grant/Award Number: 2014R1A1A2055755 and 2018R1A2B6002657

## Abstract

**Objective:** Hereditary gingival fibromatosis (HGF) is a rare oral disease characterized by either localized or generalized gradual, benign, non-hemorrhagic enlargement of gingivae. Although several genetic causes of HGF are known, the genetic etiology of HGF as a non-syndromic and idiopathic entity remains uncertain.

**Subjects and methods:** We performed exome and RNA-seq of idiopathic HGF patients and controls, and then devised a computational framework that specifies exomic/transcriptomic alterations interconnected by a regulatory network to unravel genetic etiology of HGF. Moreover, given the lack of animal model or large-scale cohort data of HGF, we developed a strategy to cross-check their clinical relevance through in silico gene–phenotype mapping with biomedical literature mining and semantic analysis of disease phenotype similarities.

**Results:** Exomic variants and differentially expressed genes of HGF were connected by members of TGF-β/SMAD signaling pathway and craniofacial development processes, accounting for the molecular mechanism of fibroblast overgrowth mimicking HGF. Our cross-check supports that genes derived from the regulatory network analysis have pathogenic roles in fibromatosis-related diseases.

**Conclusions:** The computational approach of connecting exomic and transcriptomic alterations through regulatory networks is applicable in the clinical interpretation of genetic variants in HGF patients.

KEYWORDS
hereditary gingival fibromatosis, multi-omics approach, TGF-beta signaling

## 1 | INTRODUCTION

Hereditary gingival fibromatosis (HGF) is a rare oral disease characterized by either localized or generalized gradual, benign, non-hemorrhagic enlargement of gingivae. The enlargement of gingival tissue varies in degree; in its severe form, both maxillary and mandibular gingivae cover the crowns of the teeth, resulting in displacement of teeth, disocclusion, delayed eruption, prominent lips, speech impediments, and periodontal and esthetic problems (Chaurasia, 2014). Hereditary gingival fibromatosis usually begins during or after eruption of permanent teeth and is infrequently observed at the time of eruption of deciduous teeth or birth. The current clinical treatment of choice for HGF patients is merely a surgical removal of affected tissues, however, the recurrence of gingival enlargement is frequent after the surgery (Ball, 1941; Heath & Tomes, 1879; Rapp, Nikiforuk, Donohue, & Williams, 1955; Zackin & Weisberger, 1961). Therefore, an understanding of the etiology underlying HGF is necessary to construct the therapeutic strategy with high clinical efficacy.

While gingival fibromatosis may occur as a drug-induced, hereditary, or idiopathic entity, HGF is commonly a genetic disease that may exist as an isolated condition or part of a syndrome (Häkkinen & Csiszar, 2007). Previous reports of HGF demonstrated the mode of inheritance to be autosomal dominant in most cases and autosomal recessive in a few cases (1994, 1994; Hart, Pallos, & Bozzo, 2000; Jorgenson & Cocker, 1974). Notably, idiopathic HGF is a clinically and genetically heterogeneous disorder with unclear evidence of genetic transmission in family histories; thus, its etiology and genetic mechanism have not been elucidated. To unravel the genetic basis of HGF, past studies have mainly used linkage analysis, candidate gene approaches, and whole-exome sequencing. As a result, chromosomes 2p21-p22 (*GINGF*), 2p13-p16, 2p22.3-23 (*GINGF3*), 5q13-q22 (*GINGF2*), 4q21, 4q, son of sevenless gene (*SOS1*), and RE1-silencing transcription factor (*REST*) have been associated with non-syndromic HGF (Bayram, White, & Elcioglu, 2017; DeAngelo, Murphy, Claman, Kalmar, & Leblebicioglu, 2007; Sah, Chandra, & Kaur, 2015). However, these genetic methods have not fully revealed the mechanisms underlying idiopathic HGF. It is because that (a) the gene–phenotype relationship of HGF is entangled with complex developmental traits comorbid with HGF and (b) small patient cohorts of idiopathic HGF hinders sufficient statistical power to pinpoint genetic variations indeed related to HGF.

We hypothesized that the discovery of genetic variants related to transcriptomic alterations can help to understand the genetic etiology of HGF. Genetic variants changing transcriptomic profiles can be clinically relevant because the transcriptomic profile is an "intermediate molecular phenotype," because disease phenotypes may arise from gene expression changes in disease lesions (Burga & Lehner, 2013; Lage, Hansen, & Karlberg, 2008). In addition, given the lack of animal model or large-scale cohort data of HGF, we regarded that a regulatory network analysis is the appropriate approach to find genetic variants that have causality on transcriptomic changes on such rare disease. A regulatory network represents the link between the transcriptional regulators and target genes, wherein the regulator affects target gene expressions via binding of its regulatory elements activated in a particular tissue or condition of interest (Koch, Konieczka, & Delorey, 2017; Thompson, Regev, & Roy, 2015). Recently, reproducible and high-resolution regulatory networks accounting for a particular tissue or disease state are available from advances of regulatory elements mapping and transcriptomic profiling on numerous samples (Andersson, Gebhard, & Miguel-Escalada, 2014; Forrest, Kawaji, & Rehli, 2014; Lizio, Harshbarger, & Shimoji, 2015; Maher, 2012; Marbach et al., 2016).

In this study, we analyzed both exomic and transcriptomic alterations between idiopathic HGF patients and controls, and then devised a computational framework that specify exome and transcriptomic alterations interconnected by the regulatory network to unravel the genetic etiology of idiopathic HGF. We discovered exomic variants on transcriptional regulators, with perturbed expression of their targeted genes identified from regulatory networks and confirmed by RNA-seq data of patients. Consequently, the variants on *ETV7*, *SMAD4*, *ELK4*, *SOX30*, and *ALX4* were turned out to be connected with and their regulated targets and functionally associated genes involved in TGF-β/SMAD signaling pathway and craniofacial development processes. Moreover, using the in silico gene–phenotype mapping with biomedical literature mining and semantic analysis of disease phenotype similarities, we discovered that those genes and pathways account for the molecular mechanism of fibroblast overgrowth mimicking HGF. Our findings suggest that the computational approach of connecting exomic and transcriptomic alteration through regulatory network analysis is useful for the clinical interpretation of genetic variants found in HGF patients.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients and samples

The case (HGF) participants in this study are the same as those in our previous study (Hwang, Kim, & Kang, 2017). The siblings (a 13-year-old boy and a 9-year-old girl) had enlarged gingivae whereas neither their parents nor grandparents ever had. The patients first presented to a pediatric dental clinic with chief complaint of delayed tooth eruption and lip closer incompetency when they were 2.3 and 6 years old. An intraoral examination of the girl at first dental visit revealed generalized gingival overgrowth involving attached gingiva, marginal gingiva, and interdental papilla. The posterior areas at both arches were particularly severely deformed by a large amount of gingival tissue that was sufficient to inhibit tooth eruption and impair mastication. Moreover, to exclude HGF cases confounded by poor plaque control, we followed up the patients and validated that oral hygiene had been good status during follow-up visits. Extraorally, the patient was unable to close her lips because of protrusion of the enlarged gingival tissue (Figure 1a). Clinical examination also revealed partial eruption of the teeth in the maxillae and mandible. Conventional gingivectomy localized to the maxillary anterior area was performed to facilitate eruption of the primary incisors. The removed gingival tissue was assessed histologically, and a diagnosis of gingival fibromatosis was made. The epithelium had a normal structure, with rete pegs that penetrated deep into the connective tissue. There appeared to be more collagen fiber bundles than in normal tissue, but no other unusual histologic features (Supporting Information Figure S1). Even though a parent stated no significant medical condition, siblings also had a minor retarded type facial deformation unlike their parents. We provided the details of patient selection and background in our previous research paper (Hwang et al., 2017). We obtained two fibromatic gingiva specimens from the patients for RNA sequencing. Two normal gingivae were obtained from age- and gender-matched controls (a male and a female who underwent surgical gingival resection for extraction of a supernumerary tooth and odontoma, mean age of 8.7 years). We obtained written informed consent from the parent of the children (all participants) and performed clinical investigation in accordance with the Declaration of Helsinki. The Institutional Review Board of Yonsei University College of Dentistry (Yonsei IRB No. 2-2015-0024) approved this study. The accession number for the RNA sequencing data reported in this paper is SRA: SRP154276.

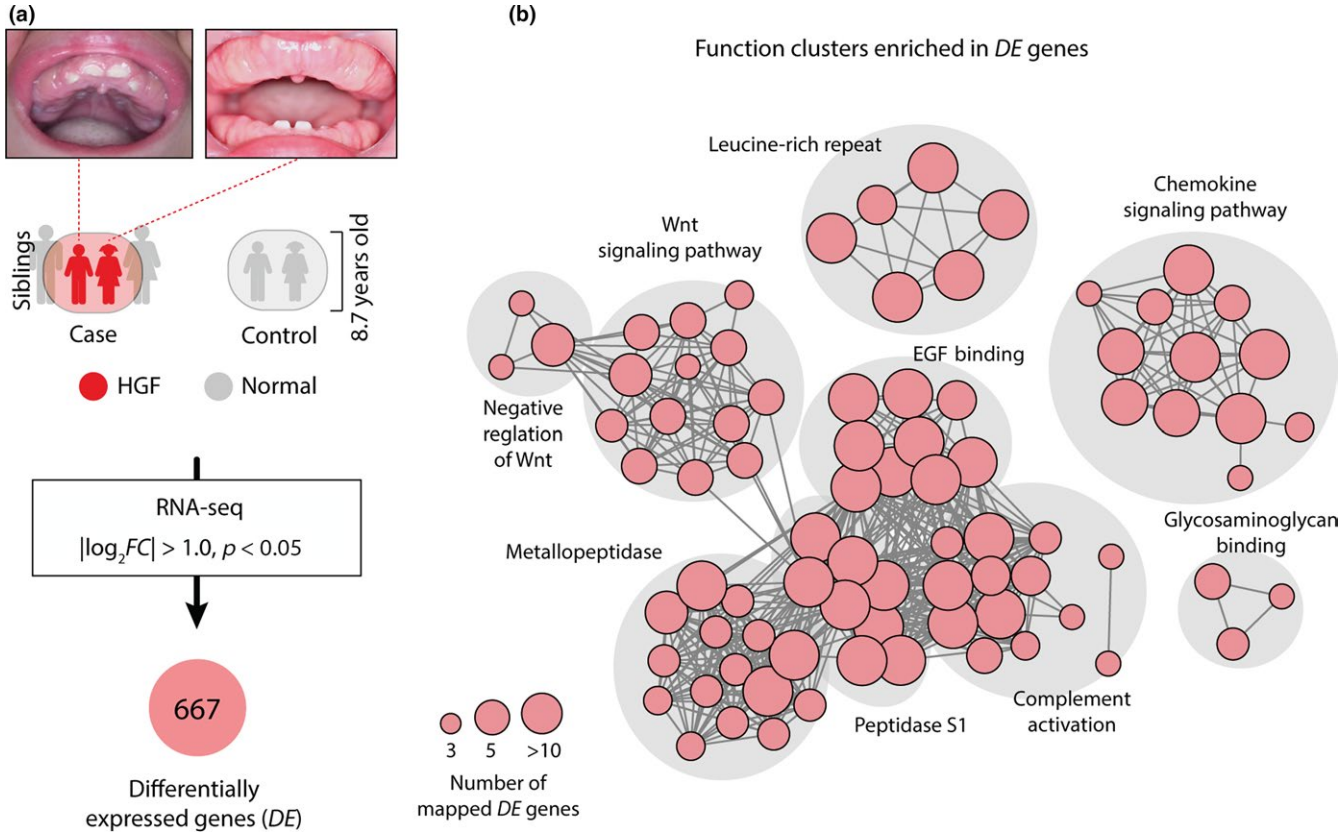


**FIGURE 1** Transcriptome analysis of hereditary gingival fibromatosis (HGF) patients. (a) Experimental scheme of RNA sequencing on HGF patients and control groups yielding extraction of differentially expressed genes (*DE*). Intraoral photographs of the two patients, taken at initial visit, show generalized gingival enlargement involving both the maxillary and mandibular arches. (b) Analysis of gene function enrichment and construction of a functional network of *DE* genes. Each cluster is indicated as a pink circle, in which nodes represent enriched clusters of gene functions. Node size represents the number of genetic variants in each functional group. Each cluster is described with a representative term, different terms being linked if they share more than three genes

## 2.2 | mRNA sequencing and mapping reads

RNA purity was determined by assaying 1 μl of total RNA extract on a NanoDrop8000 spectrophotometer. Total RNA integrity was examined using an Agilent Technologies 2,100 Bioanalyzer with an RNA integrity number (RIN) greater than eight. mRNA sequencing libraries were prepared according to the manufacturer's instructions (Illumina TruSeq RNA Prep Kit v2). mRNA was purified and fragmented from total RNA (1 μg) using poly-T oligo-attached magnetic beads with two rounds of purification. RNA fragments primed with random hexamers were reverse transcribed into first strand cDNA using reverse transcriptase and random primers. The RNA template was removed, and a replacement strand was synthesized to generate double-stranded cDNA. End repair, A-tailing, adaptor ligation, cDNA template purification, and the enrichment of purified cDNA templates using PCR were performed. The quality of the amplified libraries was verified by capillary electrophoresis (Bioanalyzer, Agilent). After performing qPCR using SYBR Green PCR Master Mix (Applied Biosystems), libraries index tagged in equimolar amounts in the pool were combined. Cluster generation occurred in the flow cell on the cBot automated cluster generation system (Illumina). The flow cell was loaded on the HISEQ 2500 sequencing system (Illumina), and sequencing performed with 2 × 100 bp read length.

## 2.3 | Identification and functional analysis of differentially expressed (*DE*) genes

The reads were mapped to the reference genome (UCSC, *hg19*) by TopHat (v2.0.13) (Kim et al., 2013). The aligned results from TopHat were added to Cuffdiff (v2.2.0) to detect DEGs (Trapnell, Roberts, & Goff, 2012). For library normalization and dispersion estimation, geometric and pooled methods were applied. For ontology analysis, up- and downregulated (twofold change and significance $p < 0.05$) genes were extracted and applied to DAVID (Jiao, Sherman, & Huang, 2012) as an input in order to acquire a comprehensive set of functional annotations. Biological process (BP) and molecular function (MF) categories of Gene Ontology (direct mapping only), the Biocarta and KEGG pathways, PFAM, and InterPro from DAVID were selected to annotate the gene groups enriched by DEGs.

## 2.4 | Network connection analysis between *EV* and *DE* genes

To map the *EV*-to-*DE* regulatory connection, we used the gingival fibroblast regulatory network associated with HGF pathology and lesions

(Marbach et al., 2016). We found a link with a directional correspondence, consisting of a source node and a target node which map to an *EV* and a *DE* gene, respectively. The regulatory connections with a link confidence score >0.05 were chosen following Marbach et al., 2016.

To expand the genes with regulatory connections based on functional association, we used STRING version 10 with a confidence score >900 (Szklarczyk, Franceschini, & Wyder, 2015). First, we constructed a functional association subnetwork (Supporting Information Figure S5 and Table S6) with a subset of *EV* and *DE* nodes which were interconnected. We then mapped genes already mapped to the *EV*-to-*DE* regulatory connections into the subnetwork and expanded into the first neighbors of the genes in the subnetwork. Finally, the *EV*-to-*DE* network was constructed by integrating the regulatory connections, genes, and expanded functional associations.

### 2.5 | Identification of HGF symptom-associated genes

To find symptom-associated genes, we used DisGeNET (Piñero, Queralt-Rosinach, & Bravo, 2015), which provides gene and symptoms connections from the integration of expert-curated databases with text-mined data and covers information on Mendelian and complex diseases. To map the disease and symptoms, we used human phenotype ontology (HPO) mapping of disease annotations covered by Online Mendelian Inheritance in Men (OMIM) (Amberger, Bocchini, Schiettecatte, Scott, & Hamosh, 2015; Groza, Köhler, & Moldenhauer, 2015). Because DisGeNET uses Unified Medical Language System (UMLS) coding to annotate symptoms, we converted the HPO to UMLS code following the Supporting Information Appendix S1 of Zhou, Menche, Barabási, and Sharma (2014).

### 2.6 | Phenotypic connection analysis of HGFs

To calculate the phenotypic similarity of diseases mapped in OMIM, we used MimMiner (van Driel, Bruggeman, Vriend, Brunner, & Leunissen, 2006), which precalculated a matrix of phenotypic similarity scores between disease pairs and quantified the co-occurrence of MeSH terms with reference to the OMIM clinical synopsis. Specifically, to calculate the phenotypic similarity score with HGF, we averaged the pairwise phenotypic similarity scores with 4 HGF diseases mapped in OMIM (Supporting Information Table S5). To quantify the disease comorbidity, we compared the co-occurrence of two diseases to random expectation in the human patient population (Hidalgo, Blumm, Barabási, & Christakis, 2009). The relative risk (RR) of diseases *i* and *j* is given by the following:

$$\mathrm{RR} = \frac{C_{ij}}{C^*_{ij}},$$

where $C_{ij}$ is the number of patients who had both disease *i* and disease *j*, and $C^*_{ij}$ is equal to $I_i I_j/N$, which represents random expectation. $I_i$ is the incidence of disease *i*. *N* is the total number of patients (13,039,018) in the Medicare record. To find the diseases caused by the mutations of a particular gene, we used the manually curated relationship data between human Mendelian disorders and genetic variants provided by OMIM (Amberger et al., 2015).

## 3 | RESULTS

### 3.1 | Transcriptomic alteration under HGF conditions

To observe gene expression changes in HGF patients, we performed an RNA sequencing analysis of the enlarged gingiva from two HGF siblings and normal ones from age- and gender-matched controls (see Materials and methods; Figure 1a and Supporting Information Figure S1). The transcriptome data of each individual yielded, on average, 62.1 million paired-end reads mapped to the human genome, with ~90% mapping to exons covered by the RefSeq Gene database (Supporting Information Table S1 for quality inspection of RNA-seq data). To find those genes differentially expressed between patients and controls, we quantified genes whose expression changed over twofold ($|\log_2 FC| > 1.0$). We employed a flexible cutoff of testing significance at $p < 0.05$ because the small number of sequenced samples ($N = 4$) entirely lowered the statistical power of the significance test; a stringent cutoff would miss many genes whose expression changes might be associated with HGF (Supporting Information Figure S2). Finally, the transcriptomic comparison of HGF patients and controls yielded 667 differentially expressed (*DE*) genes (Supporting Information Table S2).

We discovered functional groups enriched in *DE* genes that might affect the pathogenesis of HGF (Figure 1b and Supporting Information Table S3). To find these groups, we performed functional annotation clustering of *DE* genes (see Materials and methods). Specifically, diverse functional groups involved in the cell growth signaling pathway were associated with *DE* genes. For instance, EGF-binding proteins as well as the *W*nt signaling pathway and its regulatory genes were found using our functional clustering test. Functional groups of inflammatory and immune response also appeared (e.g., chemokine signaling pathway and complement activation). In particular, the inflammatory responses were found from the annotation most significantly enriched in *DE* genes ($p = 9.4 \times 10^{-9}$; Supporting Information Figure S3). These genes also had the enrichment in the functional group of metallopeptidase and peptidase S1, which regulates extracellular matrix molecules potentially associated with tissue growth and adhesion (Borden & Heller, 1997). Moreover, the enriched functional groups were consistently found even when we changed the fold change cutoff for determining *DE* genes (Supporting Information Table S3b-c). Our results suggest a relation between *DE* gene function and the main symptoms of HGF, overgrowth, and inflammation of gingival tissues.

### 3.2 | Network connections of differentially expressed genes and exome variants of HGF

Although *DE* genes comprise a molecular signature reflecting HGF symptoms, it remains unclear, in which genetic variations underlie *DE* gene expression changes. We thus analyzed regulatory connections

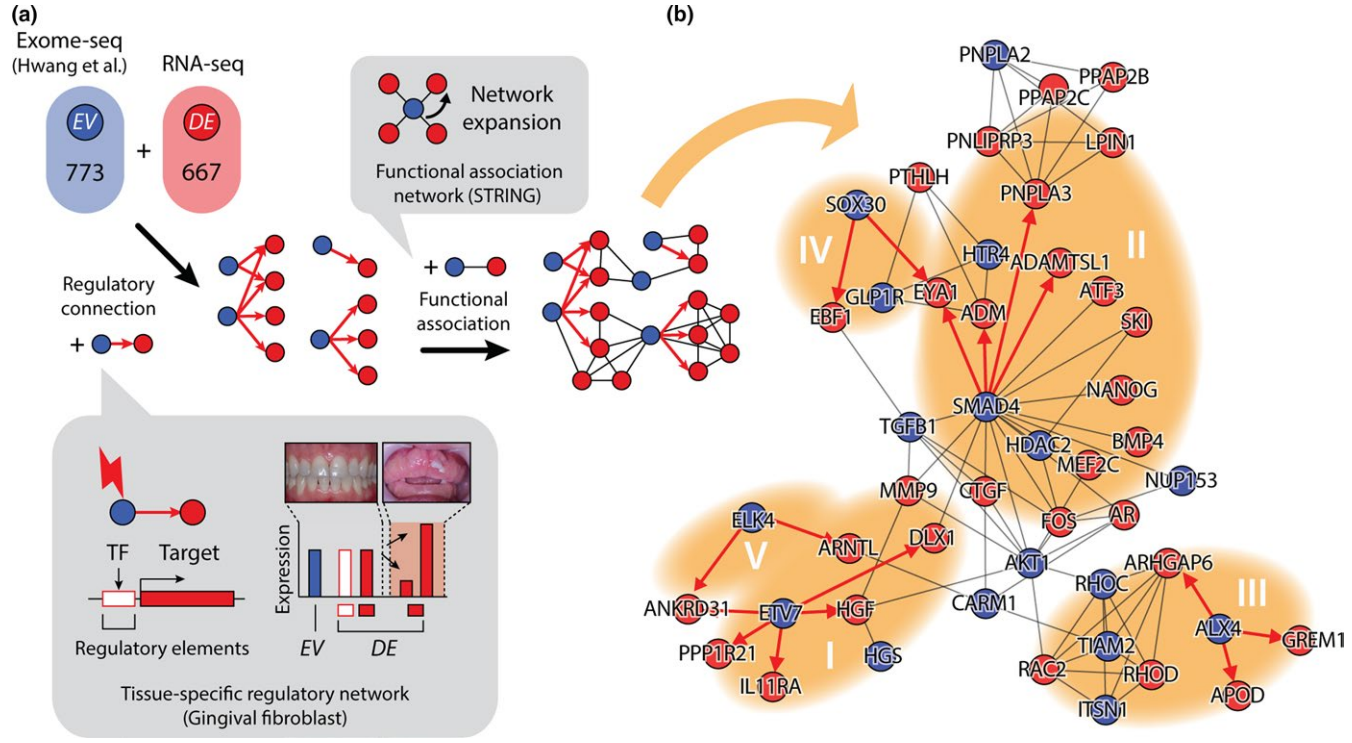


**FIGURE 2** An *EV*-to-*DE* network (a) Step-by-step description of the network connection analysis between *EV* and *DE* genes (see Materials and methods). The concept of *EV*-to-*DE* regulatory connections (red-arrowed lines) and functional associations (black-solid lines) is illustrated in the gray box. (b) The integrated network of *EV*-to-*DE* connections is displayed. Total 16 regulatory connections and 69 functional associations compose the network, including 17 *EV* and 30 *DE* genes. Each gene set (I ~ V) was chosen from the first neighbors of *EV*-to-*DE* regulatory connections derived from each *EV* gene: *ETV7*, *SMAD4*, *ALX4*, *SOX30*, and *ELK4*, respectively

between *DE* genes and exome variant (*EV*) genes derived from our previous exome-seq study of the same patients (Hwang et al., 2017). As a result, *EV*-to-*DE* regulatory connections emerged when an *EV* gene regulates the expression of a *DE* gene in the gingival fibroblast (see Materials and methods; "regulatory connection" section of Figure 2a). Three independent datasets from sequencing analysis support the influence of the *EV*-to-*DE* regulation on HGF conditions: (a) transcription factor (TF) mutations were discovered through the exome-seq of HGF patients, (b) the TF could bind and activate the *cis*-regulatory element of a targeted gene in the cell type relevant to the HGF lesion, as demonstrated by TF-binding motif sequence assays and cap analysis of gene expression (CAGE), which accurately captures transcription initiation regions, (Marbach et al., 2016) and (c) mRNA expression change of the targeted gene was confirmed by RNA-seq of HGF patients. Exomic mutations of the TF changing gene expression or DNA-binding activity would affect the activation of *cis*-regulatory elements on target genes, eventually, and changes gene expression of the target genes. Consequently, five *EV* genes, *ETV7*, *SMAD4*, *ALX4*, *SOX30*, and *ELK4*, turned out to have regulatory connections with more than one *DE* gene (Figure 2a, gray boxes in Table 1), suggesting that exomic variants in those genes potentially affect expression changes under HGF conditions.

To interpret the biological functions of the genes derived from *EV*-to-*DE* regulatory connections, we constructed an *EV*-to-*DE* network, consisting of regulatory connections and the expansion of genes mediated by functional associations between *EV* and *DE* genes (see Materials and methods; "functional association" section of Figure 2a). Five gene set clusters (I ~ V) were organized by the regulatory connections of *EV* genes and their first neighbors in the network (Figure 2b and Table 1). Specifically, throughout clusters I, II, and IV, cell proliferation and regulatory functions, for example, "Positive regulation of EGF receptor signaling," "TGF-β receptor signaling pathway," "Negative regulation of cytochrome C release" of anti-apoptotic process, and "Positive regulation of epithelial to mesenchymal transition," were enriched. The enriched functions were associated with tumorigeneses that may underlie the pathogenesis of HGF.

In particular, *TGFB1*, *SMAD4*, and *MMP9* co-occurred in more than two of the clusters and were important for interconnections within the *EV*-to-*DE* network (Figure 2b and Table 1). The selected gene sets in the network are known to be involved in fibromatosis in diverse tissue types; for example, TGF-β signaling by means of SMAD proteins regulates the matrix metalloproteinases for fibroblast proliferation (Verrecchia & Mauviel, 2002; Yuan & Varga, 2001). Moreover, "Odontogenesis of dentin-containing tooth" was an enriched function in cluster II and "craniofacial development" was annotated to *ALX4* in cluster III, implicating cell proliferation specific to teeth or gingival tissues. Note that current function annotations yield no clear insights linking cluster V with HGF. Taken together, *EV*-to-*DE* network connections are closely associated with the pathogenesis of HGF.

**TABLE 1** Transcriptomic and exomic alterations on genes in the *EV*-to-*DE* network

| Set | Gene | Description | Mutation type (*EV*) log2FC & *P* (*DE*) | | EV or DE genes newly added by net expansion | Functional annotations ($p < 0.05$) |
|---|---|---|---|---|---|---|
| I | ETV7 | Transcription factor ETV7 | Start-gained | | | Negative regulation of cytochrome C release<br>Positive regulation of EGF receptor signaling<br>Negative regulation of autophagy<br>Collagen binding<br>Negative regulation of cysteine-type endopeptidase activity involved in apoptotic process<br>Hyaluronan metabolic processes |
| | HGF | Hepatocyte growth factor receptor | 2.221 | 5.8E−03 | HGS, MMP9, AKT1 | |
| | DLX1 | Homeobox protein DLX-1 | 1.829 | 4.0E−02 | SMAD4 | |
| | ANKRD31 | Putative ankyrin repeat domain-containing protein 31 | −1.749 | 4.0E−04 | – | |
| | PPP1R21 | Phosphatase 1 regulatory subunit 21 | −1.102 | 2.2E−02 | – | |
| | IL11RA | Interleukin−11 receptor subunit alpha | 1.185 | 3.2E−02 | – | |
| II | SMAD4 | Mothers against decapentaplegic homolog 4 | Downstream, intron | | TGFB1, MMP9, CTGF, FOS, HDAC2, ATF3, SKI, AR NANOG, BMP4, NUP153 | Positive regulation of collagen biosynthetic processes<br>Regulation of cell proliferation<br>TGF-β receptor signaling pathway<br>Triglyceride lipase activity<br>Positive regulation of epithelial to mesenchymal transition<br>Branching involved in ureteric bud morphogenesis<br>Odontogenesis of dentin-containing tooth |
| | ADAMTSL1 | ADAMTS-like protein 1 | 1.898 | 1.2E−02 | – | |
| | ADM | Adrenomedullin | −1.514 | 5.5E−04 | HTR4, GLP1R, PTHLH | |
| | EYA1 | Eyes absent homolog 1 | 1.558 | 4.1E−03 | – | |
| | PNPLA3 | Patatin-like phospholipase domain-containing protein 3 | −1.583 | 3.2E−02 | PNPLA2, PPAP2C, PPAP2B, LPIN1, PNLIPRP3 | |
| III | ALX4 | Homeobox protein aristaless-like 4 | Synonymous | | | Small GTPase-mediated signaling<br>Regulation of focal adhesion assembly<br>GTP binding<br>Rho protein signaling<br>Actin filament organization |
| | GREM1 | Gremlin-1 | −1.770 | 2.5E−04 | – | |
| | APOD | Apolipoprotein D | 2.747 | 5.0E−05 | – | |
| | ARHGAP6 | Rho GTPase-activating protein 6 | 1.776 | 1.3E−02 | RHOC, TIAM2, RAC2, RHOD, ITSN1 | |
| IV | SOX30 | Transcription factor SOX-30 | Upstream | | | Positive regulation of epithelial cell proliferation |
| | EBF1 | Transcription factor COE1 | 1.641 | 1.2E−03 | TGFB1 | |
| | EYA1 | Eyes absent homolog 1 | 1.558 | 4.1E−03 | – | |
| V | ELK4 | ETS domain-containing protein Elk-4 | Downstream, synonymous | | – | |
| | ARNTL | Aryl hydrocarbon receptor nuclear translocator-like protein 1 | 1.379 | 1.1E−02 | CARM1 | |
| | ANKRD31 | Putative ankyrin repeat domain-containing protein 31 | −1.749 | 4.0E−04 | – | |

*Notes*. Gene descriptions and functional enrichments were analyzed in terms of five gene sets (I ~ V). Ambiguous, broad annotations, or those associated with the molecular function of transcription factor (TF) were excluded from the functional annotations. Detailed charts of whole enriched terms are presented in Supporting Information Table S4.

### 3.3 | Gene–phenotype association between HGF and *EV*-to-*DE* network genes

Genes and disease phenotype mapping resources provided by the text mining of biomedical literature further support the association of *EV*-to-*DE* network with HGF (Jensen, Saric, & Bork, 2006; Piñero et al., 2015). We projected *EV*-to-*DE* network genes onto those known to be associated with symptoms of HGF (see Materials and methods; Figure 3a). Ten of 47 *EV*-to-*DE* network genes were overlapped with HGF symptom-associated genes (Figure 3b and Supporting Information Table S5), and the overlap was significantly greater than that when the genes were randomly chosen ($p = 7.3 \times 10^{-7}$; Figure 3c). Specifically, *MMP9*, *HGF*, and *CTGF* were associated with "gingival fibromatosis (HP:0000169)." *TGFB1*, *EYA1*, *SMAD4*, *AKT1*, *MEF2C*, *SKI*, and *BMP4* were associated with symptoms that had been reported to occur with gingival fibromatosis

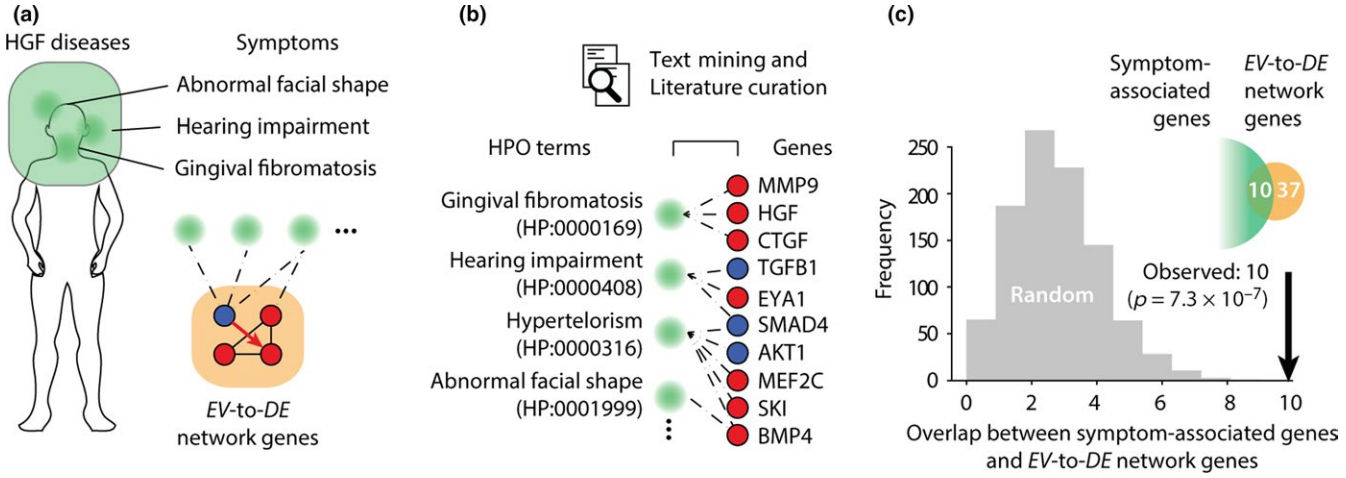


**FIGURE 3** Gene–phenotype mapping between hereditary gingival fibromatosis (HGF) diseases and the *EV*-to-*DE* network genes. (a) Schematic diagram of mapping between genes reportedly associated with HGF symptoms and *EV*-to-*DE* network genes. (b) Human phenotype ontology (HPO) terms of HGF diseases and *EV*-to-*DE* network genes associated with text mining and literature curation. (c) Overlap between HGF symptom-associated and *EV*-to-*DE* network genes. Gray bars show the distribution of overlap counts under random shuffling of symptom-associated genes

such as "hearing impairment (HP:0000408)" or "abnormal facial shape (HP:0001999) (Goldblatt & Singer, 1992; Hartsfield, Bixler, Hazen, Opitz, & Reynolds, 1985)." In particular, three genes from our network analysis, *TGFB1*, *SMAD4*, and *MMP9*, have previously been proposed as related with HGF pathogenesis (Figure 2 and Table 1) (Verrecchia & Mauviel, 2002; Yuan & Varga, 2001).

Diseases sharing underlying genetic variation may share similar phenotypes (Goh, Cusick, Valle, Childs, & Vidal, 2007). We therefore investigated whether *EV*-to-*DE* network genes, including genes putatively causing HGF, were linked to diseases phenotypically similar to HGF (see Materials and methods; Figure 4a). From semantic comparisons of clinical synopses, we found diseases caused by mutations in *EV*-to-*DE* network genes to have high phenotypic similarity scores with HGF (Figure 4b) (van Driel et al., 2006). Specifically, several fibromatosis-related diseases sharing clinical features with HGF were found. For example, "Proteus syndrome (MIM:176920)," dysplasia from the overgrowth of connective tissues, has a phenotypic similarity (*S*) score of 0.21, twofold higher than random. Moreover, craniofacial diseases such as "Myhre syndrome (MIM:139210; $S = 0.38$)" were also found, confirming the functional involvement of craniofacial development in the network gene. We confirmed that the phenotypic similarities among the diseases were unlikely to be obtained by random chance ($p = 0.03$; Figure 4c).

We discovered that the *EV*-to-*DE* network genes also account for diseases that have comorbidity with HGF by measuring the co-occurrence of diseases in hospitalized personnel (see Materials and methods; Supporting Information Figure S4a) (Hidalgo et al., 2009). The average of comorbidity scores between HGF and diseases caused by mutations in *EV*-to-*DE* network genes was significantly higher than that between diseases randomly chosen ($p = 2.9 \times 10^{-4}$; Supporting Information Figure S4b-c). Specifically, "Proteus syndrome (ICD: 759.89)" a disease phenotypically similar to HGF and "deafness (ICD:389.1)," which potentially shares an etiology (fibroblast dysplasia) with HGF, turned out to have comorbidity with HGF diseases (relative risk > 2.0). Taken together, our results suggest a close gene–phenotype association between HGF and *EV*-to-*DE* network genes.

## 4 | DISCUSSION

We conducted a genetic study of idiopathic HGF patients using regulatory network analysis connecting exomic variants and transcriptomic alterations. Our multi-omics approach using the regulatory network provided functions and pathways relevant to the genetic etiology and the putative disease-causing genes of HGF (Figures 1,2 and Table 1). Specifically, inappropriate regulation of the molecules involved in the TGF-β/SMAD signaling pathway may trigger fibroblast overgrowth mimicking HGF. Our results, supported by gene–phenotype mapping of HGF and phenotypically similar diseases (Figures 3,4), suggest that the computational approach of connecting exomic and transcriptomic alteration through regulatory networks is applicable for the clinical interpretation of genetic variants in HGF patients.

We found that regulatory connections between transcriptomic and exomic alterations of HGF were accounted for by TGF-β/SMAD signaling pathways (Figure 2 and Table 1). To confirm whether the TGF-β/SMAD signaling pathways come into play in HGF pathogenesis, we searched for previous studies where the molecular mechanism of pathways in gingival overgrowth had been thoroughly investigated. Ciclosporin A (CsA), a potent immunosuppressive drug with beneficial effects in prevention of transplant rejection and treatment of several immune-related conditions, is known to induce gingival overgrowth as a side effect (Sobral, Aseredo, & Agostini, 2012). This overgrowth arises from exaggerated accumulation of extracellular matrix due to increased expression of transforming

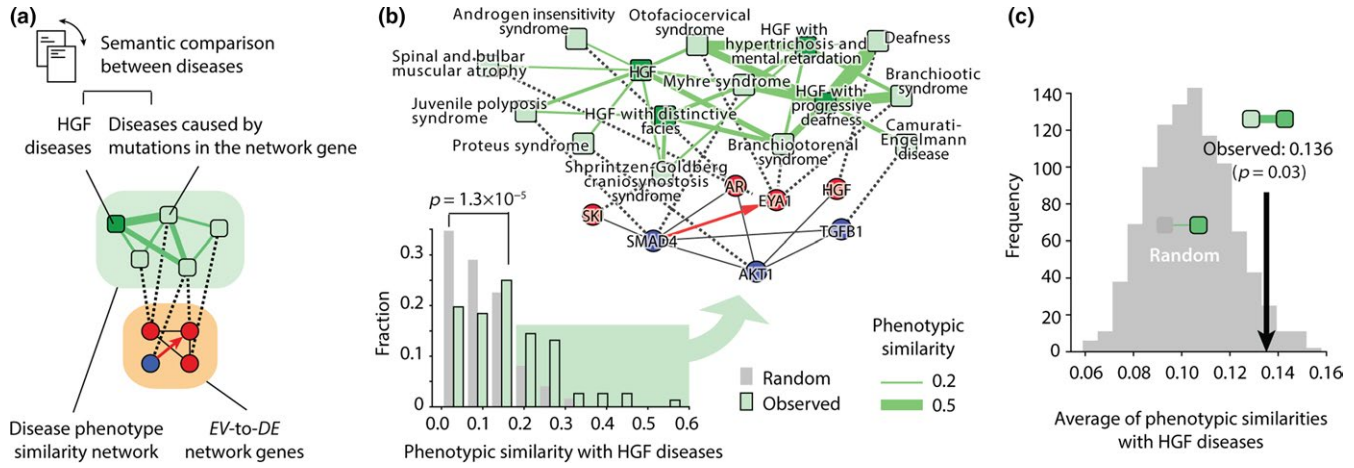


**FIGURE 4** Phenotypic relevance between hereditary gingival fibromatosis (HGF) and diseases caused by mutations in *EV*-to-*DE* network genes. (a) Schematic diagram of phenotypic similarity analysis between HGF and diseases known to be caused by mutations in the *EV*-to-*DE* network gene. Green solid, black-dashed, and solid lines present phenotypic (semantic similarity of disease phenotypes), gene–phenotype (a disease phenotype caused by mutations in a gene) and gene–gene (functional association) connections, respectively. (b) Phenotypic similarity network between four HGF diseases (Supporting Information Table S5) and the 11 *EV*-to-*DE* network gene-associated diseases. Connections of disease pairs with phenotypic similarity >0.2 are shown. (c) The average of phenotypic similarities between HGF and *EV*-to-*DE* network gene-associated diseases. Gray bars show the distribution of the similarity score calculated with randomly chosen diseases

growth factor-β1 (TGF-β1). Increased TGF-β1 expression level eventually perturbs the homeostatic equilibrium between synthesis and degradation of extracellular matrix molecules, inducing type I collagen production. Signaling by TGF-β1 involves the activation of a cytoplasmic downstream pathway composed mainly of SMAD proteins. Specifically, *SMAD4*, a crucial effector in the SMAD pathway through binding into the SMADs2/3 complex (which is phosphorylated by TGF-β1 transmembrane receptors), serves as a mediator between extracellular TGF-β signaling and target genes inside the cell nucleus via protein translocation (Shi & Massagué, 2003). Among the target genes of *SMAD4* revealed by our network analysis, *EYA1* is associated with diseases involving the malfunction of cell proliferation, including in fibroblasts (Figure 4b) (Tadjuidje & Hegde, 2013). Moreover, diverse collagen genes, whose accumulation was proposed as a hallmark of fibromatosis by a drug side effect study, were also regulated by SMAD proteins (Verrecchia, Chu, & Mauviel, 2001). Collagen genes such as *COL11A1* turned out to be differentially expressed in our RNA-seq analysis of HGF patients (Supporting Information Figure S5 and Table S2), although the genes were not directly targeted by *SMAD4* and thus not mapped into the *EV*-to-*DE* network. Taken together, our results and previous studies suggest how TGF-β signaling and the SMAD pathway may work to induce gingival overgrowth. Specifically, the transformation of TGF-β signaling by a SMAD protein might affect the regulatory gene concentration of the extracellular matrix protein which may in turn lead to fibroblast overgrowth.

Our network analysis relates odontogenesis and craniofacial development functions to HGF (Figure 2 and Table 1). Although the two development terms are seemingly irrelevant to HGF, functional matrix theory helps us explain the relationship between cranio-/odontogenesis and connective tissue growth. According to the theory, skeletal development depends on the functional needs of tissues around the bone, accounting for the dependency between the growth of gingival connective tissue and craniofacial bones (Moss, 1968). Indeed, weak facial deformation was found in HGF patients who provided samples for our analysis (data not shown to completely protect patient identities), and the space in which overgrown gingiva had been removed was filled with alveolar bone (Supporting Information Figure S6). Moreover, analysis of phenotypic and molecular connections among HGF, craniofacial diseases, and associated genes support the relevance of genetic etiology among those diseases (Figures 3,4 and Supporting Information Figure S4). The molecular mechanism underlying the relationship may arise from antagonistic crosstalk of the BMP and TGF-β signaling pathways, which are involved in the development of skeletal and soft connective tissues. It is known that *CTGF*, an effector of TGF-β signaling activation through the SMAD pathway, antagonizes BMP4 activity by preventing its binding to BMP receptors but has the opposite effect of enhancing receptor binding on TGF-β1 (Abreu, Ketpura, Reversade, & Robertis, 2002). Our network analysis reveals the molecular connection of genes involved in BMP and TGF-β signaling. *SMAD4* is known to be a co-effector activated by both BMP and TGF-β signaling (Guo & Wang, 2009). *SMAD4* has an *EV*-to-*DE* regulatory connection and a functional association with *ADM* involved in mineralization of dental and craniofacial tissues (Musson, McLachlan, Sloan, Smith, & Cooper, 2010), and at the same time *BMP4* functionally associated with *TGFB1*, *CTGF*, and *MMP9*, which play roles in connective tissue growth (Figure 2). Moreover, this regulatory connection was also likely to be activated in other connective tissues, including gingival fibroblast (Supporting Information Figure S7). Furthermore, such regulatory connections of SMAD4 pathways across multiple biological functions, reflecting involvements in various cellular activities, are

concordant with comorbidities of HGF with various developmental abnormality (Supporting Information Figure S4). Taken together, our results and functional matrix theory implicate a phenotypic connection between HGF and craniofacial diseases which might be due to cross talk between two signaling pathways.

Given the lack of animal model or large-scale cohort data on HGF, we cross-checked our results through gene–phenotype mapping and semantic analysis of disease phenotype similarity (Figures 3,4). Despite the paucity of sequenced cases and controls, our results were in agreement with a range of evidence previously reported for HGF pathogenesis, suggesting that our network approach may be valid and applicable to other diseases. Moreover, such molecular features of HGF, inappropriate regulation of TGF-β signaling, would be helpful to the estimation of incidence rate of HGF. However, to pinpoint features of HGF pathogenesis, further research into interactions between the TGF-β/SMAD signaling pathway and HGF is of paramount significance. Further, research with more cases and controls may produce legitimate biomarkers for early detection of HGF patients.

## ACKNOWLEDGEMENTS

We thank our colleagues for their insightful comments and critical readings of the manuscript. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2055755 and 2018R1A2B6002657).

## CONFLICT OF INTEREST

The authors declare that they have no conflicting interests.

## AUTHOR CONTRIBUTIONS

S.K.H., S.K., J-H.L., and D-H.H. conceived and designed the experiments. J-H.L. and D-H.H. collected patient samples. S.K.H. and J.K. performed the experiments supervised by J-H.L. and S.K.. S.K.H., J.K., J-H.K., and S.K. analyzed the data. S.K.H., J.K., S.K., and J-H.L. wrote the paper.

## ORCID

*Jae-Hoon Lee* https://orcid.org/0000-0003-2281-8885

## REFERENCES

Abreu, J. G., Ketpura, N. I., Reversade, B., & De Robertis, E. M. (2002). Connective-tissue growth factor (ctgf) modulates cell signalling by bmp and TGF-β. *Nature Cell Biology*, *4*, 599–604.

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*, *43*, D789–D798. https://doi.org/10.1093/nar/gku1205

Andersson, R., Gebhard, C., Miguel-Escalada, I., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*, 455–461. https://doi.org/10.1038/nature12787

Ball, E. L. (1941). Case of gingivoma, or elephantiasis of the gingiva. *Journal of Periodontology*, *12*, 96–100.

Bayram, Y., White, J. J., Elcioglu, N., et al. (2017). REST final-exon-truncating mutations cause hereditary gingival fibromatosis. *American Journal of Human Genetics*, *101*, 149–156. https://doi.org/10.1016/j.ajhg.2017.06.006

Borden, P., & Heller, R. A. (1997). Transcriptional control of matrix metalloproteinases and the tissue inhibitors of matrix metalloproteinases. *Critical Reviews in Eukaryotic Gene Expression*, *7*(1-2), 159–178. https://doi.org/10.1615/CritRevEukarGeneExpr.v7.i1-2.90.

Bozzo, L., paes de Almeida, O., Scully, C., & Aldred, M. J. (1994). Hereditary gingival fibromatosis: Report of an extensive four-generation pedigree. *Oral Surgery, Oral Medicine, and Oral Pathology*, *78*, 452–454.

Burga, A., & Lehner, B. (2013). Predicting phenotypic variation from genotypes, phenotypes and a combination of the two. *Current Opinion in Biotechnology*, *24*, 803–809. https://doi.org/10.1016/j.copbio.2013.03.004

Chaurasia, A. (2014). Hereditary gingival fibromatosis. *National Journal of Maxillofacial Surgery*, *5*, 42–46. https://doi.org/10.4103/0975-5950.140171

DeAngelo, S., Murphy, J., Claman, L., Kalmar, J., & Leblebicioglu, B. (2007). Hereditary gingival fibromatosis–a review. *Compendium of Continuing Education in Dentistry*. *28*:138–143; quiz 144, 152.

Forrest, A. R. R., Kawaji, H., Rehli, M., et al. (2014). A promoter-level mammalian expression atlas. *Nature*, *507*, 462–470.

Goh, K., Cusick, M. E., Valle, D., Childs, B., & Vidal, M. (2007). The human disease network. *Proceedings of the National Academy of Sciences*, *104*, 8685–8690.

Goldblatt, J., & Singer, S. L. (1992). Autosomal recessive gingival fibromatosis with distinctive facies. *Clinical Genetics*, *42*, 306–308.

Groza, T., Köhler, S., Moldenhauer, D., et al. (2015). The human phenotype ontology: Semantic unification of common and rare disease. *American Journal of Human Genetics*, *97*, 111–124. https://doi.org/10.1016/j.ajhg.2015.05.020

Guo, X., & Wang, X. F. (2009). Signaling cross-talk between TGF-β/BMP and other pathways. *Cell Research*, *19*, 71–88. https://doi.org/10.1038/cr.2008.302

Häkkinen, L., & Csiszar, A. (2007). Hereditary gingival fibromatosis: Characteristics and novel putative pathogenic mechanisms. *Journal of Dental Research*, *86*, 25–34.

Hart, T. C., Pallos, D., Bozzo, L., et al. (2000). Evidence of genetic heterogeneity for hereditary gingival fibromatosis. *Journal of Dental Research*, *79*, 1758–1764.

Hartsfield, J. K., Bixler, D., Hazen, R. H., Opitz, J. M., & Reynolds, J. F. (1985). Gingival fibromatosis with sensorineural hearing loss: An autosomal dominant trait. *American Journal of Medical Genetics*, *22*, 623–627.

Heath, C., & Tomes, C. S. (1879). Two cases of hypertrophy of the gums and alveoli treated by operation. *Transaction of the Odontological Society of Great Britain*, *11*, 18–26.

Hidalgo, C. A., Blumm, N., Barabási, A.-L., & Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS Computational Biology*, *5*, e1000353.

Hwang, J., Kim, Y. L., Kang, S., et al. (2017). Genetic analysis of hereditary gingival fibromatosis using whole exome sequencing and bioinformatics. *Oral Diseases*, *23*, 102–109. https://doi.org/10.1111/odi.12583

Jensen, L. J., Saric, J., & Bork, P. (2006). Literature mining for the biologist: From information retrieval to biological discovery. *Nature Reviews Genetics*, *7*, 119–129.

Jiao, X., Sherman, B. T., Huang, D. W., et al. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, *28*, 1805–1806. https://doi.org/10.1093/bioinformatics/bts251

Jorgenson, R. J., & Cocker, M. E. (1974). Variation in the inheritance and expression of gingival fibromatosis. *Journal of Periodontology*, *45*, 472–477. https://doi.org/10.1902/jop.1974.45.7.472

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Koch, C., Konieczka, J., Delorey, T., et al. (2017). Inference and evolutionary analysis of genome-scale regulatory networks in large phylogenies. *Cell Systems*, *4*, 543–558.e8. https://doi.org/10.1016/j.cels.2017.04.010

Lage, K., Hansen, N. T., Karlberg, E. O., et al. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 20870–20875. https://doi.org/10.1073/pnas.0810772105

Lizio, M., Harshbarger, J., Shimoji, H., et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biology*, *16*, 1–14. https://doi.org/10.1186/s13059-014-0560-6

Maher, B. (2012). ENCODE: The human encyclopaedia. *Nature*, *489*, 46–48.

Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., & Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature Methods*, *13*, 1–44. https://doi.org/10.1038/nmeth.3799

Moss, M. (1968). A theoretical analysis of the functional matrix. *Acta Biotheoretica*, *18*, 195–202.

Musson, D. S., McLachlan, J. L., Sloan, A. J., Smith, A. J., & Cooper, P. R. (2010). Adrenomedullin is expressed during rodent dental tissue development and promotes cell growth and mineralization. *Biology of the Cell*, *102*, 145–157. https://doi.org/10.1042/BC20090122

Piñero, J., Queralt-Rosinach, N., Bravo, À., et al. (2015). DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes. *Database*, *2015*, 1–17. https://doi.org/10.1093/database/bav028

Rapp, R., Nikiforuk, G., Donohue, D. W., & Williams, C. H. M. (1955). Idiopathic hyperplasia of gingivae associated with macrocheilia and ankyloglossia: A case report. *Journal of Periodontology*, *26*, 51–55.

Sah, K., Chandra, S., & Kaur, N. (2015). Hereditary gingival fibromatosis: A heterogeneous genetic disorder. *Saudi Journal for Health Sciences*, *4*, 208.

Shi, Y., & Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*, *113*, 685–700.

Sobral, L. M., Aseredo, F., Agostini, M., et al. (2012). Molecular events associated with ciclosporinA-induced gingival overgrowth are attenuated by Smad7 overexpression in fibroblasts. *Journal of Periodontal Research*, *47*, 149–158. https://doi.org/10.1111/j.1600-0765.2011.01412.x

Szklarczyk, D., Franceschini, A., Wyder, S., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*, D447–D452. https://doi.org/10.1093/nar/gku1003

Tadjuidje, E., & Hegde, R. S. (2013). The eyes absent proteins in development and disease. *Cellular and Molecular Life Sciences*, *70*, 1897–1913. https://doi.org/10.1007/s00018-012-1144-9

Thompson, D., Regev, A., & Roy, S. (2015). comparative analysis of gene regulatory networks: From network reconstruction to evolution. *Annual Review of Cell and Developmental Biology*, *31*:399–428. https://doi.org/10.1146/annurev-cellbio-100913-012908

Trapnell, C., Roberts, A., Goff, L., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*, 562–578.

van Driel, M. A., Bruggeman, J., Vriend, G., Brunner, H. G., & Leunissen, J. A. M. (2006). A text-mining analysis of the human phenome. *European Journal of Human Genetics*, *14*, 535–542.

Verrecchia, F., Chu, M. L., & Mauviel, A. (2001). Identification of Novel TGF-β/smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. *Journal of Biological Chemistry*, *276*, 17058–17062.

Verrecchia, F., & Mauviel, A. (2002). Transforming growth factor-β signaling through the smad pathway: Role in extracellular matrix gene expression and regulation. *The Journal of Investigative Dermatology*, *118*, 211–215.

Yuan, W., & Varga, J. (2001). Transforming growth factor-β repression of matrix metalloproteinase-1 in dermal fibroblasts involves Smad3. *Journal of Biological Chemistry*, *276*, 38502–38510.

Zackin, S. J., & Weisberger, D. (1961). Hereditary gingival fibromatosis: Report of a family. *Oral Surgery, Oral Medicine, and Oral Pathology*, *14*, 828–836.

Zhou, X., Menche, J., Barabási, A.-L., & Sharma, A. (2014). Human symptoms-disease Network. *Nature Communications*, *5*, 4212. https://doi.org/10.1038/ncomms5212

## SUPPORTING INFORMATION

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

**How to cite this article:** Han SK, Kong J, Kim S, Lee J-H, Han D-H. Exomic and transcriptomic alterations of hereditary gingival fibromatosis. *Oral Dis*. 2019;00:1–10. https://doi.org/10.1111/odi.13093