

## Research Article

# Identification of Novel Choroidal Neovascularization-Related Genes Using Laplacian Heat Diffusion Algorithm

Minjie Sheng <sup>1</sup>, Haiying Cai,<sup>1</sup> Ming Cheng,<sup>1</sup> Jing Li <sup>1</sup>, Jian Zhang <sup>2,3,4,5,6</sup>  
and Lihua Liu <sup>1</sup>

<sup>1</sup>Department of Ophthalmology, Yangpu Hospital, Tongji University School of Medicine, 450 Tengyue Road, Shanghai 200090, China

<sup>2</sup>Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

<sup>3</sup>Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai 200080, China

<sup>4</sup>Shanghai Engineering Center for Visual Science and Photomedicine, Shanghai 200080, China

<sup>5</sup>National Clinical Research Center for Eye Diseases, Shanghai 20080, China

<sup>6</sup>Shanghai Engineering Center for Precise Diagnosis and Treatment of Eye Diseases, Shanghai 20080, China

Correspondence should be addressed to Jian Zhang; natalieeilatan@126.com and Lihua Liu; lh\_liu@yeah.net

Received 9 May 2021; Accepted 20 August 2021; Published 7 September 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Minjie Sheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Choroidal neovascularization (CNV) is a type of eye disease that can cause vision loss. In recent years, many studies have attempted to investigate the major pathological processes and molecular pathogenic mechanisms of CNV. Because many diseases are related to genes, the genes associated with CNV need to be identified. In this study, we proposed a network-based approach for identifying novel CNV-associated genes. To execute such method, we first employed a protein-protein interaction network reported in STRING. Then, we applied a network diffusion algorithm, Laplacian heat diffusion, on this network by selecting validated CNV-related genes as the seed nodes. As a result, some novel genes that had unknown but strong relationships with validated genes were identified. Furthermore, we used a screening procedure to extract the most essential genes. Eleven latent CNV-related genes were finally obtained. Extensive analyses were performed to confirm that these genes are novel CNV-related genes.

## 1. Introduction

Choroidal neovascularization (CNV) is a typical pathogenic process that refers to the abnormal creation of blood vessels specifically in the choroid layer of the eye. As a severe pathogenesis of one subtype of age-related macular degeneration (AMD), CNV can be clinically concomitant with various ocular symptoms such as extreme myopia and malignant myopic degeneration. According to the recent epidemiological statistics provided by *Lancet*, more than 6 million people around the world suffered from AMD in 2015 [1]. Based on another independent survey, the prevalence of CNV-associated AMD was found to be 1.2% of all adults aged 43–86 years [2], indicating that CNV may be one of the major causes of vision loss.

As mentioned above, CNV is a major threat to visual health, especially in elderly people around the world. Therefore, for centuries, scientists have attempted to determine the major pathological processes and molecular pathogenic mechanisms of CNV [3, 4]. However, the detailed and comprehensive mechanisms of CNV have not been fully elucidated. According to existing literatures, the major pathogenic mechanisms of CNV can be attributed to the imbalance of antiangiogenic factors and angiogenic factors [5, 6]. The imbalance of these factors in the choroid may promote vasculogenesis and angiogenesis pathologically related to CNV [6]. In terms of regulators, PEDF (pigment epithelium-derived factor) [6] and VEGF (vascular endothelial growth factor), which are antiangiogenic and typical angiogenic factors, respectively [5] have both been confirmed

to contribute to the initiation and progression of CNV. However, the factors or initiators that drive the abnormal biological functions of PEDF and VEGF have not been confirmed. Hypoxia [7], high glucose [8], protein kinase C activation [9], advanced glycation end products [10], reactive oxygen species [11], activated oncogenes [12], and abnormal cytokine production [13] may all contribute to the pathogenesis and clinical symptoms of CNV.

Although the pathogenesis of the diseases we have discussed is complicated, we can still simply cluster all the potential pathogenic factors into two groups: genetic factors and environmental factors [3, 4]. In this study, we computationally investigated the genetic pathogenesis of CNV. According to recent publications, various genetic factors have been confirmed to contribute to CNV. Abnormal angiogenesis and antiangiogenesis are two major pathogenic processes in such disease [5, 6]. Recent publications revealed that various genes related to angiogenesis and antiangiogenesis may directly participate in the pathogenesis of CNV. *VEGF* [14] and *FGF2* [15] are two typical genes associated with angiogenesis. In 2009, these genes have been confirmed to be related to CNV and regulate its rate of progression [16]. Besides these genes, another functional gene called *CFI*, which is related to extreme myopia, has also been reported to contribute to CNV pathogenesis [17, 18], revealing the complicated genetic basis of CNV. Other functional genes associated with cell proliferation, such as *RELA* [19], *NFKB1* [20, 21], and *RELB* [19], have all been reported to promote abnormal angiogenesis during the initiation and progression of CNV.

For decades, scientists have attempted to reveal the comprehensive genetic background of CNV. However, identifying and validating CNV-associated genes one by one is quite expensive and time consuming. In recent years, with the development of high-throughput sequencing, bioinformatics algorithms have provided us a novel and more effective approach for identifying CNV-associated genes. In 2016, a systematic prediction [22] based on all the identified CNV-related genes, protein-protein interaction (PPI) network, and shortest path algorithm identified various genes associated with CNV, including *ANK1*, *ITGA4*, and *CD44*. Most of these genes have already been identified to contribute to abnormal angiogenesis or antiangiogenesis in the choroid [22], validating the efficacy and accuracy of computational prediction on disease-associated genes. Therefore, in this study, we introduced a novel computational method called *Laplacian heat diffusion* (LHD) [23] to further explore the pathogenic factors of CNV. This study not only identified potential CNV-associated genes but also revealed the detailed pathogenesis of CNV.

## 2. Materials and Methods

**2.1. CNV-Associated Genes.** Genes associated with AMD were first obtained from a previous study [24]. In detail, we downloaded the “Additional file 3” in such study, which contained these genes. Then, according to “Additional file 5”, genes in CNV up or CNV down modules were picked up, accessing 37 CNV-associated genes (Table S1). These

genes were further converted to Ensembl gene IDs to be consistent with the protein IDs in the PPI network from the STRING database [25]. These genes comprised a seed gene set  $S$ .

**2.2. PPIs.** In general, proteins interact with each other to regulate biological process; thus, they share similar biological functions. Based on this assumption, many studies have been devoted to infer protein functions. Therefore, potential CNV-associated genes can be identified from the known CNV-associated genes and their interaction network.

We downloaded 4,274,001 human protein-protein interactions (PPIs) for 19,247 proteins from STRING (<https://www.string-db.org/>, version 10) [26]. These interactions were derived from genomic context predictions, high-throughput lab experiments, (conserved) coexpression, automated text mining, and previous knowledge in databases. Thus, PPIs reported in STRING can widely measure the associations of proteins compared with those in some other databases [27, 28], in which PPIs were only determined by solid experiments. For each PPI, both proteins are represented by Ensembl IDs, and a score ranging from 150 to 999 is assigned. A high score indicates that the corresponding interaction is supported by high-quality evidence. The interaction score between two proteins ( $P_1$  and  $P_2$ ) was denoted as  $I(P_1, P_2)$ . Using the abovementioned data, we can construct a PPI network consisting of 19,247 nodes and 4,274,001 edges, which connects two nodes with interaction score as the weight if and only if two proteins interact. The PPI network is denoted as  $G$ . Such PPI network has been widely used in many researches [29–37].

**2.3. Laplacian Heat Diffusion.** Nowadays, network methods are more and more popular to deal with different biological and medical problems [30, 32, 36, 38–41]. This study also adopted a powerful network method, LHD algorithm. As a type of network diffusion method, heat diffusion follows some rules to transmit heat on the seed nodes to surrounding nodes in the network. The heat on a node indicates its connections to seed nodes. In this study, the LHD algorithm [23] was applied to search for novel CNV-related genes, which was a heat diffusion process on a Laplacian matrix constructed from protein-protein network.

Given a PPI network  $G$ , we can first construct its adjacent matrix  $A$  based on the edge weights. Then, we normalize it column wisely as follows:

$$A'[i, j] = \frac{A[i, j]}{\sum_{k=1}^n A[k, j]}, \quad (1)$$

where  $i$  is the column index of 19,247 nodes in  $G$  and  $j$  is the index of CNV-related genes. Each column in  $A'$  was a 19,247-dimensional vector. Each element was the heat of a node in the network  $G$ . Initially, the component in  $A'$  corresponding to 37 CNV-related genes was configured to be  $1/37$ ; other components were set to 0. Then, the values of each vector were updated as follows:

$$H_t[i] = H_0[i] \exp(-\lambda_i t), \quad (2)$$

where  $H_t$  is the heat distribution at time  $t$  and  $\lambda_i$  is the  $i$ th eigenvalue of matrix  $A'$ . We updated the vectors until the heat distribution vectors at two consecutive time points change as small as a defined threshold. After the diffusion process, each node was assigned a heat value. A larger heat value indicates that the node is more important. Thus, we selected nodes with heat values greater than the defined cut-off and mapped those nodes back to the corresponding genes.

In this study, we used the LHD algorithm (<https://CRAN.R-project.org/package=diffusr>) to perform the analysis with default parameters on the PPI network  $G$ .

**2.4. Postprocessing of CNV-Related Candidate Genes.** According to the LHD-based method, we can obtain a large number of candidate CNV-related genes. However, some of them are essential genes, while others are nonessential genes. A three-stage method was applied to select the essential genes by integrating other biological information: (1)  $Z$ -score based on permutation test to exclude false positives, (2) maximum interaction score (MIS) based on PPI information to exclude genes with few connections to the validated CNV-related genes, and (3) maximum function score (MFS) based on biological function annotation information to filter functional genes.

**2.4.1.  $Z$ -Score.** To evaluate the significance of the produced heat values, we randomly sampled 1000 gene sets and calculated the mean and standard deviation of these heat values. Then, we calculated the  $Z$ -score for all CNV-related candidate genes. In detail, 1000 gene sets with a size of 37 were randomly generated. For each gene set, we performed the LHD algorithm on the PPI network  $G$  by using it as the seed set. Then, each gene  $g$  was assigned a heat value. The above process was run for the produced 1000 gene sets. Each  $g$  received 1000 heat values and a real heat value based on 37 validated CNV-related genes. We calculated the measurement  $Z$ -score as follows:

$$Z\text{-score}(g) = \frac{h - \bar{h}}{\text{sd}}, \quad (3)$$

where  $h$  is the real heat value of gene  $g$  and  $\bar{h}$  and  $\text{sd}$  are the mean and standard deviation of 1000 heat values of the 1000 randomly produced gene sets, respectively. The higher the  $Z$ -score of one gene is, the more likely it is a real CNV-related gene. In this study, we selected genes with  $Z$ -score greater than 1.96.

**2.4.2. MIS.** After the permutation test, some CNV-related candidate genes were further verified to have strong associations with the validated CNV-related genes. In general, interacting proteins always exhibit similar functions. Based on this observation, we calculated MIS as follows:

$$\text{MIS}(g) = \max \left\{ I(g, g') \mid g' \text{ is a validated CNV-related gene} \right\}, \quad (4)$$

where  $I(g, g')$  is the interaction score between two genes from the STRING database. A high MIS value indicates that this gene is strongly connected to the validated CNV-related genes; thus, it is more likely to be true CNV-associated gene. Here, we set a threshold of 900 (the highest confidence score in the STRING database) to filter out genes with low MIS values.

**2.4.3. MFS.** To be CNV-related genes, they must highly contribute to certain biological processes involved in CNV. To further select more reliable CNV-related candidate genes, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) [42] pathways were used. We extracted important candidate CNV-related genes with similar GO terms and KEGG pathways to validate CNV-related genes. The enrichment theory [43, 44] was applied to estimate the relationships between genes and GO/KEGG pathways. It encodes a gene as a vector. The relationship between two genes can be calculated as follows:

$$Q(g, g') = \frac{E(g) \cdot E(g')}{\|E(g)\| \cdot \|E(g')\|}, \quad (5)$$

where  $E(g)$  is the column vector obtained according to enrichment theory.

Similarly, for each gene,  $\text{MFS}(g)$  was calculated as follows:

$$\text{MFS}(g) = \max \left\{ Q(g, g') \mid g' \text{ is a validated CNV-related gene} \right\}. \quad (6)$$

The higher the MFS of one gene is, the more GO/KEGG pathways it shares with the validated CNV-related genes. The final candidate CNV-related genes were extracted with an MFS value greater than a defined cutoff of 0.9.

### 3. Results

In this study, we presented a computational approach to infer novel CNV-associated genes using the LHD-based method. The entire procedures are illustrated in Figure 1. This approach collected verified CNV-related genes, which were extrapolated to identify novel candidate genes on the PPI network using Laplacian heat diffusion. Next, these identified candidate genes were further screened to filter out false positive genes that are not associated with any CNV-related biological process.

We first selected genes with a heat value  $> e - 10$ , and a total of 19,218 genes were obtained. Then, these genes were evaluated using the permutation test with 1000 randomly generated sets. We selected genes with a  $Z$ -score greater than 1.96 and obtained a list of 153 genes. We further filtered out genes with fewer connections to the validated CNV-related genes by MIS score. We kept genes with MIS value greater than 900 and obtained 27 genes. Finally, for each of the 27 genes, we calculated the MFS and selected genes with an MFS value greater than 0.9, resulting in a final list of 11

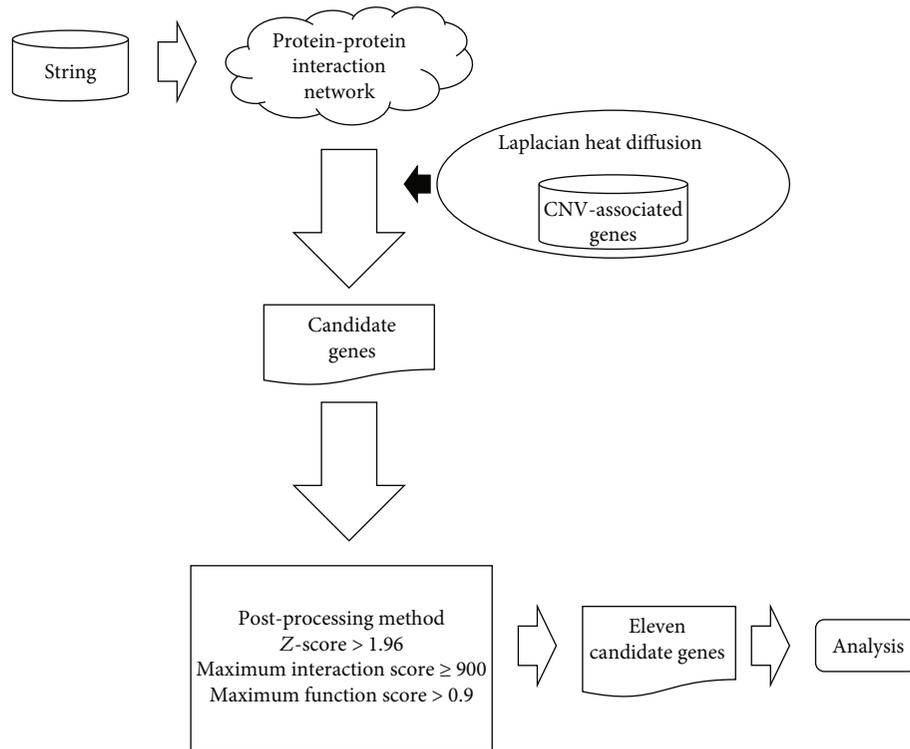


FIGURE 1: Entire procedures to identify novel choroidal neovascularization- (CNV-) related genes. A protein-protein interaction network reported in STRING is employed. Laplacian heat diffusion (LHD) with validated CNV-related genes as seed nodes is applied to such network for extracting raw candidate genes. They are further filtered by a postprocessing method, resulting in eleven candidate genes. These genes were extensively analyzed.

CNV-related genes, which is totally different from previous discoveries [22]. The selected numbers of putative genes in different steps of LHD are shown in Table 1, and the detailed information of the 11 final candidate CNV-associated genes is listed in Table 2. The interaction network between the 11 candidate genes and 37 verified CNV-related genes is shown in Figure 2. All measurements mentioned above are listed in sheets 1–4 of Table S2.

#### 4. Discussion

As we have analyzed above, we applied a novel computational method named *Laplacian heat diffusion* [45] to identify potential CNV-related genes based on the existing PPI network provided in STRING [25]. According to such algorithm and the database, we screened out eleven functional genes that may directly or indirectly participate in the pathogenesis of CNV. To validate the efficacy and accuracy of our newly applied computational method, we performed a systematic datamining on the biological functions and CNV relevance of all predicted genes. The predicted genes have been validated by recent publications. The detailed analysis on each gene can be seen below. For a clear description, we classified these genes into some classes, which is illustrated in Figure 3.

**4.1. Matrix Metalloproteinases (MMPs).** *MMP3* (ENSP0000 0299855), which ranks the highest in the prediction list, has been predicted to be related to the pathogenesis of

TABLE 1: Number of candidate CNV-related genes in different stages of LHD-based method.

| Method           | Network diffusion algorithm | Z-score | MIS | MFS |
|------------------|-----------------------------|---------|-----|-----|
| LHD-based method | 19,218                      | 153     | 27  | 11  |

CNV. Generally, it has been widely reported to contribute to the activation of procollagenase [46] and matrix remodeling [47]. In terms of its potential pathogenic functions in CNV, this gene has been confirmed to act abnormally in the choroidal neovascular membranes, implying its pathogenic potential [48]. Further studies on the contribution of hypoxia to CNV confirmed that our predicted gene *MMP3* may contribute to hypoxia-induced apoptosis and secretion of proangiogenic factors in the choroid under hypoxia microenvironment, which further initiates CNV [49]. Therefore, our predicted gene *MMP3* may functionally be a potential driving factor for CNV, demonstrating the accuracy of our prediction result. Apart from *MMP3*, three other components of the MMP family, namely, *MMP13* (ENSP00000260302), *MMP7* (ENSP00000260227), and *MMP10* (ENSP00000279441), have also been predicted to contribute to the pathogenesis of CNV in our prediction list with a high rank. With similar biological functions as *MMP3*, all of such three genes (*MMP13*, *MMP7*, and *MMP10*) have been reported to participate in the abnormal

TABLE 2: Eleven candidate genes yielded by LHD-based method.

| Ensemble ID     | Gene symbol | Description                            | Heat         | Z-score | MIS | MES    |
|-----------------|-------------|--|--------------|---------|-----|--------|
| ENSP00000299855 | MMP3        | Matrix metalloproteinase 3             | $8.79E - 05$ | 2.0806  | 999 | 0.9761 |
| ENSP00000230990 | HBEGF       | Heparin-binding EGF-like growth factor | $9.01E - 05$ | 3.0964  | 989 | 0.9608 |
| ENSP00000260302 | MMP13       | Matrix metalloproteinase 13            | $1.54E - 04$ | 3.9922  | 964 | 0.9600 |
| ENSP00000260227 | MMP7        | Matrix metalloproteinase 7             | $1.25E - 04$ | 3.8042  | 975 | 0.9580 |
| ENSP00000262768 | TIMP2       | TIMP metalloproteinase inhibitor 2     | $1.20E - 04$ | 3.2461  | 994 | 0.9569 |
| ENSP00000222390 | HGF         | Hepatocyte growth factor               | $1.09E - 04$ | 4.0224  | 922 | 0.9487 |
| ENSP00000304408 | COL3A1      | Collagen type III alpha 1 chain        | $1.40E - 04$ | 2.6753  | 951 | 0.9461 |
| ENSP00000279441 | MMP10       | Matrix metalloproteinase 10            | $1.31E - 04$ | 3.2940  | 977 | 0.9186 |
| ENSP00000347665 | COL18A1     | Collagen type XVIII alpha 1 chain      | $1.25E - 04$ | 2.5902  | 991 | 0.9183 |
| ENSP00000277480 | LCN2        | Lipocalin 2                            | $9.96E - 05$ | 2.1251  | 985 | 0.9141 |
| ENSP00000297904 | VEGFD       | Vascular endothelial growth factor D   | $2.13E - 04$ | 6.6122  | 939 | 0.9078 |

angiogenesis of choroidal tissues, validating their specific contribution to CNV. In 2011, a study [50] on CNV in a mouse model confirmed that the deficiency of *MMP13* contributed to the impairment of neovascularization formation in choroid tissues, and such pathogenesis could be restored by injecting mesenchymal cells secreting *MMP13*, validating the specific role of this gene during CNV initiation and progression. As for *MMP7*, basal laminar and linear deposits are typical complications of CNV, contributing to the constitution of the CNV microenvironment [51, 52]. A recent study [53] on the typical basal laminar and linear deposits of CNV confirmed that *MMP7* together with its homologue *MMP13* may contribute to CNV by regulating the inflammatory processes in the microenvironment of choroidal tissues. Furthermore, *MMP10* has also been validated by recent publications. Although no reports connected *MMP10* and CNV directly, the specific contribution of all metalloproteinases including *MMP10* on choroidal microenvironment remodeling and inflammation mediation implies the specific biological function of *MMP10* during the progression of CNV [54].

**4.2. Growth Factors.** *HBEGF* (ENSP00000230990) has also been predicted to contribute to the progression of CNV. As a typical growth factor, *HBEGF* participates in the ERBB2 signaling pathway and interacts with functional genes such as EGFR and ERBB4 [55, 56]. A recent study confirmed that *HBEGF* may affect the production and biological functions of *VEGF* in CNV [57]. Therefore, although no direct reports confirmed the detailed biological function of *HBEGF* in CNV, this gene may interact with *VEGF* and play a crucial pathogenic role during the progression of CNV. Another functional growth factor encoding gene *HGF* (ENSP00000222390) has also been predicted to contribute to the pathogenesis of CNV. Generally, the binding of *HGF* to its target receptor (hepatocyte growth factor) contributes to the regulation of cell growth, cell motility, and morphogenesis in various cell and tissue subtypes [58, 59]. As for its unique pathogenic contribution to CNV, a paired experimental study [60] on CNV confirmed that compared with normal tissues, the pathogenic tissues of the choroid during CNV

initiation and progression have different expression profiling of growth factors including *VEGF*, *HGF*, and *FGF*, implying the potential pathogenic role of *HGF* in such disease. In 2011, a specific study on the biological and pathogenic functions of cytokines in CNV confirmed that *HGF* has a mitogenic effect on choroidal cells, promoting neovascularization processes [61]. Therefore, such gene may be a potential CNV-associated gene. As the next predicted growth factor in the predicted list of genes, *VEGFD* (ENSP00000297904) has been widely reported to be a member of the platelet-derived growth factor family. This gene has been reported to promote angiogenesis [62], lymph angiogenesis [62], and endothelial cell growth [63]. As the homologue of the identified key driver gene of CNV (*VEGF*) generated by differential alternative splicing, *VEGFD* directly participates in the pathogenesis of CNV, regulating the same biological processes of *VEGF* [64]. Recent clinical studies [65] confirmed that *VEGFD* may also be a candidate marker for the diagnosis and treatment of CNV, and drugs that target *VEGF* to relieve symptoms may also target the products of *VEGFD*.

**4.3. MMP Inhibitors.** *TIMP2* (ENSP00000262768) has been widely reported to act as a natural inhibitor for MMPs [66]. With a specific expression pattern in vitreous and sub-retinal fluid, this gene has been found to be expressed in choroid tissues [67] and directly contribute to the activation of the hypoxia-induced VEGF signaling pathway and MMP regulation [68]. Considering the irreplaceable role of *VEGF* in CNV, *TIMP2* may be a potential CNV-associated gene.

**4.4. Collagens.** Based on our newly presented computational methods, we also obtained two collagen coding genes that may contribute to the pathogenesis of CNV, namely, *COL3A1* (ENSP00000304408) and *COL18A1* (ENSP00000347665). *COL3A1* encodes the pro-alpha1 chain of type III collagen, a fibrillary collagen. Based on existing literatures, this gene contributes to the regulation of cortical development together with type I collagen in soft connective tissues [69, 70]. As for its specific pathogenic contribution to CNV, a specific study [71] confirmed that *COL3A1* may

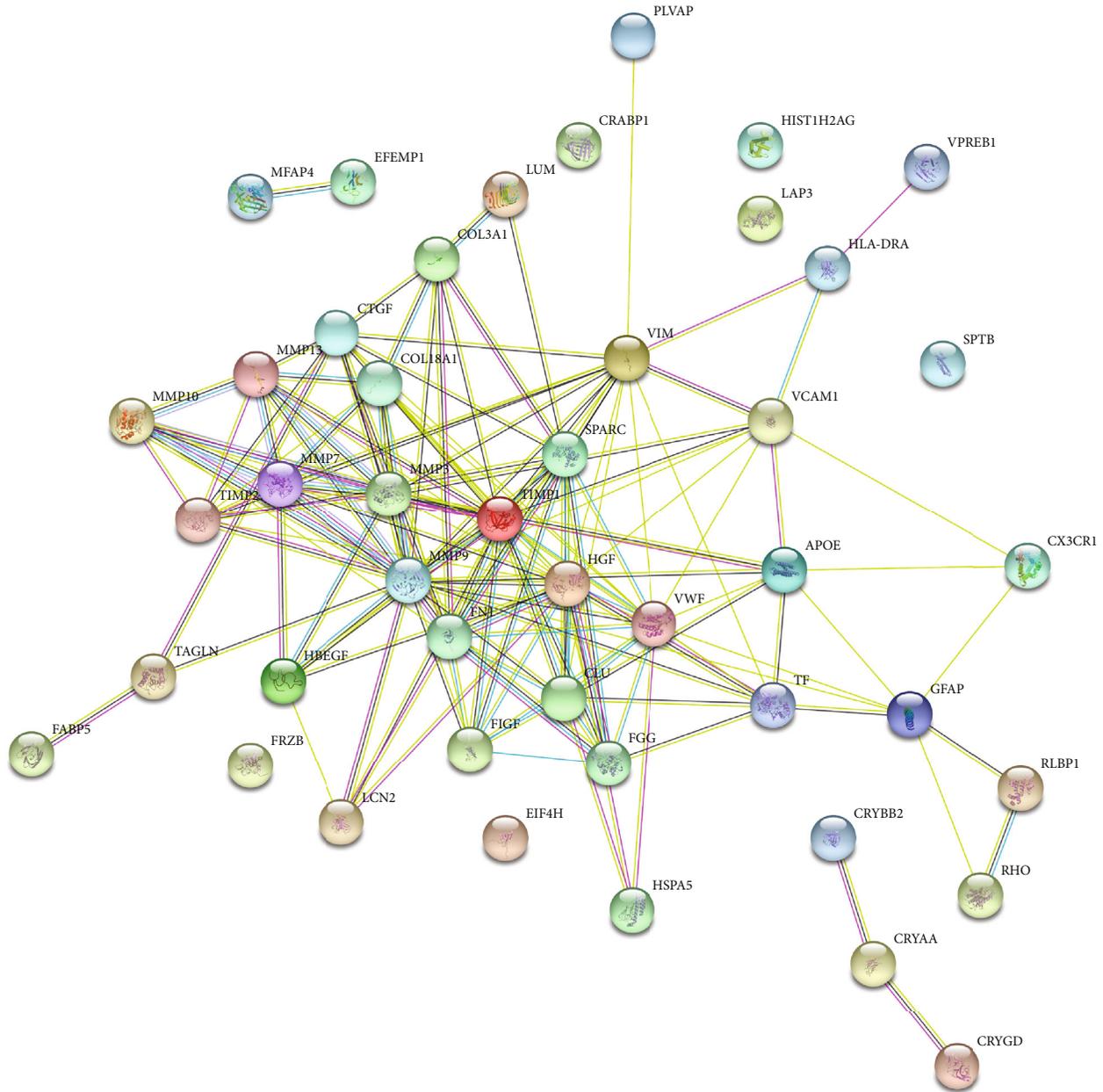


FIGURE 2: Interaction network between the 11 candidate genes and verified 37 CNV-related genes.

contribute to actin cytoskeleton remodeling and affect the specific lesion size and fibrosis of CNV. Similarly, the next predicted gene (*COL18A1*) has also been reported to participate in collagen-associated CNV pathogenesis [72]. Currently, no direct pathogenic experiment has confirmed that *COL18A1* can induce the progression of CNV independently. Other studies on collagen families including collagen XVIII [48, 73, 74] in CNV and their respective angiogenic functions have validated the potential pathogenic role of our predicted collagen encoding genes.

**4.5. Lipocalins.** Apart from MMPs, collagen, and growth factor-associated genes, we also obtained a specific lipocalin encoding gene, namely, *LCN2* (ENSP00000277480). Generally, this gene has been identified in the lung, breast [75],

and eye secretions [76] and contribute to the transport of hydrophobic ligands [77]. As for its specific contribution to CNV, this gene may promote angiogenesis and neovascularization under pathogenic conditions [78, 79]. With a high-expression pattern in choroid tissues [80] and its interaction with MMPs [81], *LCN2* has been confirmed to participate in the pathogenic activation of the *AKT2-NF- $\kappa$ B-lipocalin-2* axis in CNV [82].

Taken together, the predicted functional genes are enriched in MMP-, growth factor-, collagen-, and lipocalin-related genes, implying the specific role of such components during the initiation and progression of CNV. The predicted genes have all been confirmed by recent publications as we have described above. Therefore, the computational approach in this study may be quite effective and accurate for

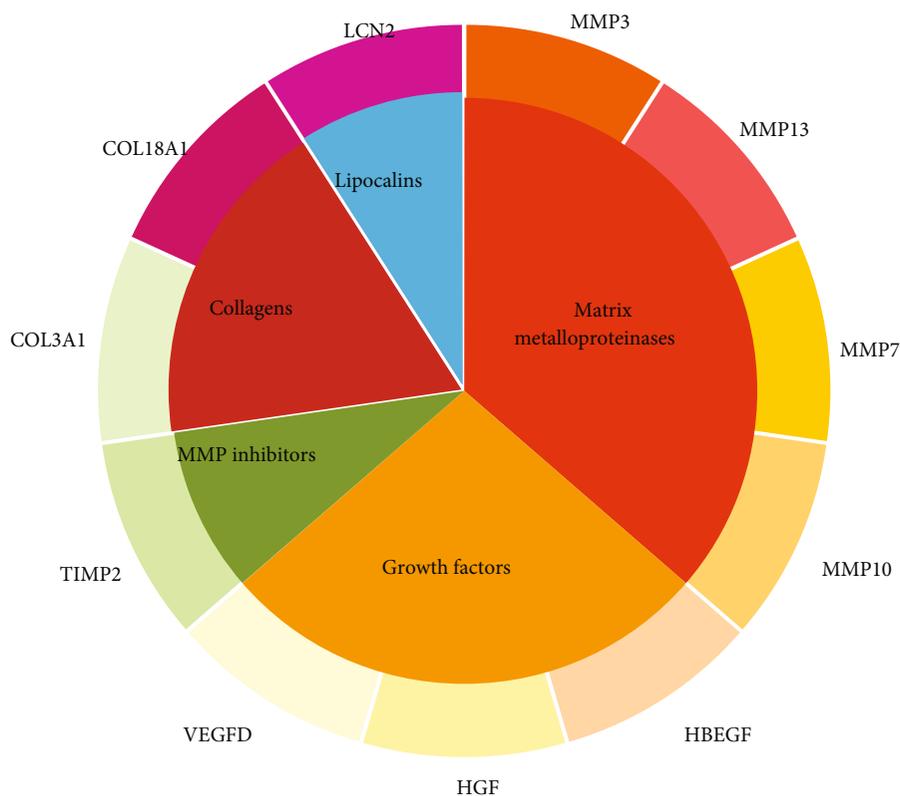


FIGURE 3: Classification of eleven candidate CNV-associated genes.

identifying CNV-associated genes. This study not only identified a group of functional CNV-associated genes and potential related biological processes but also contributed to the improvement of current computational prediction approaches on the genetic background of diseases.

## 5. Conclusions

This study employed a powerful network diffusion method to identify possible CNV-related genes in a PPI network. To obtain reliable genes, a three-stage method followed to screen out key latent CNV-related genes. The analysis on final obtained genes indicate that they can be novel CNV-related genes with high likelihood. It is hopeful that the new findings reported in this study can provide new insights for investigating CNV.

## Data Availability

The data used to support the findings of this study are included within the supplementary information files.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Authors' Contributions

Minjie Sheng and Haiying Cai contributed equally to this work.

## Supplementary Materials

Table S1: curated CNV-associated genes. Table S2: measurements of genes in each step. (*Supplementary Materials*)

## References

- [1] G. B. D. Disease, I. Injury, and C. Prevalence, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1545-1602, 2016.
- [2] R. Klein, B. E. Klein, and K. L. Linton, "Prevalence of Age-related Maculopathy: The Beaver Dam Eye Study," *Ophthalmology*, vol. 99, no. 6, pp. 933-943, 1992.
- [3] E. D'Ambrosio, P. Tortorella, and L. Iannetti, "Management of uveitis-related choroidal neovascularization: from the pathogenesis to the therapy," *Journal of Ophthalmology*, vol. 2014, Article ID 450428, 6 pages, 2014.
- [4] D. V. Do, "Detection of new-onset choroidal neovascularization," *Current Opinion in Ophthalmology*, vol. 24, no. 3, pp. 244-247, 2013.
- [5] K. Kinnunen and S. Yla-Herttuala, "Vascular endothelial growth factors in retinal and choroidal neovascular diseases," *Annals of Medicine*, vol. 44, no. 1, pp. 1-17, 2012.

- [6] J. P. Tong and Y. F. Yao, "Contribution of VEGF and PEDF to choroidal angiogenesis: a need for balanced expressions," *Clinical Biochemistry*, vol. 39, no. 3, pp. 267–276, 2006.
- [7] S. Takata, T. Masuda, S. Nakamura et al., "The effect of triamcinolone acetonide on laser-induced choroidal neovascularization in mice using a hypoxia visualization bio-imaging probe," *Scientific Reports*, vol. 5, no. 1, p. 9898, 2015.
- [8] Y. Cai, X. Li, Y. S. Wang et al., "Hyperglycemia promotes vasculogenesis in choroidal neovascularization in diabetic mice by stimulating VEGF and SDF-1 expression in retinal pigment epithelial cells," *Experimental Eye Research*, vol. 123, pp. 87–96, 2014.
- [9] Y. Saishin, R. L. Silva, Y. Saishin et al., "Periocular injection of microspheres containing PKC412 inhibits choroidal neovascularization in a porcine model," *Investigative Ophthalmology & Visual Science*, vol. 44, no. 11, pp. 4989–4993, 2003.
- [10] L. Sun, T. Huang, W. Xu, J. Sun, Y. Lv, and Y. Wang, "Advanced glycation end products promote VEGF expression and thus choroidal neovascularization via Cyr61-PI3K/AKT signaling pathway," *Scientific Reports*, vol. 7, no. 1, p. 14925, 2017.
- [11] R. S. Eshaq, W. S. Wright, and N. R. Harris, "Oxygen delivery, consumption, and conversion to reactive oxygen species in experimental models of diabetic retinopathy," *Redox Biology*, vol. 2, pp. 661–666, 2014.
- [12] K. Nochioka, H. Okuda, K. Tatsumi, S. Morita, N. Ogata, and A. Wanaka, "Hedgehog signaling components are expressed in choroidal neovascularization in laser-induced retinal lesion," *Acta Histochemica et Cytochemica*, vol. 49, no. 2, pp. 67–74, 2016.
- [13] H. Yin, X. Fang, J. Ma et al., "Idiopathic choroidal neovascularization: intraocular inflammatory cytokines and the effect of intravitreal ranibizumab treatment," *Scientific Reports*, vol. 6, no. 1, p. 31880, 2016.
- [14] Y. Zhang, Q. Han, Y. Ru, Q. Bo, and R. H. Wei, "Anti-VEGF treatment for myopic choroid neovascularization: from molecular characterization to update on clinical application," *Drug Design, Development and Therapy*, vol. 9, pp. 3413–3421, 2015.
- [15] J. Kusari, E. Padillo, S. X. Zhou et al., "Effect of brimonidine on retinal and choroidal neovascularization in a mouse model of retinopathy of prematurity and laser-treated rats," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 8, pp. 5424–5431, 2011.
- [16] K. Nakai, M. S. Rogers, T. Baba et al., "Genetic loci that control the size of laser-induced choroidal neovascularization," *FASEB Journal*, vol. 23, no. 7, pp. 2235–2243, 2009.
- [17] K. Fang, P. Gao, J. Tian et al., "Joint effect of CFH and ARMS2/HTRA1 polymorphisms on neovascular age-related macular degeneration in Chinese population," *Journal of Ophthalmology*, vol. 2015, Article ID 821918, 8 pages, 2015.
- [18] C. Saade, B. Ganti, M. Marmor, K. B. Freund, and R. T. Smith, "Risk characteristics of the combined geographic atrophy and choroidal neovascularisation phenotype in age-related macular degeneration," *The British Journal of Ophthalmology*, vol. 98, no. 12, pp. 1729–1732, 2014.
- [19] K. Izumi-Nagai, N. Nagai, K. Ohgami et al., "Macular pigment lutein is antiinflammatory in preventing choroidal neovascularization," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 27, no. 12, pp. 2555–2562, 2007.
- [20] M. Hirasawa, K. Takubo, H. Osada et al., "Angiopoietin-like Protein 2 Is a Multistep Regulator of Inflammatory Neovascularization in a Murine Model of Age-related Macular Degeneration," *The Journal of Biological Chemistry*, vol. 291, no. 14, pp. 7373–7385, 2016.
- [21] L. Paneghetti and Y. S. Ng, "A novel endothelial-derived anti-inflammatory activity significantly inhibits spontaneous choroidal neovascularisation in a mouse model," *Vascular Cell*, vol. 8, no. 1, p. 2, 2016.
- [22] J. Zhang, Y. Suo, Y. H. Zhang et al., "Mining for genes related to choroidal neovascularization based on the shortest path algorithm and protein interaction information," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1860, no. 11, pp. 2740–2749, 2016.
- [23] D. E. Carlin, B. Demchak, D. Pratt, E. Sage, and T. Ideker, "Network propagation in the cytoscape cyberinfrastructure," *PLoS Computational Biology*, vol. 13, no. 10, article e1005598, 2017.
- [24] A. M. Newman, N. B. Gallo, L. S. Hancox et al., "Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks," *Genome Medicine*, vol. 4, no. 2, p. 16, 2012.
- [25] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, 2015.
- [26] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, 2014.
- [27] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, and D. Eisenberg, "DIP: the database of interacting proteins," *Nucleic Acids Research*, vol. 28, no. 1, pp. 289–291, 2000.
- [28] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets," *Nucleic Acids Research*, vol. 34, no. 9, pp. D535–D539, 2006.
- [29] J. Gao, B. Hu, and L. Chen, "A path-based method for identification of protein phenotypic annotations," *Current Bioinformatics*, 2021.
- [30] H. Liu, B. Hu, L. Chen, and L. Lu, "Identifying protein subcellular location with embedding features learned from networks," *Current Proteomics*, vol. 17, 2021.
- [31] X. Zhang, L. Chen, Z. H. Guo, and H. Liang, "Identification of human membrane protein types by incorporating network embedding methods," *IEEE Access*, vol. 7, pp. 140794–140805, 2019.
- [32] X. L. Zhang and L. Chen, "Prediction of membrane protein types by fusing protein-protein interaction and protein sequence information," *Biochimica Et Biophysica Acta-Proteins and Proteomics*, vol. 1868, no. 12, article 140524, 2020.
- [33] Y. Zhang, T. Zeng, L. Chen, S. J. Ding, T. Huang, and Y. D. Cai, "Identification of COVID-19 infection-related human genes based on a random walk model in a virus-human protein interaction network," *BioMed Research International*, vol. 2020, Article ID 4256301, 7 pages, 2020.
- [34] X. Pan, H. Li, T. Zeng et al., "Identification of protein subcellular localization with network and functional embeddings," *Frontiers in Genetics*, vol. 11, article 626500, 2021.
- [35] Y. H. Zhang, T. Zeng, L. Chen, T. Huang, and Y. D. Cai, "Determining protein-protein functional associations by functional rules based on gene ontology and KEGG pathway,"

- Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1869, no. 6, p. 140621, 2021.
- [36] R. Zhao, L. Chen, B. Zhou, Z. H. Guo, S. Wang, and Aorigele, "Recognizing novel tumor suppressor genes using a network machine learning strategy," *IEEE Access*, vol. 7, pp. 155002–155013, 2019.
- [37] R. Zhao, B. Hu, L. Chen, and B. Zhou, "Identification of latent oncogenes with a network embedding method and random forest," *BioMed Research International*, vol. 2020, Article ID 5160396, 11 pages, 2020.
- [38] H. Liang, L. Chen, X. Zhao, and X. Zhang, "Prediction of drug side effects with a refined negative sample selection strategy," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 1573543, 16 pages, 2020.
- [39] H. Y. Liang, B. Hu, L. Chen, S. Wang, and Aorigele, "Recognizing novel chemicals/drugs for anatomical therapeutic chemical classes with a heat diffusion algorithm," *Biochimica et Biophysica Acta-Molecular Basis of Disease*, vol. 1866, no. 11, article 165910, 2020.
- [40] Y. Zhu, B. Hu, L. Chen, and Q. Dai, "iMPTCE-Hnetwork: A Multilabel Classifier for Identifying Metabolic Pathway Types of Chemicals and Enzymes with a Heterogeneous Network," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6683051, 12 pages, 2021.
- [41] J.-P. Zhou, L. Chen, and Z.-H. Guo, "iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs," *Bioinformatics*, vol. 36, no. 5, pp. 1391–1396, 2020.
- [42] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [43] J. Yang, L. Chen, X. Kong, T. Huang, and Y. D. Cai, "Analysis of tumor suppressor genes based on Gene Ontology and the KEGG Pathway," *PLoS One*, vol. 9, no. 9, article e107202, 2014.
- [44] P. Carmona-Saez, M. Chagoyen, F. Tirado, J. M. Carazo, and A. Pascual-Montano, "GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists," *Genome Biology*, vol. 8, no. 1, p. R3, 2007.
- [45] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, "Walking the interactome for prioritization of candidate disease genes," *The American Journal of Human Genetics*, vol. 82, no. 4, pp. 949–958, 2008.
- [46] Z. Yu, R. Visse, M. Inouye, H. Nagase, and B. Brodsky, "Defining Requirements for Collagenase Cleavage in Collagen Type III Using a Bacterial Collagen System," *The Journal of Biological Chemistry*, vol. 287, no. 27, pp. 22988–22997, 2012.
- [47] X. Ji, L. Wang, B. Wu et al., "Associations of MMP1, MMP2 and MMP3 genes polymorphism with coal workers' pneumoconiosis in Chinese Han population," *International Journal of Environmental Research and Public Health*, vol. 12, no. 11, pp. 13901–13912, 2015.
- [48] B. Steen, S. Sejersen, L. Berglin, S. Seregard, and A. Kvanta, "Matrix metalloproteinases and metalloproteinase inhibitors in choroidal neovascular membranes," *Investigative Ophthalmology & Visual Science*, vol. 39, no. 11, pp. 2194–2200, 1998.
- [49] J. Zhang, J. Zhao, Y. Bai, L. Huang, W. Yu, and X. Li, "Effects of p75 neurotrophin receptor on regulating hypoxia-induced angiogenic factors in retinal pigment epithelial cells," *Molecular and Cellular Biochemistry*, vol. 398, no. 1-2, pp. 123–134, 2015.
- [50] J. Lecomte, K. Louis, B. Detry et al., "Bone marrow-derived mesenchymal cells and MMP13 contribute to experimental choroidal neovascularization," *Cellular and Molecular Life Sciences*, vol. 68, no. 4, pp. 677–686, 2011.
- [51] S. Sarks, S. Cherepanoff, M. Killingsworth, and J. Sarks, "Relationship of basal laminar deposit and membranous debris to the clinical presentation of early age-related macular degeneration," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 3, pp. 968–977, 2007.
- [52] V. K. Katsi, M. E. Marketou, D. A. Vrachatis et al., "Essential hypertension in the pathogenesis of age-related macular degeneration: a review of the current evidence," *Journal of Hypertension*, vol. 33, no. 12, pp. 2382–2388, 2015.
- [53] A. Lommatzsch, P. Hermans, K. D. Müller, N. Bornfeld, A. C. Bird, and D. Pauleikhoff, "Are low inflammatory reactions involved in exudative age-related macular degeneration? Morphological and immunohistochemical analysis of AMD associated with basal deposits," *Graefes' Archive for Clinical and Experimental Ophthalmology*, vol. 246, no. 6, pp. 803–810, 2008.
- [54] M. Singh and S. C. Tyagi, "Metalloproteinases as mediators of inflammation and the eyes: molecular genetic underpinnings governing ocular pathophysiology," *International Journal of Ophthalmology*, vol. 10, no. 8, pp. 1308–1318, 2017.
- [55] K. J. Thornton, E. Kamange-Sollo, M. E. White, and W. R. Dayton, "Role of G protein-coupled receptors (GPCR), matrix metalloproteinases 2 and 9 (MMP2 and MMP9), heparin-binding epidermal growth factor-like growth factor (hbEGF), epidermal growth factor receptor (EGFR), erbB2, and insulin-like growth factor 1 receptor (IGF-1R) in trenbolone acetate-stimulated bovine satellite cell proliferation," *Journal of Animal Science*, vol. 93, no. 9, pp. 4291–4301, 2015.
- [56] R. Roskoski Jr., "The ErbB/HER family of protein-tyrosine kinases and cancer," *Pharmacological Research*, vol. 79, pp. 34–74, 2014.
- [57] K. Nakai, K. Yoneda, T. Moriue, J. Igarashi, H. Kosaka, and Y. Kubota, "HB-EGF-induced VEGF production and eNOS activation depend on both PI3 kinase and MAP kinase in HaCaT cells," *Journal of Dermatological Science*, vol. 55, no. 3, pp. 170–178, 2009.
- [58] H. Ota, A. Itaya-Hironaka, A. Yamauchi et al., "Pancreatic  $\beta$  cell proliferation by intermittent hypoxia via up-regulation of Reg family genes and HGF gene," *Life Sciences*, vol. 93, no. 18-19, pp. 664–672, 2013.
- [59] C. C. Chang, J. J. Chiu, S. L. Chen et al., "Activation of HGF/c-Met signaling by ultrafine carbon particles and its contribution to alveolar type II cell proliferation," *American Journal of Physiology. Lung Cellular and Molecular Physiology*, vol. 302, no. 8, pp. L755–L763, 2012.
- [60] W. Hu, M. H. Criswell, S. L. Fong et al., "Differences in the temporal expression of regulatory growth factors during choroidal neovascular development," *Experimental Eye Research*, vol. 88, no. 1, pp. 79–91, 2009.
- [61] J. R. de Oliveira Dias, E. B. Rodrigues, M. Maia, O. Magalhaes, F. M. Penha, and M. E. Farah, "Cytokines in neovascular age-related macular degeneration: fundamentals of targeted combination therapy," *The British Journal of Ophthalmology*, vol. 95, no. 12, pp. 1631–1637, 2011.
- [62] N. I. Bower, A. J. Vogrin, L. le Guen et al., "VEGFD modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development," *Development*, vol. 144, no. 3, pp. 507–518, 2017.

- [63] T. Duong, K. Koltowska, C. Pichol-Thievend et al., "VEGFD regulates blood vascular development by modulating SOX18 activity," *Blood*, vol. 123, no. 7, pp. 1102–1112, 2014.
- [64] R. M. Hussain and T. A. Ciulla, "Emerging vascular endothelial growth factor antagonists to treat neovascular age-related macular degeneration," *Expert Opinion on Emerging Drugs*, vol. 22, no. 3, pp. 235–246, 2017.
- [65] T. Cabral, L. H. Lima, L. G. M. Mello et al., "Bevacizumab injection in patients with neovascular age-related macular degeneration increases angiogenic biomarkers," *Ophthalmol Retina*, vol. 2, no. 1, pp. 31–37, 2018.
- [66] K. Nakamura, K. Shinozuka, and N. Yoshikawa, "Anticancer and antimetastatic effects of cordycepin, an active component of *Cordyceps sinensis*," *Journal of Pharmacological Sciences*, vol. 127, no. 1, pp. 53–56, 2015.
- [67] T. Matsuo, Y. Okada, F. Shiraga, and T. Yanagawa, "TIMP-1 and TIMP-2 levels in vitreous and subretinal fluid," *Japanese Journal of Ophthalmology*, vol. 42, no. 5, pp. 377–380, 1998.
- [68] P. Ottino, J. Finley, E. Rojo et al., "Hypoxia activates matrix metalloproteinase expression and the VEGF system in monkey choroid-retinal endothelial cells: involvement of cytosolic phospholipase A2 activity," *Molecular Vision*, vol. 10, pp. 341–350, 2004.
- [69] S. J. Jeong, S. Li, R. Luo, N. Strokes, and X. Piao, "Loss of Col3a1, the gene for Ehlers-Danlos syndrome type IV, results in neocortical dyslamination," *PLoS One*, vol. 7, no. 1, article e29767, 2012.
- [70] D. Horn, E. Siebert, U. Seidel et al., "Biallelic COL3A1 mutations result in a clinical spectrum of specific structural brain anomalies and connective tissue abnormalities," *American Journal of Medical Genetics. Part A*, vol. 173, no. 9, pp. 2534–2538, 2017.
- [71] S. Caballero, R. Yang, M. B. Grant, and B. Chaqour, "Selective blockade of cytoskeletal actin remodeling reduces experimental choroidal neovascularization," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 5, pp. 2490–2496, 2011.
- [72] Y. Bai, M. Zhao, C. Zhang et al., "Anti-angiogenic effects of a mutant endostatin: a new prospect for treating retinal and choroidal neovascularization," *PLoS One*, vol. 9, no. 11, article e112448, 2014.
- [73] M. Aikio, M. Hurskainen, G. Brideau et al., "Collagen XVIII short isoform is critical for retinal vascularization, and overexpression of the Tsp-1 domain affects eye growth and cataract formation," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 12, pp. 7450–7462, 2013.
- [74] K. S. Moulton, B. R. Olsen, S. Sonn, N. Fukai, D. Zurakowski, and X. Zeng, "Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis," *Circulation*, vol. 110, no. 10, pp. 1330–1336, 2004.
- [75] L. Colina-Vegas, W. Villarreal, M. Navarro et al., "Cytotoxicity of Ru(II) piano-stool complexes with chloroquine and chelating ligands against breast and lung tumor cells: interactions with DNA and BSA," *Journal of Inorganic Biochemistry*, vol. 153, pp. 150–161, 2015.
- [76] W. Tang, J. Ma, R. Gu et al., "Lipocalin 2 suppresses ocular inflammation by inhibiting the activation of NF- $\kappa$ B pathway in endotoxin-induced uveitis," *Cellular Physiology and Biochemistry*, vol. 46, no. 1, pp. 375–388, 2018.
- [77] B. Wu, C. Li, Z. Du et al., "Network based analyses of gene expression profile of LCN2 overexpression in esophageal squamous cell carcinoma," *Scientific Reports*, vol. 4, p. 5403, 2014.
- [78] M. Shen, Y. Tao, Y. Feng, X. Liu, F. Yuan, and H. Zhou, "Quantitative proteomic analysis of mice corneal tissues reveals angiogenesis-related proteins involved in corneal neovascularization," *Biochimica et Biophysica Acta*, vol. 1864, no. 7, pp. 787–793, 2016.
- [79] L. Wu, Y. Du, J. Lok, E. H. Lo, and C. Xing, "Lipocalin-2 enhances angiogenesis in rat brain endothelial cells via reactive oxygen species and iron-dependent mechanisms," *Journal of Neurochemistry*, vol. 132, no. 6, pp. 622–628, 2015.
- [80] S. D. Mesquita, A. C. Ferreira, A. M. Falcao et al., "Lipocalin 2 modulates the cellular response to amyloid beta," *Cell Death and Differentiation*, vol. 21, no. 10, pp. 1588–1599, 2014.
- [81] K. M. Rood, I. A. Buhimschi, K. Rodewald Millen et al., "Evidence for participation of neutrophil gelatinase-associated lipocalin/matrix metalloproteinase-9 (NGAL•MMP-9) complex in the inflammatory response to infection in pregnancies complicated by preterm birth," *American Journal of Reproductive Immunology*, vol. 76, no. 2, pp. 108–117, 2016.
- [82] S. Ghosh, P. Shang, M. Yazdankhah et al., "Activating the AKT2-nuclear factor- $\kappa$ B-lipocalin-2 axis elicits an inflammatory response in age-related macular degeneration," *The Journal of Pathology*, vol. 241, no. 5, pp. 583–588, 2017.