Research Article

Computational Approach to Investigating Key GO Terms and KEGG Pathways Associated with CNV

YuanYuan Luo,[1] Yan Yan,[1] Shiqi Zhang,[2] and Zhen Li [1]

[1]*Department of Ophthalmology, School of Medicine, Renji Hospital, Shanghai Jiao Tong University, Shanghai 200127, China*
[2]*Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark*

Correspondence should be addressed to Zhen Li; lizhen1981_1@126.com

Choroidal neovascularization (CNV) is a severe eye disease that leads to blindness, especially in the elderly population. Various endogenous and exogenous regulatory factors promote its pathogenesis. However, the detailed molecular biological mechanisms of CNV have not been fully revealed. In this study, by using advanced computational tools, a number of key gene ontology (GO) terms and KEGG pathways were selected for CNV. A total of 29 validated genes associated with CNV and 17,639 nonvalidated genes were encoded based on the features derived from the GO terms and KEGG pathways by using the enrichment theory. The widely accepted feature selection method—maximum relevance and minimum redundancy (mRMR)—was applied to analyze and rank the features. An extensive literature review for the top 45 ranking features was conducted to confirm their close associations with CNV. Identifying the molecular biological mechanisms of CNV as described by the GO terms and KEGG pathways may contribute to improving the understanding of the pathogenesis of CNV.

1. Introduction

Choroidal neovascularization (CNV) is a serious eye disease involving the abnormal growth of blood vessels in the choroid region [1–3]. The growth originates from a break in Bruch's membrane, and subsequently the new blood vessels penetrate into the subretinal pigment epithelium [4]. From an epidemiological perspective, CNV is a major cause of pathological visual loss in aging populations [5]. Clinically, age-related macular degeneration (ARMD), myopia, and presumed ocular histoplasmosis syndrome (POHS) are the three major pathogeneses attributed to CNV [6–8]. The Wisconsin Beaver Dam Eye Study [9] confirmed that up to 90% of visual loss in ARMD is secondary to CNV. Given that ARMD is the most common cause of visual loss in people older than 50 years, CNV is speculated to be directly linked to such pathological visual loss. Aside from ARMD, the other two pathological processes—myopia [10] and POHS [11]—are also linked to pathological visual loss. This finding validates the specific role of CNV in pathological visual loss.

Clinically, most patients with CNV share a group of characteristic signs and symptoms, including painless loss of vision, metamorphopsia, paracentral or central scotoma, and apparent changes in image size perception [12, 13]. Generally, patients with these complaints need further physical examinations on blood, fluid, lipid exudation, and retinal pigment epithelial detachment for accurate diagnoses [14]. However, for the final differential diagnosis, laboratory tests are the golden criteria. Generally, the laboratory studies for CNV involve three main techniques, namely, fluorescein angiography [15], indocyanine green angiography [16], and spectral domain optical coherence tomography [17]. For the patients with confirmed diagnosis, due to the unclear pathological mechanisms of CNV, anti-VEGF treatment that counters angiogenesis is the only preferred clinical therapeutic approach [17]. However, the injection burden limits the long-term application of such anti-VEGF treatment. Therefore, more detailed pathological mechanisms of CNV need to be revealed to promote the development and application of new drugs against the disease.

Recent publications have partially revealed the detailed pathological mechanisms of CNV, which involve the interactions between genetic factors and exogenous environments. For the environmental factors, the personal physical factors

induced by the exogenous factors are directly involved in the pathogenesis [18]. Age, obesity, high cholesterol, and high blood pressure aggravate the progression of CNV and further contribute to the occurrence of complications [19, 20]. Aside from these so-called physical exogenous factors, various genetic factors are also connected to the initiation and progression of CNV. Given that CNV is a highly specific disease with an abnormal angiogenesis, genes associated with angiogenesis, such as VEGF [21] and FGF2 [22], definitely participate in the pathological processes, which have been widely confirmed by reliable experiments. In addition to these genes, a specific gene called CFI participates in CNV and induces gradual visual loss and myopia; this finding is based on the sequencing data of CNV families [23]. Furthermore, a specific study [24] on the East Asian population with 2119 patients and 5691 controls revealed a group of effective hereditary and sporadic virulence genes that participate in CNV, mapping out the detailed genetic blueprint of CNV. Some trials were also conducted in the bioinformatics field. Zhang et al. [25] presented a specific computational routine for the identification of CNV-associated genes, indicating the efficacy and accuracy of computational application in such field.

As mentioned earlier, the genetic basis and the environmental influences of CNV have been revealed. However, its biological molecular mechanisms have not been explained thoroughly. Here, the detailed biological processes, cellular components, and molecular functions that may participate in the pathogenesis of CNV were screened out by using computational methods. In this study, GO [26] and KEGG [26, 27] pathways were introduced as two effective bioinformatics tools to accurately describe such items [27]. Based on widely known biological processes associated with CNV, an effective network was rebuilt, and novel biological processes described by the GO and KEGG items were screened out. Recent publications have validated these highly correlative biological processes, thus supporting the efficacy and accuracy of our prediction. With the use of computational methods, a group of functional biological processes that may participate in the potential pathogenesis of CNV were screened out, and for the first time, the detailed pathological mechanisms of CNV were described at the level of comprehensive biological processes instead of genes. The results contributed to the understanding of the development and progression of CNV.

2. Materials and Methods

This study aimed to extract some key GO terms and KEGG pathways that share close biological associations with CNV by using a computational framework. The flowchart of our method is illustrated in Figure 1 for the easy understanding of this work.

2.1. Materials. In 2012, Newman et al. [28] reported a number of genes that are related to AMD. We downloaded the "Additional file 3" in their study [28], in which genes associated with AMD in literature either by genetic linkage or as expression biomarker were listed. Since CNV was a subtype of AMD, we further filtered the genes. Only the genes in
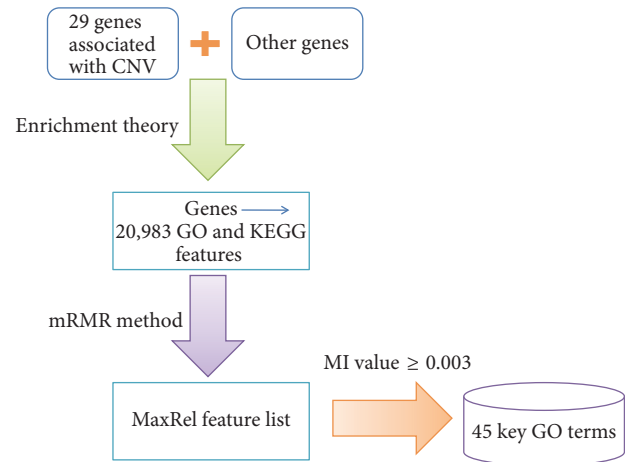

FIGURE 1: Flowchart of selecting the key GO terms and KEGG pathways related to CNV.

CNV Up or CNV Down modules from "Additional file 5" in Newman et al.'s study were kept and at last, 35 CNV genes were obtained. CNV Up or CNV Down modules were generated by network clustering of differentially expressed genes with a permuted $p < 0.1$ and fold change ≥ 1.5 among 31 normal, 7 MD1, 4 MD2, 17 Dry AMD, 2 GA, 4 CNV, and 3 GA/CNV samples. Therefore, the final CNV genes we used were both reported by literature and differentially expressed.

The obtained 35 CNV genes were mapped onto their Ensembl IDs. We excluded IDs that are not in the PPI network reported in STRING (Version 10.0) [29]. 38 Ensembl IDs were accessed. The GO terms and KEGG pathways were used to investigate the difference between CNV-related genes and others; thus, the Ensembl IDs without a GO term and KEGG pathway information were excluded. A total of 29 Ensembl IDs were left. These IDs were the positive samples in this study. The other 17,639 Ensembl IDs were the negative samples and comprised the dataset together with the positive samples in this study. The genes belonging to the positive and negative samples are provided in Supplementary Material S1.

2.2. Feature Vector. The goal of this study was to refine important GO terms and KEGG pathways that are associated with CNV genes. To fulfill that goal, all the genes in the dataset were needed to be represented by all the GO terms and KEGG pathways. Here, the enrichment theory [30] of the GO term and KEGG pathway was used to transform the genes into numeric values, which indicated the biological relationships between the genes and GO terms (KEGG pathways). Comparing with the direct binary annotation of whether a gene has a specific GO term or KEGG pathway, the score obtained by the enrichment theory can indicate the significance of overlap between a gene set and a GO or KEGG function in the genome background. It is more robust than the binary qualitative measurement [31]. To date, this theory has been widely applied to investigate different gene- or protein-related problems [30, 32–41]. After each gene was represented by a larger number of features, by applying a

feature selection method described in Section 2.3, the key GO term or KEGG pathway features were extracted to distinguish the difference between the positive and negative samples. The encoding procedure follows.

GO Enrichment Score. The GO enrichment score was utilized to represent the association between a GO term and an involved gene as a numeric value. For a given GO term, such as GO_j, and a gene g, the gene set G_1 consisted of genes annotated to GO_j and gene set G_2 consisted of the neighbor genes of g in the protein–protein interaction network reported in STRING (http://string-db.org/) [29], a well-organized database providing known and predicted protein–protein interactions. On the basis of the preceding items, the GO enrichment score of GO_j and g can be defined as the $-\log_{10}$ of the hypergeometric test p value [30, 32–35] of G_1 and G_2 according to the following equation:

$$\mathrm{ES_{GO}}\left(g, \mathrm{GO}_j\right) = -\log_{10}\left(\sum_{k=m}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\right), \qquad (1)$$

where N is the total number of genes in humans, M is the number of genes in G_1, n is the number of genes in G_2, and m is the number of the common genes of G_1 and G_2. A large enrichment score of GO_j and g indicated a close relationship between them. In this study, 20,686 GO terms were considered. Thus, 20,686 GO enrichment scores were calculated for each gene in the dataset, which were obtained by using an in-house program using R function phyper. The R code is "score $\leftarrow -\log10(\mathrm{phyper(numWdrawn} - 1, \mathrm{numW}, \mathrm{numB}, \mathrm{numDrawn}, \mathrm{lower.tail} = \mathrm{FALSE}))$," where numW, numB, and numDrawn correspond to the number of genes annotated to GO_j, number of genes not annotated to GO_j, and number of neighbor genes g, respectively.

KEGG Enrichment Score. Similar to the GO enrichment score, the KEGG pathway score was calculated using the same theory to represent the quantitative associations between the KEGG pathways and genes in the dataset. For a given KEGG pathway K_j and a gene g, G_1 was a gene set containing genes in K_j and G_2 had the same meaning as described in preceding paragraph. The KEGG enrichment score shared a similar definition with the GO enrichment score between K_j and g, which was formulated as

$$\mathrm{ES_{KEGG}}\left(g, K_j\right) = -\log_{10}\left(\sum_{k=m}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\right), \qquad (2)$$

where N, M, n, and m share similar definitions as described in (1). In addition, a high score yielded by a KEGG pathway K_j and a gene g indicated their strong associations. Here, 297 KEGG pathways were considered and resulted in 297 KEGG enrichment scores for each gene, which were also obtained by using an in-house program using R function phyper.

Accordingly, each gene in the dataset was encoded by a combination of 20,686 GO term and 297 KEGG pathway

features and was defined as a feature vector with a total of 20,983 elements:

$$
\begin{aligned}
f\left(g\right) = &\left(\mathrm{ES_{GO}}\left(g, \mathrm{GO}_1\right), \ldots, \mathrm{ES_{GO}}\left(g, \mathrm{GO}_{20686}\right),\right.\\
&\left.\mathrm{ES_{KEGG}}\left(g, K_1\right), \ldots, \mathrm{ES_{KEGG}}\left(g, K_{291}\right)\right)^T.
\end{aligned}
\qquad (3)
$$

2.3. Feature Selection. As described in Section 2.2, each gene in the dataset was encoded with 20,983 features derived from the GO terms and KEGG pathways. Some of them shared closer biological associations with CNV. Thus, advanced tools were necessary to extract these important features that played essential roles in the development of CNV. Here, a reliable and widely accepted feature selection method, namely, maximum relevance and minimum redundancy (mRMR) [42], was adopted to analyze all 20,983 features. The mRMR method, proposed by Peng et al. [42], is a useful tool to analyze the feature space of complicated biological problems. To date, many investigations related to complicated biological systems or problems have applied this method to analyze their feature space and extract important information [34, 36, 43–52].

In the mRMR method, two excellent criteria were proposed to rank the features: (1) maximum relevance and (2) minimum redundancy. According to their names, the former criterion measures the importance of features by relying on their correlation to target variable, whereas the latter criterion provides a guarantee that the selected features also have minimum redundancies. If one decides to construct an optimal feature subspace, both maximum relevance and minimum redundancy should be used. In this study, the purpose was to extract important features that are closely related to CNV rather than construct an optimal feature subspace. Therefore, only the criterion of maximum relevance was employed to rank the features in this study. The maximum relevance of each feature was measured by the mutual information (MI) between the feature and the target variable. For each feature, f was a variable representing the values in all samples and c was the target variable. The MI was calculated as follows:

$$I\left(c, f\right) = \iint p\left(c, f\right) \log \frac{p\left(c, f\right)}{p\left(c\right) p\left(f\right)} dc\, df, \qquad (4)$$

where $p(c)$ and $p(f)$ are the marginal probabilities of c and f and $p(c, f)$ is their joint probabilistic distribution. According to (4), MI measures the mutual dependence between two variables.

Based on the MI value assigned to each feature, the feature ranking list called MaxRel feature list was obtained and formulated as follows:

$$F = \left[f_1, f_2, \ldots, f_N\right], \qquad (5)$$

where N is the total number of features in the feature space. A high rank received by a feature indicates a strong association with CNV. Based on the properties of the top ranked features, a new insight into the CNV can be proposed for the investigation of the corresponding GO terms and KEGG pathways.
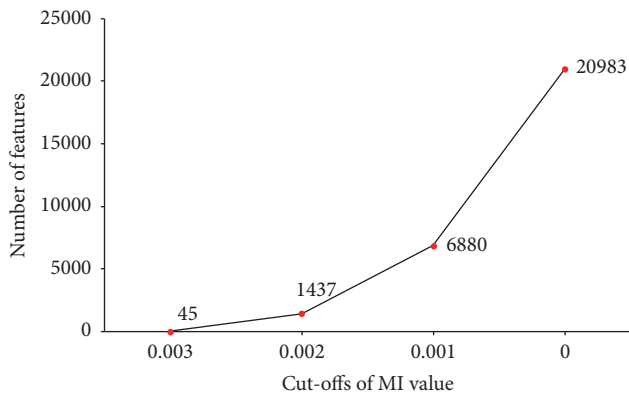
Figure 2: The number of selected features under different cut-offs of MI values.

3. Results and Discussion

3.1. Results. As described in Section 2.2, a total of 20,983 GO terms and KEGG pathway features were encoded in each gene in dataset. Then, according to their relevance to the target variables, these features were ranked in the descending order by using the maximum relevance criterion described in Section 2.3. The output feature list, called MaxRel feature list, was built and obtained (Supplementary Material S2).

As mentioned in the preceding paragraphs, not all GO terms or KEGG pathways shared equal roles on influencing the progression of CNV. Thus, extracting the key GO terms or KEGG pathways was necessary. By applying the maximum relevance criterion, all the features were ranked by their relevance to the target variables, and the rank of a corresponding feature in the output MaxRel feature list for a GO term or KEGG pathway indicated its association with CNV. According to their MI values, some GO terms or KEGG pathways received high MI values in the MaxRel feature list; these features were extracted and their importance was further investigated. To determine the cut-off of MI value, a curve was plotted in Figure 2, which shows the number of selected features under different cut-offs of MI value. It can be observed that the cut-off 0.003 was a proper choice, resulting in 45 features. These features would be given further literature review. Their detailed information is listed in Table 1. All the 45 features corresponded to important GO terms. The following section provides a detailed discussion on these GO terms.

3.2. Analysis of Key GO Terms. As mentioned earlier, based on our current computational methods, a group of functional biological processes that may directly contribute to the initiation and progression of CNV as a pathological mechanism were screened out. In the prediction list, the top 45 biological processes described by the GO terms as optimal CNV-associated biological processes were selected. Due to the limitation of such manuscript, an individual analysis of all the items was not feasible. Therefore, the top terms were chosen, and their respective connection with CNV according to recent publications was discussed. According to recent publications, such GO terms can be summarized into three major subgroups: angiogenesis, local neural metabolism, and immune-associated biological processes. The detailed discussion follows.

3.2.1. Analysis of Angiogenesis Associated Biological Processes. The two GO terms in our prediction list—*GO: 0031091* and *GO: 0031093*—both describe the functional cellular components of platelet alpha-granules. In 2015, a specific study [53] on proangiogenic responses confirmed that the release of platelet alpha-granules promotes angiogenesis. No direct connections were revealed between platelet alpha-granules and CNV; however, abnormal angiogenesis plays an irreplaceable role and may be the core pathological biological process during the initiation and progression of CNV [54]. Therefore, GO items associated with platelet alpha-granules, such as cellular components GO: 0031091 and GO: 0031093, are definitely associated with CNV. This result validated the efficacy and accuracy of our prediction.

GO term *GO: 0038133* describes a detailed pathway called the ERBB2-ERBB3 signaling pathway. According to recent publications, this signaling pathway contributes to the regulation of cell survival and tumorigenesis [55, 56]. As for the detailed connections between the ERBB2-ERBB3 signaling pathway and CNV, mediated by miR-199a and miR-125b, ERBB2 and ERBB3 as two functional components of our predicted biological process have been confirmed to contribute to the regulation of vascular endothelial growth factor secretion and the stimulation of angiogenesis in multiple tissues, including the eyes [57–59]. Given the core initiative functions of angiogenesis for CNV, the predicted biological process called the ERBB2-ERBB3 signaling pathway is a potential CNV-associated GO term. Moreover, the next predicted GO term, called *GO: 0038129*, also describes the ERBB3-associated signaling pathway. This finding not only implied the prediction consistency of the current computational methods but also further confirmed the specific role of such pathways during the initiation and progression of CNV.

GO: 0031983 was the next predicted GO term and describes the vesicle lumen as a functional cellular component. As the parent term of GO: 0060205 describing the cytoplasmic vesicle lumen, such cellular component definitely is associated with the initiation and progression of CNV. As for detailed literature evidence, in 2009, a specific study on the vascular permeability and pathological angiogenesis of CNV confirmed that the vesicle lumen in living cells is related to the vascular hyperpermeability and abnormal angiogenesis [60]. Vascular hyperpermeability [61] and abnormal angiogenesis are both specific symptoms of CNV [62]; thus, such biological processes are potential CNV-associated biological processes.

The next GO term, called *GO: 0005576*, describes a general term called extracellular region. Various extracellular substances participate in the pathogenesis of CNV, including LOX [63], LOXL2 [63], Thy-1 [64], and integrins [65]. Such specific substances may play irreplaceable roles during the initiation and progression of CNV; thus, this GO term that describes the extracellular regions of a certain focus is a potential CNV-associated biological process. *GO: 0035767*, as the next predicted GO in our prediction list, describes an effective biological process called endothelial cell chemotaxis.

TABLE 1: Top 45 key GO terms associated with CNV.

GO term ID	GO term	GO description	MI value	Rank
GO: 0031091	Platelet alpha-granule	Cellular component	0.003	1
GO: 0031093	Platelet alpha-granule lumen	Cellular component	0.003	2
GO: 0060205	Cytoplasmic membrane-bounded vesicle lumen	Cellular component	0.003	3
GO: 0038133	ERBB2-ERBB3 signaling pathway	Biological process	0.003	4
GO: 0038129	ERBB3 signaling pathway	Biological process	0.003	5
GO: 1902847	Regulation of neuronal signal transduction	Biological process	0.003	6
GO: 0061517	Macrophage proliferation	Biological process	0.003	7
GO: 1902949	Positive regulation of tau protein kinase activity	Biological process	0.003	8
GO: 0061518	Microglial cell proliferation	Biological process	0.003	9
GO: 0031983	Vesicle lumen	Cellular component	0.003	10
GO: 0005576	Extracellular region	Cellular component	0.003	11
GO: 0035767	Endothelial cell chemotaxis	Biological process	0.003	12
GO: 0002580	Regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	Biological process	0.003	13
GO: 0044421	Extracellular region part	Cellular component	0.003	14
GO: 0007603	Phototransduction, visible light	Biological process	0.003	15
GO: 0001948	Glycoprotein binding	Molecular function	0.003	16
GO: 0072562	Blood microparticle	Cellular component	0.003	17
GO: 0044650	Adhesion of symbiont to host cell	Biological process	0.003	18
GO: 0019062	Virion attachment to host cell	Biological process	0.003	19
GO: 0010466	Negative regulation of peptidase activity	Biological process	0.003	20
GO: 0009584	Detection of visible light	Biological process	0.003	21
GO: 0010951	Negative regulation of endopeptidase activity	Biological process	0.003	22
GO: 0001654	Eye development	Biological process	0.003	23
GO: 0002581	Negative regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	Biological process	0.003	24
GO: 0052547	Regulation of peptidase activity	Biological process	0.003	25
GO: 0052548	Regulation of endopeptidase activity	Biological process	0.003	26
GO: 0005791	Rough endoplasmic reticulum	Cellular component	0.003	27
GO: 0050839	Cell adhesion molecule binding	Molecular function	0.003	28
GO: 0071307	Cellular response to vitamin K	Biological process	0.003	29
GO: 1902430	Negative regulation of beta-amyloid formation	Biological process	0.003	30
GO: 0005604	Basement membrane	Cellular component	0.003	31
GO: 0030023	Extracellular matrix constituent conferring elasticity	Molecular function	0.003	32
GO: 2000768	Positive regulation of nephron tubule epithelial cell differentiation	Biological process	0.003	33
GO: 1903002	Positive regulation of lipid transport across blood brain barrier	Biological process	0.003	34
GO: 1903000	Regulation of lipid transport across blood brain barrier	Biological process	0.003	35
GO: 1903001	Negative regulation of lipid transport across blood brain barrier	Biological process	0.003	36
GO: 1902951	Negative regulation of dendritic spine maintenance	Biological process	0.003	37
GO: 1902999	Negative regulation of phospholipid efflux	Biological process	0.003	38
GO: 1901627	Negative regulation of postsynaptic membrane organization	Biological process	0.003	39
GO: 2001139	Negative regulation of phospholipid transport	Biological process	0.003	40
GO: 0046911	Metal chelating activity	Molecular function	0.003	41
GO: 0030574	Collagen catabolic process	Biological process	0.003	42
GO: 0007423	Sensory organ development	Biological process	0.003	43
GO: 0001968	Fibronectin binding	Molecular function	0.003	44
GO: 0051346	Negative regulation of hydrolase activity	Biological process	0.003	45

Based on recent publications, such biological process is involved in the activation of platelets [66] and exosome-mediated antiangiogenesis [67]. Platelet activation [68] and angiogenesis [69] are directly connected to the initiation and progression of CNV; therefore, this predicted GO term is quite significant in the pathogenesis of CNV.

3.2.2. Analysis of Local Neural Metabolism Associated Biological Processes. *GO: 0060205*, as another cellular component associated item, describes the cytoplasmic vesicle lumen. Based on recent publications, such cellular component participates in autophagy and secretion-associated biological processes in living cells [70, 71]. As for the biological connections between the cytoplasmic vesicle lumen and CNV, the predicted GO-associated biological processes, such as autophagy and substance secretion, have all been confirmed to be involved in the initiation and progression of CNV [72, 73]. This result implied the accuracy and efficacy of our prediction. *GO: 1902847* describes the regulation of neuronal signal transduction. In the biological process of neuronal signal transduction, a specific gene called IKK2 has been confirmed to be significant [74]. Coincidentally, the inhibition of IKK2 has been widely used in the treatment against CNV, indicating the specific role of IKK2 during the pathogenesis of CNV [75, 76]. Therefore, connected by such functional gene IKK2, the predicted biological processes associated with neuronal signal transduction may also be related to CNV. This finding validates the efficacy and accuracy of our prediction. As the next predicted GO, *GO: 1902949* describes the positive regulation of tau protein kinase activity. Tau protein is a major pathological factor that contributes to the initiation and progression of Alzheimer's disease (AD) [77–79]. During the initiation and progression of AD, another specific protein called apolipoprotein E4 (apoE4) interacts with our predicted tau protein [80] and participates in the pathogenesis of AD [81]. Given that recent studies also validated the specific role of apoE4 in neovascularization [82] and its potential functions in CNV [23], tau protein associated kinase activity is reasonably connected to CNV-associated biological characteristics.

3.2.3. Analysis of Immune-Associated Biological Processes. The GO term *GO: 0061517* describes the proliferation of a specific immune-associated cell subgroup: macrophage. Based on recent publications, macrophages contribute to CNV by regulating CCR2-dependent and proangiogenic biological processes [83, 84], indicating that the proliferation of such gene is definitely related to the progression of the disease [85]. Apart from the proliferation of macrophage, the proliferation of another effective cell subgroup called microglial cells is also predicted to contribute to CNV by *GO: 0061518* in our prediction list. Mediated by neuroprotectin D1, microglial ramifications and redistribution participate in the pathological processes of CNV [86]. Therefore, as a functional neuronal cell subtype with specific microglial ramifications [86], this predicted GO is reasonably connected to CNV [87]. Besides these predicted biological processes, several functional GOs in the top 45 predicted GO terms have been reported to participate in CNV-associated biological

processes. These functional GO terms include GO: 0002580 (regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II) [88], GO: 0044421 (extracellular region) [89], and GO: 0007603 (phototransduction, visible light) [90, 91]. These results confirmed the efficacy and accuracy of our prediction.

4. Conclusion

Based on our presented computational method, a group of functional biological functions that have been confirmed by recent publications to be related to the pathogenesis of CNV were screened out. Such predicted biological processes not only further revealed the detailed pathological mechanisms of CNV but also provided a new tool to identify potential functional disease-associated biological processes in multiple categories of the disease. Finally, we will try our best to develop a computational method based on some extracted features in this study to predict novel CNV genes in the future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

YuanYuan Luo and Yan Yan contributed equally to this work.

Supplementary Materials

The supplementary materials consist of two files. Supplementary Material S1: positive and negative samples in the dataset represented by their gene symbols. Supplementary Material S2: MaxRel feature list as an output of the mRMR method with ranked 20,983 features. *(Supplementary Materials)*

References

[1] F. B. Pereira, C. E. Veloso, G. T. Kokame, and M. B. Nehemy, "Characteristics of Neovascular Age-Related Macular Degeneration in Brazilian Patients," *Ophthalmologica*, vol. 234, no. 4, pp. 233–242, 2015.

[2] D. Karagiannis, G. A. Kontadakis, K. Kaprinis et al., "Treatment of myopic choroidal neovascularization with intravitreal ranibizumab injections: The role of age," *Clinical Ophthalmology*, vol. 11, pp. 1197–1201, 2017.

[3] C. M. G. Cheung, J. J. Arnold, F. G. Holz et al., "Myopic choroidal neovascularization: review, guidance, and consensus statement on management," *Ophthalmology*, 2017.

[4] J. J. Jung, C. Y. Chen, S. Mrejen et al., "The incidence of neovascular subtypes in newly diagnosed neovascular age-related macular degeneration," *American Journal of Ophthalmology*, vol. 158, no. 4, pp. 769–779.e2, 2014.

[5] K. Neelam, C. M. G. Cheung, K. Ohno-Matsui, T. Y. Y. Lai, and T. Y. Wong, "Choroidal neovascularization in pathological myopia," *Progress in Retinal and Eye Research*, vol. 31, no. 5, pp. 495–525, 2012.

[6] D. P. Han, J. T. McAllister, D. V. Weinberg, J. E. Kim, and W. J. Wirostko, "Combined intravitreal anti-VEGF and verteporfin

photodynamic therapy for juxtafoveal and extrafoveal choroidal neovascularization as an alternative to laser photocoagulation," *Eye*, vol. 24, no. 4, pp. 713–716, 2010.

[7] M. El Mellaoui, A. El Ouafi, Z. El Hansali, A. Bouzidi, S. Iferkhas, and A. Laktaoui, "Presumed ocular histoplasmosis syndrome," *Journal Français d'Ophtalmologie*, vol. 38, no. 9, pp. 892-893, 2015.

[8] M. Laghmari and O. Lezrek, "Presumed ocular histoplasmosis syndrome (POHS)," *Pan African Medical Journal*, vol. 18, article no. 268, p. 268, 2014.

[9] R. Klein, B. E. K. Klein, and K. L. P. Linton, "Prevalence of age-related maculopathy. The Beaver Dam Eye Study," *Ophthalmology*, vol. 99, no. 6, pp. 933–943, 1992.

[10] I. M. Ghafour, D. Allan, and W. S. Foulds, "Common causes of blindness and visual handicap in the west of Scotland," *British Journal of Ophthalmology*, vol. 67, no. 4, pp. 209–213, 1983.

[11] S. S. Feman, S. F. Podgorski, and M. K. Penn, "Blindness from presumed ocular histoplasmosis in Tennessee," *Ophthalmology*, vol. 89, no. 12, pp. 1295–1298, 1982.

[12] K. Hatz and C. Prünte, "Polypoidal choroidal vasculopathy in Caucasian patients with presumed neovascular age-related macular degeneration and poor ranibizumab response," *British Journal of Ophthalmology*, vol. 98, no. 2, pp. 188–194, 2014.

[13] C. M. G. Cheung, B. K. Loh, X. Li et al., "Choroidal thickness and risk characteristics of eyes with myopic choroidal neovascularization," *Acta Ophthalmologica*, vol. 91, no. 7, pp. e580–e581, 2013.

[14] M. A. Bonini Filho, T. E. de Carlo, D. Ferrara et al., "Association of choroidal neovascularization and central serous chorioretinopathy with optical coherence tomography angiography," *JAMA Ophthalmology*, vol. 133, no. 8, pp. 899–906, 2015.

[15] W. Liu, H. Li, R. S. Shah et al., "Simultaneous optical coherence tomography angiography and fluorescein angiography in rodents with normal retina and laser-induced choroidal neovascularization," *Optics Expresss*, vol. 40, no. 24, pp. 5782–5785, 2015.

[16] A. Kawamura, M. Yuzawa, R. Mori, M. Haruyama, and K. Tanaka, "Indocyanine green angiographic and optical coherence tomographic findings support classification of polypoidal choroidal vasculopathy into two types," *Acta Ophthalmologica*, vol. 91, no. 6, pp. e474–e481, 2013.

[17] P. Iacono, M. Battaglia Parodi, A. Papayannis et al., "Fluorescein angiography and spectral-domain optical coherence tomography for monitoring Anti-VEGF therapy in myopic choroidal neovascularization," *Ophthalmic Research*, vol. 52, no. 1, pp. 25–31, 2014.

[18] I. C. Munch, A. Linneberg, and M. Larsen, "Precursors of age-related macular degeneration: associations with physical activity, obesity, and serum lipids in the Inter99 Eye Study," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 6, pp. 3932–3940, 2013.

[19] C.-C. Hsu, S.-J. Chen, A.-F. Li, and F.-L. Lee, "Systolic blood pressure, choroidal thickness, and axial length in patients with myopic maculopathy," *Journal of the Chinese Medical Association*, vol. 77, no. 9, pp. 487–491, 2014.

[20] X. Zhang, M. Li, F. Wen et al., "Different impact of high-density lipoprotein-related genetic variants on polypoidal choroidal vasculopathy and neovascular age-related macular degeneration in a Chinese Han population," *Experimental Eye Research*, vol. 108, pp. 16–22, 2013.

[21] P. S. Muether, I. Neuhann, C. Buhl, M. M. Hermann, B. Kirchhof, and S. Fauser, "Intraocular growth factors and cytokines in patients with dry and neovascular age-related macular degeneration," *Retina*, vol. 33, no. 9, pp. 1809–1814, 2013.

[22] J. R. O. De Dias, E. B. Rodrigues, M. Maia, O. Magalhat~es Jr., F. M. Penha, and M. E. Farah, "Cytokines in neovascular age-related macular degeneration: Fundamentals of targeted combination therapy," *British Journal of Ophthalmology*, vol. 95, no. 12, pp. 1631–1637, 2011.

[23] N. Leveziel, Y. Yu, R. Reynolds et al., "Genetic factors for choroidal neovascularization associated with high myopia," *Investigative Ophthalmology & Visual Science*, vol. 53, no. 8, pp. 5004–5009, 2012.

[24] C. Y. Cheng, K. Yamashiro, and L. J. Chen, "New loci and coding variants confer risk for age-related macular degeneration in East Asians," *Nature Communications*, vol. 6, p. 6063, 2015.

[25] 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.

[26] The Gene Ontology Consortium, "Gene ontology consortium: going forward," *Nucleic Acids Research*, vol. 43, no. 1, pp. D1049–D1056, 2015.

[27] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Research*, vol. 44, no. 1, pp. D457–D462, 2016.

[28] 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, article 16, 2012.

[29] 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, pp. D447–D452, 2015.

[30] 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, article R3, 2007.

[31] T. Huang, C. Wang, G. Zhang, L. Xie, and Y. Li, "SySAP: a system-level predictor of deleterious single amino acid polymorphisms," *Protein & Cell*, vol. 3, no. 1, pp. 38–43, 2012.

[32] T. Huang, L. Chen, Y. Cai, and K. Chou, "Classification and analysis of regulatory pathways using graph property, biochemical and physicochemical property, and functional property," *PLoS ONE*, vol. 6, no. 9, Article ID e25297, 2011.

[33] T. Huang, J. Zhang, Z. Xu et al., "Deciphering the effects of gene deletion on yeast longevity using network and machine learning approaches," *Biochimie*, vol. 94, no. 4, pp. 1017–1025, 2012.

[34] L. Chen, Y.-H. Zhang, G. Lu, T. Huang, and Y.-D. Cai, "Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways," *Artificial Intelligence in Medicine*, vol. 76, pp. 27–36, 2017.

[35] L. Chen, Y.-H. Zhang, S. Wang, Y. Zhang, T. Huang, and Y.-D. Cai, "Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways," *PLoS ONE*, vol. 12, no. 9, Article ID e0184129, 2017.

[36] L. Chen, Y. Zhang, M. Zheng, T. Huang, and Y. Cai, "Identification of compound–protein interactions through the analysis of gene ontology, KEGG enrichment for proteins and molecular fragments of compounds," *Molecular Genetics and Genomics*, vol. 291, no. 6, pp. 2065–2079, 2016.

[37] J. Li, L. Chen, S. Wang et al., "A computational method using the random walk with restart algorithm for identifying novel epigenetic factors," *Molecular Genetics and Genomics*, vol. 293, no. 1, pp. 293–301, 2018.

[38] S. Lu, Y. Yan, Z. Li et al., "Determination of genes related to uveitis by utilization of the random walk with restart algorithm on a protein–protein interaction network," *International Journal of Molecular Sciences*, vol. 18, no. 5, article no. 1045, 2017.

[39] L. Chen, J. Yang, Z. Xing et al., "An integrated method for the identification of novel genes related to oral cancer," *PLoS ONE*, vol. 12, no. 4, p. e0175185, 2017.

[40] Y. Zhou, B. Li, Y. Zhang, L. Chen, and X. Kong, "Feature classification and analysis of lung cancer related genes through gene ontology and KEGG pathways," *Current Bioinformatics*, vol. 11, no. 1, pp. 40–50, 2016.

[41] 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 ID e107202, 2014.

[42] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[43] S. Wang, Y. Zhang, J. Lu, W. Cui, J. Hu, and Y. Cai, "Analysis and identification of aptamer-compound interactions with a max-imum relevance minimum redundancy and nearest neighbor algorithm," *BioMed Research International*, vol. 2016, Article ID 8351204, 9 pages, 2016.

[44] L. Chen, C. Chu, T. Huang, X. Kong, and Y. Cai, "Prediction and analysis of cell-penetrating peptides using pseudo-amino acid composition and random forest models," *Amino Acids*, vol. 47, no. 7, pp. 1485–1493, 2015.

[45] T. Huang, M. Wang, and Y.-D. Cai, "Analysis of the preferences for splice codes across tissues," *Protein & Cell*, vol. 6, no. 12, pp. 904–907, 2015.

[46] Z. Li, X. Zhou, Z. Dai, and X. Zou, "Classification of G-protein coupled receptors based on support vector machine with maximum relevance minimum redundancy and genetic algorithm," *BMC Bioinformatics*, vol. 11, article 325, 2010.

[47] Q. Ni and L. Chen, "A feature and algorithm selection method for improving the prediction of protein structural classes," *Combinatorial Chemistry High Throughput Screening*, vol. 20, no. 7, pp. 612–621, 2017.

[48] B.-Q. Li, L.-L. Zheng, K.-Y. Feng, L.-L. Hu, G.-H. Huang, and L. Chen, "Prediction of linear B-cell epitopes with mRMR feature selection and analysis," *Current Bioinformatics*, vol. 11, no. 1, pp. 22–31, 2016.

[49] H. Mohabatkar, M. M. Beigi, K. Abdolahi, and S. Mohsenzadeh, "Prediction of allergenic proteins by means of the concept of Chou's pseudo amino acid composition and a machine learning approach," *Medicinal Chemistry*, vol. 9, no. 1, pp. 133–137, 2013.

[50] L. Chen, S. Wang, Y. Zhang et al., "Identify Key Sequence Features to Improve CRISPR sgRNA Efficacy," *IEEE Access*, vol. 5, pp. 26582–26590, 2017.

[51] L. Chen, C. Chu, and K. Feng, "Predicting the types of metabolic pathway of compounds using molecular fragments and sequential minimal optimization," *Combinatorial Chemistry & High Throughput Screening*, vol. 19, no. 2, pp. 136–143, 2016.

[52] L. Chen, Y.-H. Zhang, G. Huang et al., "Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection," *Molecular Genetics and Genomics*, vol. 293, no. 1, pp. 137–149, 2018.

[53] J. Etulain, H. A. Mena, S. Negrotto, and M. Schattner, "Stimulation of PAR-1 or PAR-4 promotes similar pattern of VEGF and endostatin release and pro-angiogenic responses mediated by human platelets," *Platelets*, vol. 26, no. 8, pp. 799–804, 2015.

[54] H. Xu, F. Zeng, D. Shi, X. Sun, X. Chen, and Y. Bai, "Focal choroidal excavation complicated by choroidal neovasculariza-tion," *Ophthalmology*, vol. 121, no. 1, pp. 246–250, 2014.

[55] C. F. McDonagh, A. Huhalov, B. D. Harms et al., "Antitumor activity of a novel bispecific antibody that targets the ErbB2/ErbB3 oncogenic unit and inhibits heregulin-induced activation of ErbB3," *Molecular Cancer Therapeutics*, vol. 11, no. 3, pp. 582–593, 2012.

[56] V. Fock, K. Plessl, P. Draxler et al., "Neuregulin-1-mediated ErbB2-ErbB3 signalling protects human trophoblasts against apoptosis to preserve differentiation," *Journal of Cell Science*, vol. 128, no. 23, pp. 4306–4316, 2015.

[57] J. He, Y. Jing, W. Li et al., "Roles and Mechanism of miR-199a and miR-125b in Tumor Angiogenesis," *PLoS ONE*, vol. 8, no. 2, Article ID e56647, 2013.

[58] L. Yen, X. You, A. Al Moustafa et al., "Heregulin selectively upregulates vascular endothelial growth factor secretion in cancer cells and stimulates angiogenesis," *Oncogene*, vol. 19, no. 31, pp. 3460–3469, 2000.

[59] K. S. Russell, D. F. Stern, P. J. Polverini, and J. R. Bender, "Neuregulin activation of ErbB receptors in vascular endothe-lium leads to angiogenesis," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 277, no. 6, pp. H2205–H2211, 1999.

[60] S.-H. Chang, D. Feng, J. A. Nagy, T. E. Sciuto, A. M. Dvorak, and H. F. Dvorak, "Vascular permeability and pathological angiogenesis in caveolin-1-null mice," *The American Journal of Pathology*, vol. 175, no. 4, pp. 1768–1776, 2009.

[61] P. Daull, C. A. Paterson, B. D. Kuppermann, and J.-S. Garrigue, "A preliminary evaluation of dexamethasone palmitate emul-sion: A novel intravitreal sustained delivery of corticosteroid for treatment of macular Edema," *Journal of Ocular Pharmacology and Therapeutics*, vol. 29, no. 2, pp. 258–269, 2013.

[62] H. Du, X. Sun, M. Guma et al., "JNK inhibition reduces apop-tosis and neovascularization in a murine model of age-related macular degeneration," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 110, no. 6, pp. 2377–2382, 2013.

[63] T. V. Bergen, R. Spangler, D. Marshall et al., "The role of LOX and LOXL2 in the pathogenesis of an experimental model of choroidal neovascularization," *Investigative Ophthalmology & Visual Science*, vol. 56, no. 9, pp. 5280–5289, 2015.

[64] H. Wang, X. Han, E. Kunz, and M. Elizabeth Hartnett, "Thy-1 regulates VEGF-mediated choroidal endothelial cell activation and migration: Implications in neovascular age-related macular degeneration," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 13, pp. 5525–5534, 2016.

[65] T. Nakajima, M. Hirata, T. R. Shearer, and M. Azuma, "Mech-anism for laser-induced neovascularization in rat choroid: accumulation of integrin alpha chain-positive cells and their ligands," *Molecular Vision*, vol. 20, pp. 864–871, 2014.

[66] A. Janowska-Wieczorek, M. Wysoczynski, J. Kijowski et al., "Microvesicles derived from activated platelets induce metas-tasis and angiogenesis in lung cancer," *International Journal of Cancer*, vol. 113, no. 5, pp. 752–760, 2005.

[67] A. R. Hajrasouliha, G. Jiang, Q. Lu et al., "Exosomes from reti-nal astrocytes contain antiangiogenic components that inhibit

laser-induced choroidal neovascularization," *The Journal of Biological Chemistry*, vol. 288, no. 39, pp. 28058–28067, 2013.

[68] K. Birke, E. Lipo, M. T. Birke, and R. Kumar-Singh, "Topical Application of PPADS Inhibits Complement Activation and Choroidal Neovascularization in a Model of Age-Related Macular Degeneration," *PLoS ONE*, vol. 8, no. 10, Article ID e76766, 2013.

[69] J. H. Qi, Q. Ebrahem, M. Ali et al., "Tissue Inhibitor of Metalloproteinases-3 Peptides Inhibit Angiogenesis and Choroidal Neovascularization in Mice," *PLoS ONE*, vol. 8, no. 3, Article ID e55667, 2013.

[70] M. Zhang, S. J. Kenny, L. Ge, K. Xu, and R. Schekman, "Translocation of interleukin-1beta into a vesicle intermediate in autophagy-mediated secretion," *eLife*, vol. 4, 2015.

[71] V. Malhotra and P. Erlmann, "The Pathway of Collagen Secretion," *Annual Review of Cell and Developmental Biology*, vol. 31, pp. 109–124, 2015.

[72] A. Klettner, A. Kauppinen, J. Blasiak, J. Roider, A. Salminen, and K. Kaarniranta, "Cellular and molecular mechanisms of age-related macular degeneration: from impaired autophagy to neovascularization," *The International Journal of Biochemistry & Cell Biology*, vol. 45, no. 7, pp. 1457–1497, 2013.

[73] R. Zhang, Z. Liu, H. Zhang, Y. Zhang, and D. Lin, "The COX-2-selective antagonist (NS-398) inhibits choroidal neovascularization and subretinal fibrosis," *PLoS ONE*, vol. 11, no. 1, Article ID e0146808, 2016.

[74] E. Tsaousidou, L. Paeger, B. F. Belgardt et al., "Distinct Roles for JNK and IKK Activation in Agouti-Related Peptide Neurons in the Development of Obesity and Insulin Resistance," *Cell Reports*, vol. 9, no. 4, pp. 1495–1506, 2014.

[75] S. Gaddipati, Q. Lu, R. B. Kasetti et al., "IKK2 inhibition using TPCA-1-Loaded PLGA microparticles attenuates laser-induced choroidal neovascularization and macrophage recruitment," *PLoS ONE*, vol. 10, no. 3, Article ID e0121185, 2015.

[76] H. Lu, Q. Lu, S. Gaddipati et al., "IKK2 inhibition attenuates laser-induced choroidal neovascularization," *PLoS ONE*, vol. 9, no. 1, Article ID e87530, 2014.

[77] S. Crunkhorn, "Antisense oligonucleotide reverses tau pathology," *Nature Reviews Drug Discovery*, vol. 16, no. 3, pp. 166-166, 2017.

[78] T. Uchihara, K. Endo, H. Kondo et al., "Tau pathology in aged cynomolgus monkeys is progressive supranuclear palsy/corticobasal degeneration- but not Alzheimer disease-like -Ultrastructural mapping of tau by EDX," *Acta Neuropathologica Communications*, vol. 4, no. 1, p. 118, 2016.

[79] H. Malkki, "Alzheimer disease: BACE1 inhibition could block CSF tau increase," *Nature Reviews Neurology*, vol. 13, no. 1, p. 6, 2017.

[80] O. Liraz, A. Boehm-Cagan, and D. M. Michaelson, "ApoE4 induces Aβ42, tau, and neuronal pathology in the hippocampus of young targeted replacement apoE4 mice," *Molecular Neurodegeneration*, vol. 8, no. 1, article no. 16, 2013.

[81] J.-T. Yu, L. Tan, and J. Hardy, "Apolipoprotein e in Alzheimer's disease: An update," *Annual Review of Neuroscience*, vol. 37, pp. 79–100, 2014.

[82] R. Antes, S. Salomon-Zimri, S. C. Beck et al., "VEGF mediates apoE4-induced neovascularization and synaptic pathology in the choroid and retina," *Current Alzheimer Research*, vol. 12, no. 4, pp. 323–334, 2015.

[83] T. A. Krause, A. F. Alex, D. R. Engel, C. Kurts, and N. Eter, "VEGF-production by CCR2-dependent macrophages contributes to laser-induced choroidal neovascularization," *PLoS ONE*, vol. 9, no. 4, Article ID e94313, 2014.

[84] S. Horie, S. J. Robbie, J. Liu et al., "CD200R signaling inhibits pro-angiogenic gene expression by macrophages and suppresses choroidal neovascularization," *Scientific Reports*, vol. 3, article no. 3072, 2013.

[85] L. He and A. G. Marneros, "Doxycycline inhibits polarization of macrophages to the proangiogenic M2-type and subsequent neovascularization," *The Journal of Biological Chemistry*, vol. 289, no. 12, pp. 8019–8028, 2014.

[86] K. G. Sheets, B. Jun, Y. Zhou, and etal., "Microglial ramification and redistribution concomitant with the attenuation of choroidal neovascularization by neuroprotectin D1," *Molecular Vision*, vol. 19, pp. 1747–1759, 2013.

[87] M.-C. Carrasco, J. Navascués, M. A. Cuadros et al., "Migration and ramification of microglia in quail embryo retina organotypic cultures," *Developmental Neurobiology*, vol. 71, no. 4, pp. 296–315, 2011.

[88] P. L. Penfold, J. G. Wong, J. Gyory, and F. A. Billson, "Effects of triamcinolone acetonide on microglial morphology and quantitative expression of MHC-II in exudative age-related macular degeneration," *Clinical & Experimental Ophthalmology*, vol. 29, no. 3, pp. 188–192, 2001.

[89] S. Binder, B. V. Stanzel, I. Krebs, and C. Glittenberg, "Transplantation of the RPE in AMD," *Progress in Retinal and Eye Research*, vol. 26, no. 5, pp. 516–554, 2007.

[90] V. P. Papastefanou, V. Nogueira, G. Hay et al., "Choroidal naevi complicated by choroidal neovascular membrane and outer retinal tubulation," *British Journal of Ophthalmology*, vol. 97, no. 8, pp. 1014–1019, 2013.

[91] M. B. Parod, P. Iacono, and F. Bandello, "Correspondence of leakage on fluorescein angiography and optical coherence tomography parameters in diagnosis and monitoring of myopic choroidal neovascularization treated with bevacizumab," *Retina*, vol. 36, no. 1, pp. 104–109, 2016.